JOHN LAW: A LIFETIME'S CONTRIBUTION TO MONETARY ECONOMICS

BY ROISIN DUFF

Senior Sophister

The name of John Law is usually associated with the consequences of the failure of his Mississippi System. However, as Roisin Duff argues in this essay, his contribution to monetary thought is of great significance. The author reviews the life and analyses the work of John Law to conclude that his status as a monetary theorist is exceptional.

"He worked out the economics of his projects with a brilliance and, yes, profundity which places him in the front ranks of monetary theorists of all times"

(Schumpeter, 1954).

Introduction

A gambler, a womaniser and a fugitive from the law. This description is not exactly associated with an economist and yet this is a perfect illustration of John Law, particularly in the early years of his life. From this questionable background emerged a monetary theorist and policy maker who would create and subsequently destroy an economy. To some, he is regarded as a failure. His contributions to monetary thought are often overlooked. This is primarily on account of the devastating application of Law's theory to the French economy. In his principal works, *Essay on a Land Bank* (c. 1704) and *Money and Trade Considered with a Proposal for Supplying the Nation with Money* (1705), Law introduces many contemporary concepts. These alone should have guaranteed him a place in the front ranks of monetary theorists of all time. These concepts will be analysed accordingly. The Mississippi System will subsequently be analysed and the flaws identified. I aim to show that Law was in fact well advanced in his thoughts on monetary issues and that he has been misrepresented as a failure. His contributions to monetary theory alone have earned him an elevated status within the monetary realm.

Background to John Law

Law was born the son of a Scottish financier in Edinburgh in 1671. He was educated in political economy, commerce and economics in London. Much, of his younger years were occupied with gambling and women. 'Jessamy' and 'Beau Law' were some of the diminutives that he acquired due to his activities, which were not associated with banking at this time. In 1694, a duel took place in Bloomsbury Square, which resulted in the death of Edward Wilson, at the hand of Law. He was imprisoned for murder. It was at this time that Patterson established the Bank of England. This illustrates the advances that were being made in monetary issues at this time, to which Law was currently detached from. However, he subsequently escaped from prison, and fled to the continent and vanished. When he re-appeared he was a changed man. It was John Law the banker, but it was banking with a difference, involving as it did acting as the banker at the gaming table (Murphy, 1997).

Analysis of Law's work

Essay on a Land Bank is the first known work written by John Law on monetary economics. The Essay shows clarity of thought in monetary issues considerably superior to contemporary writings and clearly distinguishable from previous works on money (Murphy, 1994; 1997). In the context of his life, this was produced when Law had fled England. In 1704-1705, he resurfaced and took part in the land bank debate. Law was not the originator of this line of thought on land banks, Hugh Chamberlen being Law's chief rival in this area (Ibid, 1997). He presented many proposals between 1703-1707 for banks, which were subsequently rejected in Scotland, England and France (Ibid, 1994).

Essay on a Land Bank concentrates on defining and analysing the functions and qualities of money. In order for Law to unveil the merits of his bank, he firstly needed to discuss monetary issues. In doing this he produced an outstanding analysis of the functions, qualities and types of money (Ibid, 1994). Within the opening folio of the manuscript Law defines money: "Money is used as the measure by which goods are valued, as the value by which goods are exchanged, in which contracts are made payable and payments are made". This definition clearly outlines the three functions that Law deems necessary for money to possess. Money is firstly a measure of value, secondly it is a means of payment or medium of exchange and thirdly it is a standard of deferred payment. Still today these functions of money exist, which is a further justification for elevating Law's status within monetary economics.

Law proceeds to segregate money into categories. Category one includes money that fulfils the functions of money listed, while possessing six qualities – stability of value, homogeneity, durability, divisibility, transferability and a stamp

that denotes its value. Category two includes money that fulfils only the payment and exchange functions of money and possesses all the qualities with the exception of a stable value. This would include Bank of England Shares. Law expresses his ideas with a profound understanding of monetary issues. It was the first comprehensive assessment of the functions of money.

He links the important relationship between the demand for goods and the quantity available of them. This was not an original concept. As far back as Plato the paradox of value was dwelled upon "Only what is rare is valuable, and water which is the best of all things…is also the cheapest". Davanzati, Pindar and Locke have used the paradox previously in their work also. However, Law approached this analysis from an alternative viewpoint in that he used supply and demand analysis while explaining the paradox. Diamonds and water are the goods he used. He proceeded to explain that water is a necessity of life and the quantity available of it is in abundance. However, diamonds have limited use and yet because the quantity available of them is limited, the demand for them is great. Therefore: "the value of goods is rated not as the uses they are applied to are more or less necessary but as they are in quantity in proportion to the demand for them." (Law, 1704). Law's contribution to monetary thought in this respect was overlooked because this analysis was borrowed by Adam Smith and due credit was not given to Law.

Law's essay proceeds, when he has clarified the monetary issues necessary, to focus on the replacement of specie money with money based on land, his proposal for a land bank. Having previously analysed money in detail, he argues that a money based on land can fulfil those functions and qualities of money better than silver: "Land is a commodity and may be made money but, it serves the uses of money and produces at the same time" (Ibid). The analysis is based on money demand and money supply, which was a revolutionary concept at the time. Silver is liable to changes in its value. This he explains by illustrating that although the quantity, quality and demand for silver is the same in two consecutive years, if the good – barley in his illustration, is better quality, in less quantity or if demand is greater this will result in a greater price of silver than the previous year (Ibid). This is one such argument that Law makes. The analysis is extended to each of the functions and in each case, by the use of supply and demand analysis; Law substantiates his argument in a clear and concise fashion.

Previous to the application of his ideas to a system, Law, in 1705 published *Money and Trade Considered with a Proposal for Supplying the Nation with Money*. In this he amalgamated money and trade and assessed how far money affects trade. If Law had to end his monetary career at this stage in his life, at 34 years of age, he would no doubt have been listed with the great monetary theorists of all time. *Money and Trade* clearly outlines the issues to be discussed – three in total. Firstly, the nature of money, this is mostly a reiteration of many of the issues in *Essay on a*

Land Bank and as will be seen the parallels between them are unquestionable. Secondly, the relationship between money and trade is examined, which in Law's work was a new concept. Thirdly, a policy issue of how to produce a new monetary structure capable of expanding the money supply is introduced. Law wishes to expand the money supply, but is aware of the need for a new monetary structure to be put in place in order to achieve this. He examines how to go about this task.

Nature of money

Law reiterates that goods have a value derived from the uses they are applied to. Their value depends on the quantity supplied in relation to the quantity demanded of them. Excessive supply could make a highly useful product valueless. The definition of money is restated with a slight alteration: "Money is the measure by which goods are valued, the value by which goods are exchanged and in which contracts are made payable". In this definition, payment and exchange are treated as synonyms, while the functions remain the same (Murphy, 1997). Law reiterates the flaws that silver, as a form of money possesses within chapter 1. By the concluding paragraph in chapter 1 trade is introduced for the first time:

> "As money encreas'd, the Disadvantages and Inconvencies of Barter were remov'd; the Poor and Idle were employ'd, more of the Land was Labour'd, the Product encreas'd, Manufactures and Trade improv'd, the Landed-men Lived better, and the People with less Dependence on them" (Ibid).

This illustrates the symbiotic relationship between money and trade. His train of thought had developed to include the exchange of goods with other countries. Essentially, he introduces four further concepts – the circular flow of income, the money in advance requirement, further analysis of international inflation in a money supply/money demand framework and finally the formulation of the law of one price for a small open economy (Ibid).

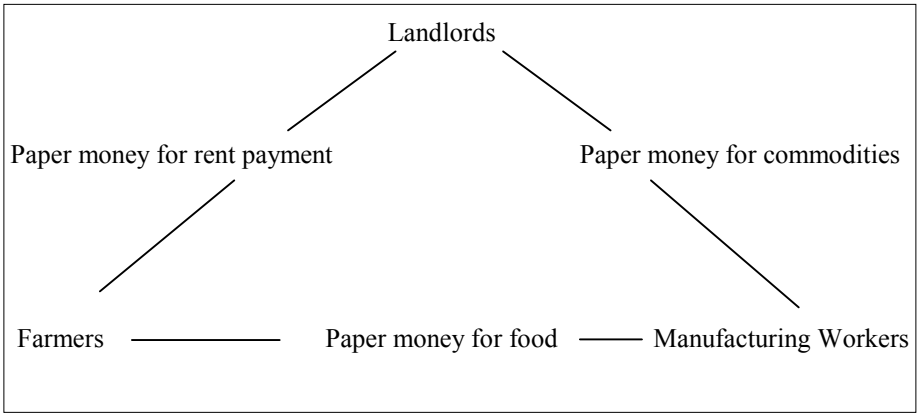
Law begins to consider money in a wider macro-economic context. Law believed that 'trade depends on money'. If money is relatively plentiful, more labour can be employed. Since production should increase with the addition to labour, this will increase the potential for trade. Employment and output are included in the discourse. The interrelationship between these issues is clear to Law: "More money, by employing more people, will make an overplus to export". An inadequate money supply is then seen as the cause of the low level of trade, which in turn suggested unemployment and under utilisation of resources. "As trade depends on money, so the increase or decrease of the people depends on Trade" (Law, 1705). Next Law introduces exports and imports and identifies trade surpluses, when exports are larger than imports. Divergences between the domestic and foreign prices of goods can occur and when this does it can cause discrepancies in the cost of these goods.

There is a danger if the price is lower in other countries of it being imported at a lower cost and so undercutting domestic production. The Law of One Price is introduced. Irrespective of the money in circulation in a country, the price of traded goods must be equal to those in the world market (Murphy, 2003).

The availability of money is a key ingredient in the analysis with the use of Banks being "the best method yet practis'd for the increase of money" (Law, 1705). Law attributed unemployment to a shortage of money, but not to laziness. "A part of the people then employed being now idle, not for want of inclination to work, or for want of employers, but for want of money to pay them with". The lack of money creates unemployed resources, which will hinder trade and production, as they are not realising their full potential.

Chapter 7 is occupied with a formal presentation of the circular flow of income model. The essence of this is that money flows between three groups – the landlords, farmers and manufacturing workers. The landlord or proprietor initiates the circular flow. They pay the labourers in the newly created manufacturing sector with paper money for their goods and services. The labourers use the money to purchase corn and other agricultural goods from the tenant farmers. The tenant farmers use the money to pay the landlord his rent. Therefore, money flows between the three groups, financing the payment of goods and services as well as the payment of the rent. The money returns to the landlord at the end of the process enabling him to start a further round of economic activity (Murphy, 1997). The expansion of this model will need an injection of cash. Law is suggesting that money is needed in advance to create a stimulus to the trading process. Robert Clower develops this to a greater extent. Law's additional monetary contributions in Money and Trade reinforce his capabilities as a monetary economist.

Figure 1: The circular flow of income as suggested by Law



The Mississippi System created by Law

Law's proposals were rejected on numerous occasions, although he persisted. He wanted a further ambition to be realised, that of John Law the policy maker. Law's success in this area was ultimately the cause of his downfall. In the early 18th century the economy of France was heavily burdened with debt and intolerably high taxes, internal commerce was virtually at a standstill, exports almost reduced to vanishing point and the treasury was depleted, so much so that the armies had neither regular pay, food nor clothing (Hyde, 1969). The country was in utter devastation after the War of the Spanish Succession, which broke out in 1701. This depressed French economic environment was fertile ground for Law to experiment with his monetary and economic ideas. When Louis XIV died in 1715, Phillipe Duke of Orleans served as ruler until the 5-year-old Louis XV became of age. This was to be of considerable advantage to Law, as they were acquaintances. This association with the Duke of Orleans would ensure Law's place in history (Moen, 2003). He convinced the Duke to allow him to open a conventional, note-issuing bank in June 1716, the *Banque Generale* (Garber, 2000). This was the beginning of Law's system, a system, which was the genuine ancestor of the idea of managed currency (Schumpeter, 1954). The bank was the first milestone on the road towards the goal he had set himself, the financial and commercial rehabilitation of his adopted country (Hyde, 1969).

The foundation of this bank was built on Law's knowledge of economic and monetary theory to date. It was authorised to take part in banking functions exclusively, to issue notes, to carry and transfer demand deposits and to discount promissory notes and bills of exchange (Hamilton, 1936). Law was the appointed managing director. Individuals would deposit their specie money with the bank and in return they were issued notes to that value. They could reclaim the specie on demand, but the bank notes would circulate as a medium of exchange in the meantime. The idea seemed somewhat simplistic because where something so important is involved; a deeper mystery seems only decent (Galbraith, 1975). Initially the public was inclined to be sceptical of such a novel institution as a bank of issue, and it regarded the bank notes with suspicion (Hyde, 1969). However the responsibilities of the bank were increased after a year in existence. In April 1717, the notes were to be accepted in payment of taxes. Obviously this widened the banks sphere of activity and with it gave the bank a greater credibility. However, sceptics still existed. Law was determined to prove them wrong. His ambitious nature thrived for success: "My bank is not the only one of my ideas not the greatest. I will create something which will surprise Europe by the changes it will bring about in favour of France – changes more profound than have resulted from the discovery of the Indies or the introduction of credit" (Hyde, 1969).

Law's grand scheme was not nearly established it seemed. Antoine Crozat, a wealthy financier, had obtained from the King in 1712 an exclusive monopoly trade for fifteen years in the present day states of Louisiana, Mississippi, Arkansas, Missouri, Illinois, Iowa, Wisconsin and Minnesota. It also extended to the French establishments in Canada (Hyde, 1969). This monopoly was conditional on the payment of taxes to the royal treasury and that the new colony was peopled with French emigrants (Ibid). Disputes with the French governor concerning taxes resulted in Crozat surrendering control of the region. Law was not to miss this opportunity. He created the Company of the West to which the French government gave the control of trade between France and its Louisiana and Canadian colonies, as had been given to Crozat (Moen, 2003). This was only another addition to Law's final objective. In May of 1718, gold and silver coinage was debased by 40% (Hamilton, 1936). Law's scheme was to profit from this event. By December, the Duke of Orleans converted the General Bank to the Royal Bank, with Law as the director once more. The difference this made was that note issues were sanctioned by the Council of the State and specifically guaranteed by the crown (Hamilton, 1936). From this Law's empire thrived. There were two remaining French trading corporations of substance, the East India Company and the China Company. He succeeded in acquiring these in May 1719. The conglomerate was renamed the *Compagnie des Indes*, but it was still commonly known as the Mississippi Company (Moen, 2003). It monopolised all French trade outside Europe (Garber, 2000). Law's achievements were mounting and the more he achieved, the more ambitious he became. In September 1719, the *Compagnie* bought the right to collect all French indirect taxes and subsequently direct taxes also, for a payment to the Government of 52 million livres per year (Garber, 2000). Also they purchased the right to mint new coins for France. By 1720, he had assembled and fused together all of the French trading companies, the tax firms, the tobacco farm, the mint, the French national debt and a quasi-Central Bank under a giant holding conglomerate popularly known as the Mississippi Company (Murphy, 1997). To finance the company shares were issued and paid for in government debt (Smant, 2003). Shares of 500 livres each, to be subscribed in state bonds-billets etat (Hyde, 1969). Law's proposal to exploit the apparently limitless resources of the region caused a tremendous wave of interest, not just in France, but also throughout Europe.

Credibility in the system had been firmly established as in January 1720 Law was announced as the Controller General of Finance with the responsibility of controlling all of France's financial and monetary affairs. However, Law had made the mistake that was to lead him to the nadir of his career. Law's fatal mistake was undertaking to refund the huge national debt. To finance this debt acquisition Law sold stocks in the company. Initially the shares (meres) were priced at 500 livres. The shares were sold at a premium of 10% and were made payable in twenty

monthly instalments of 5% each. Basically, the price was 550 livres per share and by paying 75 livres in cash the investor was allotted a single share (Ibid). The price of the shares rose continuously. A second issuing of shares (*filles*) were issued at a price of 550 livres. Preceding this a third issuing of shares (*les petites filles*) were issued at 1,000 livres. In order to obtain shares in the Company of the Indies it was necessary to possess original shares in the Company of the West. This resulted in a run on the old shares, which sent up their market value (Ibid). Finally, the soumissions were issued for 5,000. The drastic increase in the price encouraged huge speculation and at the zenith of the company's success the shares reached an amazing 10,000 livres. It must be noted that it was not only the wealthy that speculated in Law's system. As the price of shares rose, individuals from all walks of life invested what they had in the hope of benefiting from such an extraordinary event. The word millionaire can be attributed to Law as he created them in this period in history. The rich and poor alike were converted to millionaires within weeks. Fortunes were made overnight; coachmen, cooks and lackeys became millionaires (Hamilton, 1936). Unfortunately this newfound fortune was not to last for very long. Money circulation of specie and notes had increased by 186% with the issuing of all the shares and the resulting monetary expansion (Smant, 2003). However, because Law did not view shares as having the ability to increase the money supply, he neglected the inflationary effects it was to have on prices. Cantillon identified the flaws in the system as the permanent excessive increases in the money supply. He was aware of the unsustainable character of such increases. The printing presses were also under severe pressure to print out notes. The inflation added to the prevalent rage for speculation in the Mississippi stock, which had seized the whole country, unless checked, was bound, sooner or later to lead to disaster (Hyde, 1969).

The Collapse of the System

> *My shares which on Monday I bought*
> *Were worth millions on Tuesday, I thought.*
> *So on Wednesday I chose my abode;*
> *In my carriage on Thursday I rode;*
> *To the ball-room on Friday I went;*
> *To the workhouse next day I was sent.*
> (Hyde, 1969)

This passage from Hyde shows that the era of unbridled speculation and millionaires was to come to an abrupt end. Law needed to devaluate the shares to rectify the inflation problem, which had reached a monthly rate of 23% in January 1720 (Moen, 2003). The shares were to be reduced in price. After pegging shares at

9,000 on the 5th of March, Law proposed to deflate the shares further to 5,000 within a specified period. At the height of the system the shares were valued at 10,000, this policy was to have the effect of reducing their value by 50%. Shareholders were furious. This decree by Law in March was the last nail in his coffin and sealed the fate of the System. It led to a run on the banks and not surprisingly the specie was not available to support the notes issued. Law was dismissed and subsequently reinstated following the disorder it caused (Hamilton, 1936). The euphoric state turned to pure panic as the people rushed to the banks for the return of their coins. The intensification of his efforts to force the circulation of bank notes, the incessant changes in the tariffs of gold and silver coins, and other drastic measures adopted to draw gold and silver into the Bank further discredited the system (Ibid). Paper money became inconsequential and so coins were gradually reintroduced. Law fled France to Venice and died there nine years later in 1729. It was to take France more than one generation to recover fully from the effect of Law's system (Hyde, 1969).

Speculative Bubbles

Law's theory was highly advanced for his time. His application of it was flawed to say the least. The major problem with his monetary expansion was the over-issuing of shares and the fixing of their prices. This generated a period of unsustainable hyperinflation (Smant, 2003). Law's theory backdated to *Money and Trade*, which was that an expansion of the money supply could stimulate the economy, but his continuous expansion ultimately caused the demise of the system. This financial bubble as it is referred to in the world of finance refers to an unusually rapid increase in stock prices followed by an equally rapid collapse in prices (Moen, 2003). The driving force behind which is speculation. If the reason that the price is high today is only because investors believe that the selling price will be high tomorrow, when fundamental factors do not seem to justify such a price, then a bubble exists (Stiglitz, 1990). Inevitably, it will burst. In Law's system, individuals were not aware of the financial bubble that they were a part of and so they did not expect its sudden collapse.

Many economists are highly critical of Law and disregard his contributions to economics. His incompetence as an economic policy-maker is what tarnished his reputation. The system did not last it is true, but the theoretical frameworks in his writings provided the foundations for economics to come. Schumpeter is one of the few that appreciated the advanced thought, which Law possessed and placed him in the 'front ranks of monetary theorists of all times' (Emmett, 2000). Keynes' conversely makes no mention of him in the *General Theory*. While other economists remained critical of him, such as, Paris-Duverney, Turgot and Smith and Marshall, to name but a few.

Conclusion

Law possessed vision on economic matters comparable to no others. His ambition and motivation for success were the elements, which eventually led to his demise. However, awry his system went, Law cannot be held fully accountable. Yes, he did destroy millionaires, but he also created them in the first instance. This did not give him the right to subsequently destroy them of course, but it was not his intention to do this. His clarity of thought only extended so far. Ultimately flaws in his theory produced its collapse, which others were to learn from. If Law had not applied his theory to a working model, he would today have an elevated status as an economist. His previous works: *Essay on a Land Bank*, and *Money and Trade*, should have earned him this status regardless.

Bibliography

Emmett, R. (2000), *Great Bubbles.* London: Pickering and Chatto.

Galbraith, J.K (1975), *Money: Whence it came, where it went.* London: Deustch

Garber, P.M. (2000), *Famous First Bubbles.* London: MIT Press

Hamilton, Earl J. (1936), "Prices and Wages at Paris under John Law's System." *Quarterly Journal of Economics*. Vol. 51.

Hamilton, Earl J. (1969), "The Political Economy of France at the Time of John Law." *History of Political Economy*, Vol. 1(1)

Hyde, M. (1969), *John Law: The History of an Honest Adventurer.* London: W.H. Allen.

Law, J. (1705), "Money and Trade Considered With a Proposal for Supplying the Nation with Money" [Online]. *Archive for the History of Economic Thought.* Available from: <http://socserv2.socsci.mcmaster.ca/~econ/ugcm/3113/law/mon.txt>

Moen, J. (2003), "John Law and the Mississippi Bubble 1718-1720" [Online]. *Mississippi History Now.* Available From: <http://mshistory.k12.ms.us/features/feature22/law2>

Murphy, A. E. (1994), *John Law's 'Essay on a Land Bank.* Dublin: Aeon Publishing.

Muphy, A. E. (1997), *John Law: economic theorist and policy-maker.* Oxford: Clarendon.

Roll, E. (1961), *A History of Economic Thought.* London: Faber and Faber Ltd

Smant, D.J. (2003), "Famous First Bubbles?: Mississippi Company Bubble" [Online]. *Rotterdam School of Economics.* Available from: <http://www.few.eur.nl/few/people/smant/m-economics/johnlaw.htm>

Specialist Antique Map Magazine (2003), **"**John Law and the Mississippi Scheme" [Online]. *Specialist Antique Map Magazine.* Vol. 1 (5). Available from: <http://www.mapforum.com/05/mayind.htm>

Stiglitz, J.E. (1990), "Symposium on Bubble's." *Journal of Economic Perspectives*, Vol. 4 (2).

AN INVESTIGATION INTO WOMEN'S EMPLOYMENT IN BRITAIN DURING THE PERIOD OF INDUSTRIALISATION.

BY ROWENA GRAY

Senior Sophister

In this thorough research of women's employment patterns in 19th century Britain Rowena Gray examines historical factors explaining the fall in participation rates of married women in the labour market during the period of industrialisation.

The first question that arises immediately from the title of this project is: what is industrialisation and when did it occur in Britain? This is a much-debated issue in economic history but I take it to mean the gradual and often incomplete manner in which machinery and technical innovation was introduced into production in all areas of the economy.[1] I will deal primarily with the 19th century period (1780 to 1914), though admittedly it could be argued that the 'Industrial Revolution' began some time before this, at least in some sectors, such as textiles. Industrialisation is a historical phenomenon that is very difficult to pin down to one starting point, and so I feel justified in taking this arbitrary date. Industrialisation brought much social and economic change, which must have altered the outlook of the British people and presented challenges for the way in which society should be organised, that no country in the world had ever confronted before. The volume of contemporary inquiry into the problems of industrialisation is testament to this curiosity and to the attempts that were made to cope with the changing environment. I hope to answer the specific questions outlined above, as well as why it was "not unusual for women to change occupations several times during their lives or to perform several jobs simultaneously" (Honeyman, 2000). The focus here is on women, who appear to have been somewhat neglected in the existing literature, which is mainly concerned with how the Factory Acts and other measures affected the employment of children in industrial Britain (Nardinelli, 1980).

As this is a research paper, I have chosen a number of primary sources as evidence to support my arguments and to direct me to the most pertinent questions that need to be resolved about female employment in the nineteenth century. These

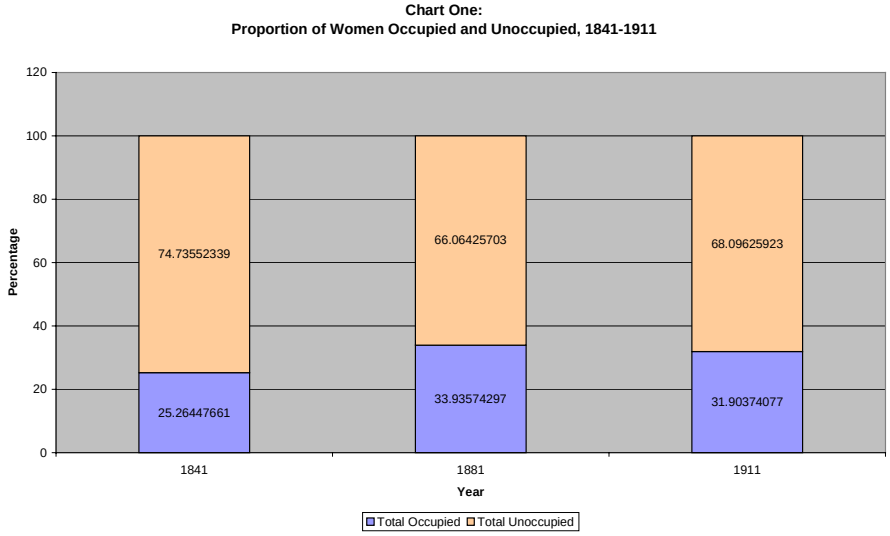
[1] Some have argued that the pace of industrialisation did not speed up until after the First World War. (Jones, 1971).

include selections from newspaper articles, pamphlets and diaries. J.M. Golby's (1986) collection of documents provided the impetus for this study, and led me to the survey of poverty among the working classes conducted by Henry Mayhew in 1850 (Thompson and Yeo, 1971). The use of the medium of newspaper meant that the survey was accessible to a wide range of people, and created much controversy. It made public and explicit what had been tacitly and privately acknowledged by so many - that the world's richest capital city was the scene of appalling poverty, which affected a large proportion of ordinary workers. It is not possible to determine conclusively what the consequences of this kind of survey may have been, but it was probably the middle-classes that reacted most strongly. It represents one of the first pieces of research into the area of homework, as previously only women's factory work had been investigated, even though an indeterminate but significant number of women worked in unregulated jobs, outside factories. Working-class women rarely leave diary evidence so Mayhew's interviews with them are even more valuable (Roberts, 1985). Patricia Hollis' (1979) collection lends further insight into what the impoverished working women themselves thought of their employment situation, the Factory Acts and their standing in society in general. It contains abstracts from meetings of the various governmental commissions formed to look into the problem of the double coincidence of poverty and industry in Britain, and in particular the plight of child and female labour at this time. The purpose of analysing these documents is to try to ascertain "the views and attitudes of the working-class women themselves" (Roberts, 1988), which is often overlooked both in modern studies and Victorian investigations such as are found in Engels' work. I will use the documentary literature throughout the essay, whilst realising that I may not have a fully representative sample of sources and that some of them may have embellished their claims in order to sell newspapers or convince the government or employers of the expediency of a particular policy.

I have made use of another primary source in constructing graphs from data on women's employment in the period 1841-1911,[2] to investigate the extent to which some occupations saw a rise in the numbers of females it employed, while others witnessed a notable contraction. One point of interest is that the proportion of all women who were engaged in paid work increased notably between 1841 and 1881, but that this level decreased slightly in the thirty years to 1911:

[2] All data derives from (Mitchell, 1988).

Figure 1: Proportion of Women Occupied and Unoccupied 1841 – 1911



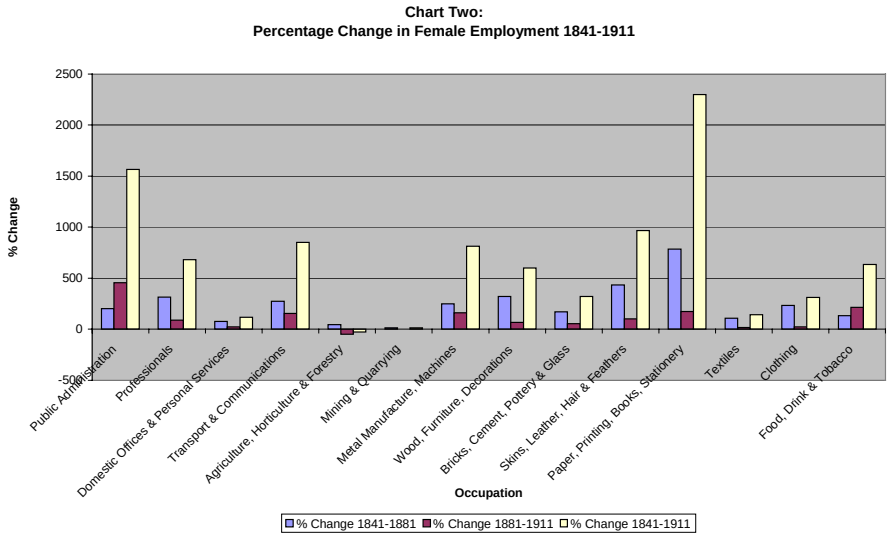
Source: Mitchell, 1988

The contraction in female employment in the latter period maybe due to inaccuracy in the data, as it may be that women switched out of officially recorded occupations into less formal employment, including working from home. One aim of this project is to identify broad trends in women's work in factories, smaller workshops, homework etc. Homework may be defined as the form of work whereby raw materials were collected from a central office and finished goods completed in the homes of the workforce (Pennington and Westover, 1989) it was referred to by some, including the Select Committee on Homework (1908), as sweating, due to the low pay and poor conditions in which the mainly female workforce toiled, which was rendered "quite insufficient to enable an adult person to obtain anything like proper food, clothing and house accommodation." However, due to intensified industrialisation, homework may have been disappearing from the 1870s.

The more rapid increase in the female population and growth in school attendance among girls after the 1870 Education Act (Ibid) may have caused the proportion of women employed to fall, despite the fact that absolute numbers were rising rapidly. Either way, over the 60 years from 1841 to 1911, an increase in the female population at work is clearly visible. The chart below illustrates the

percentage changes in women's employment across various occupations and draws attention to the main growth areas, administration and the professions, as well as printing and transport. In fact, another noteworthy growth area was in chemicals, but the percentage increase was so huge that it could not be included.

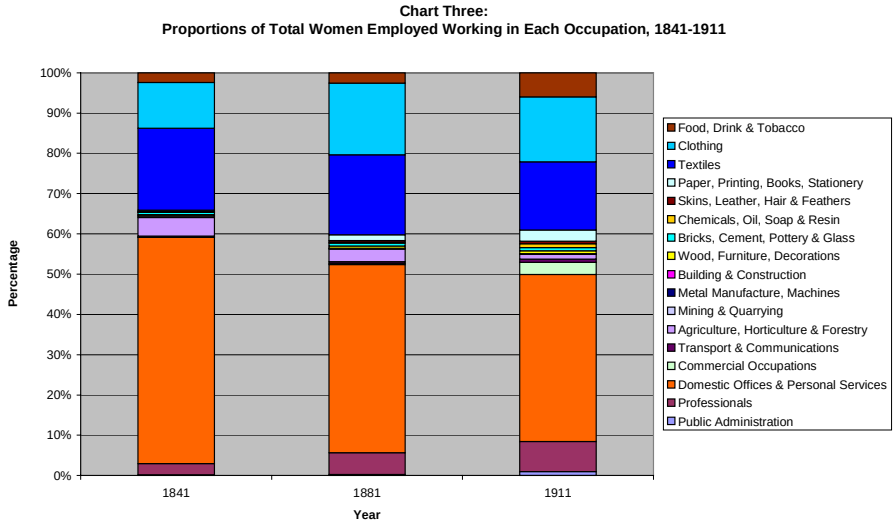
Figure 2: Percentage Change in Female Employment 1841 - 1911



Source: Mitchell, 1988

It is more useful perhaps to look at the change in the composition of occupations that women worked at over this same time period. Chart three does just that, revealing how domestic service, despite its increasing numbers employed, had less relative prominence in terms of proportions of women working in each occupation. Broadly speaking, clothing and textiles maintained their positions as significant sources of employment for women, while other sectors such as food and drink, the professions and printing saw considerable gains. Agriculture was the main area that declined as an employer of women over the period, as technology replaced female labour. This section has provided a summary of the main trends in female employment. The next will aim to explain how these trends came about.

Figure 3: Occupation breakdown of working women



Source: Mitchell, 1988.

Technology played a pivotal role in creating employment for women during industrialisation. It allowed employers to subdivide skilled work processes, and to set women, who were generally unskilled, to work on individual tasks, with the use of machinery. This is called 'dilution' and it facilitated mass production of the components of the final product. One example comes from the tailoring trade, which, once the sewing and buttonhole machines were invented, employed a large quantity of women on finishing tasks (Morris, 1986). It was condemned by Engels, who argued that men were forced out of work, or remained only as supervisors, as "feeble women or even a child" could complete the work with the help of the new machinery (Engels, 1993). The introduction of technology did not happen at one point in time, across all industries, as inventions took time to be perfected and adopted. This was especially important for the persistence of home working, which was mainly a female mode of production. Improved efficiency in one area of the business could often lead to an enlarged demand for labour in the more labour-intensive sectors, which makes sense, as some processes can never be mechanised. Factor endowments may also have influenced the extent to which home working persisted, as London, which lacked the space of the northern industrial centres, continued to use labour-intensive methods of production for longer into the nineteenth century (Jones, 1971). This may account for the acute poverty of the

capital. To counter the dilution process, unions attempted to establish a tradition of skilled work in many occupations (Lown, 1990). Even where machinery was specifically introduced to allow for female workers, such as in 1835 when Richard Roberts invented the self-acting mule for the textile trade, men successfully resisted this intrusion (Honeyman, 2000). Their exclusion of women and monopoly on skill yielded a hysteresis effect, whereby women lost the techniques and knowledge necessary for certain trades. With the onset of complete industrialisation and especially power-driven machinery, however, central production became the only profitable option and homework was phased out, with diminished wages (Honeyman, 2000). The position of female home workers became more tenuous[3] and led many to try to find alternative employment.

One of the major factors effecting women's employment and their occupation decision was the introduction of legislation, which labelled women as "protected persons",[4] in the Factory Acts. Some welcomed the legislation due to their genuine concern for women and children and the often-harsh conditions in which they worked. This group may have been predominantly middle-class, with little direct knowledge of the realities of work in factories or workshops around Britain. Others believed that legislation needed to be introduced to prevent women from working in jobs, which they considered to be unsuitable for them, as they took away feminine qualities. The main proponent of the Factory Acts, Lord Shaftesbury, told the House of Commons in 1846 that factory women disobey their husbands and meet together to drink, sing and smoke (Roberts, 1988). There is a difficulty here in distinguishing between people who genuinely lobbied for the Factory Acts for these reasons, and those who supported them with the hope that they would result in the decreased employment of women (leaving jobs free for men), particularly in skilled trades, and who used this excuse as propaganda for their cause. Official sources often tried to 'sell' the Factory Acts to the people by convincing them that they would necessitate a fall in labour supply that would in turn raise wages (Golby, 1986).

Some women took the contrary view to that outlined above, stressing the detrimental effects that such protective, but prohibitive legislation might have, in causing the unemployment and destitution of women, and in resulting in the erosion of their already meagre wages. Among these was Emma Paterson, head of the Women's Protective and Provident League until 1886. She convinced the authorities to establish an alternative method of guaranteeing proper working conditions for women by employing female factory inspectors (Roberts, 1998). Intuitively, this

[3] The 1908 Select Committee on Homework found the earnings of homeworkers to be very low, with workers having to work long hours just to make a living. (Morris, 1986).
[4] This classification was first brought in under the 1844 Act. (Roberts, 1988).

would have benefited female employees of factories more than the Factory Acts, as their welfare would be monitored directly and did not depend on the cooperation of employers with the legislation. The argument against the Factory Acts was that it placed an implicit tax on the employment of women. As two workers interviewed by the Royal Commission on Factories and Workshops said: "making it less profitable to employ women tends to drive them to unprotected occupations which are more laborious" (Hollis, 1979). It was not only the factory laws that had this effect. Many women had been forced into prostitution at least on a part-time basis, as Mayhew's interviews have graphically highlighted. One woman declared that "there is not one young girl that works at slop-work that is virtuous" due to the fact that they cannot make enough money at the more socially respectable but unpredictable endeavours of needlework (Thompson and Yeo, 1971). The 1864 Contagious Diseases Acts (Hollis, 1979) intervened in this trade, making it more difficult to do business. Many of the Classical economists of the period backed the regulation of child labour, arguing that children could be easily exploited and made to work long hours. Most followed the view of John Stuart Mill, who believed that "women were 'free agents' as much as men" (Blaug, 1958), and condemned the 1847 Ten Hours Act, which limited the working-day of women. For those who foresaw no augmentation in productivity arising from the reduction in working hours, such as Torrens, a fall in wages was predicted, given the narrow profit margins in most industries (Ibid.) These factors forced many women out of formal employment and into unregulated activity (Morris, 1986).

"A woman finds it difficult to earn her own living on account of the hostility of her male competitors" (Marcus, 1982). This outlines the importance of prejudicial male attitudes in determining women's access to paid work. Humphries calls this the social-control hypothesis, adding that it allowed all family members to work (as financial reasons necessitated) whilst maintaining the segregation of the sexes. Her index on the segregation of workplaces does appear to support the thesis that women were confined to certain jobs within industries, and that this worsened over time (Humphries, 1987). Women had always worked in the pre-industrial period, mainly in domestic industry, where production was based round the family unit. Moving into the 19th century, their role became much less clearly defined and women tried to move into all areas of employment. This raised new issues about the interaction between men and women. It was the working of women in public that was a particular problem for many in society, and the notion of a family wage was constructed to encourage women to stay at home, supported by their spouses (Pennington and Westover, 1989) making marriage the ultimate aim for women (Butler, 1979). The belief that the labour market had a fixed number of positions meant that a fear built up that cheap female labour that could replace men (Hunt, 1981). The unions are perhaps the best example of a body who played on the fear of

poverty in society (which had become more widespread with industrialisation), advising married women to stay at home and avoid stealing jobs from men, who were assumed to have families to support. Women were encouraged to enter only occupations that were perceived to require domestic and nurturing qualities, such as teaching and nursing (Lown, 1990), while domestic service was sometimes held up as the optimal job for women. There was a certain amount of hypocrisy in this as it was tacitly recognised that the ideal of the family wage did not exist, and so women and children were never fully prohibited from the workplace, only from higher wages thus only men were truly protected (Nardinelli, 1980). This was seen even in the work of Engels, who belonged to a movement that would later advocate women's liberation, as he recounts the story of a Lancashire man who was reduced to housework, while his wife worked in a local factory (Engels, 1993). This echoes the words of a deputation to Sir Robert Peel regarding the Ten Hours Act, when they described women's work as "an inversion of the order of nature and of Providence" (Hollis, 1979) Contemporary feminists railed against these notions, arguing that "this degradation of women is simply the accident of a new social order that has not yet righted itself" (Marcus, 1982) society would eventually remember that men and women had worked side by side, for equal pay, in the pre-industrial era. I believe that attitudes were important in affecting women's employment decisions; but that this has been over-stated by feminist historians, imposing their own paradigm on the nineteenth century.

In 1879, Emma Paterson described how there was a "great disinclination" among women for domestic service, condemning it as "at the best but a kind of slavery" (Golby, 1986). This may be contrasted with the perception of domestic service earlier in the century, which was summed up by Edward Higgs when he stated that domestic service was relatively well paid during this period, and was advocated by some as the optimal profession for young girls: "as an improving and educative process" (Higgs, 1986). The apparent diminution in popularity of domestic service, which was born out in the figures above, may have been due to changes in how working-class women perceived it. This may be explained by the fact that institutions were sending most of their young girls into domestic service, thus diminishing its respectability (Marcus, 1982). It may be concluded that only women with no alternative employment available, such as rural girls who migrated to the city, were attracted to domestic service in the later period. There are other explanations for the drop in the proportion of women being engaged as servants in this era (one reason given is that middle-class demand for servants had fallen off). This is unlikely, as real incomes were rising at the time, so much so that domestic

tasks had been contracted out to laundries, bakeries etc.[5] A more plausible explanation is that even the increase in wages was not sufficient to compensate women for the long hours and lack of freedom of a domestic servant. This example demonstrates the way in which perceptions changed over time and how this may partly explain why some occupations gained relative to others as employers of women, up to 1914.

Female activists of the day blamed the failure of women to organise effectively and bargain with employers for the erosion of wages. In part, this may have been due to apathy, as the power of the male unions was perceived as too great to be overcome. Some women even supported the notion of women staying out of the labour market and allowing men to take the available employment, as long as they could earn a family wage and thus keep the family from impoverishment. One instance of this occurred in 1853, when wives went out on strike with the Preston mule spinners, demanding a family wage for the men that would permit their wives to stay at home (Roberts, 1988). This undoubtedly had an adverse effect on women's wages as it sent a signal to employers that women did not deserve a full wage. *The Girl's Own Paper* spoke out against married women accepting low paid work for these reasons, as this pushed down the rates for single women trying to support themselves, but there was still no mobilisation of women behind this (Pennington and Westover, 1989). Beatrice Webb said that women did more damage to themselves than men, in the labour market, due to the lack of solidarity (Hollis, 1979). Perhaps though, this is simply a further sign that the working-classes did not wish to mobilise as a political entity, which was recognised by the *Pall Mall Gazette* during the depression years of the 1880s, when middle-class Socialists were attempting to rally the workers, in claiming that "they do not want revolution, they want a job" (Jones, 1971).

Female employment could not only have been determined by ideological and sociological factors. Supply and demand for labour and for goods undoubtedly played their part. The casual and seasonal trades in the Victorian era, with their fluctuations in demand, depended on "a abundance of the unskilled, 'degraded', or inefficient labour" (Ibid). Due to the barriers to entry to many trades, home working was often their only option and employers were eager to tap into a plentiful, immobile and un-unionised workforce (Pennington and Westover, 1989). The *laissez-faire* system that prevailed in nineteenth-century Britain became increasingly competitive over time, as other countries caught up with British development and began to challenge their firms for markets abroad. This created a constant downward pressure on wages, which was compounded by the nature of home working. It

[5] West mentions how "man has invaded woman's sphere" in setting up these service industries that are replacing the live-in domestic servant (Marcus, 1982).

involved information asymmetries in that the employer could not directly observe the employee at work, making it difficult to determine productivity and easier for workers to embezzle raw materials (Bythell, 1978). Unskilled female labour was particularly abundant because of institutional barriers to women acquiring skill, as unions controlled access to apprenticeship. The expectation that women would look after children and any elderly parents also prevented many from taking up full-time jobs outside the home. The supply of unskilled labour between ages 20 and 30 was pushed out by the tremendous amount of rural-urban immigration among these age groups (Jones, 1971). This compounded the unprecedented explosion in the population, which had produced a surplus of women.[6] The new Poor Law of 1834 ensured that all women had to work, unless they had an independent fortune or a husband. There was no subsidisation of meagre wages as there had been previously- only a place in an unpleasant workhouse for those on the extreme margins of society. In order to avoid the fate of the workhouse, men and women worked long hours for a low return, often in their own homes, in an effort to make enough to subsist that Mayhew frequently mentions this in his interviews (Thompson and Yeo, 1971). As Mayhew pointed out: "there are so many to work at it that one will under work the other" (Ibid) - with such a glut of labour, workers will be so desperate as to undercut each other.

The Courtauld silk factory provides the contradiction to the norm of employer hostility to female factory labour.[7] Technology had arrived early in the textile industry, as far back as 1720, and factories had been depending on the inexpensive labour of women and children since that time. Indeed, the 1851 census shows more women working in the silk industry than men (Lown, 1990) Courtauld wanted to keep employing women but were caught in the "ideological crossfire" with those who were offended by the notion of females working outside the home and in supposed unwomanly activities. They were firmly opposed to the Factory Acts, which made their female employees relatively more expensive, as women were their employees of choice replacing them with men would not improve efficiency. Thus, in some areas, women were the workers of choice in the formal factory setting.

These points have highlighted the negative aspects of the story of female employment up to 1914. There were some positive outcomes, as women were forced to search for new areas of work and some were aided by the widening educational opportunities after 1870. The emergence of the consumer society created a service sector that employed a growing number of women in the latter half of the 19th

[6] It is suggested that there is an excess of 405,00 women over twenty in Britain in 1851, by W.R. Greg, in 'Why are women redundant?' *National Review*, April 1862 (Hollis, 1979).
[7] This is the subject of Lown's (1990) investigation into women's employment.

century, as shop assistants and personal aides to the social elites. Admittedly, many of the more interesting job opportunities became available only for the relatively more educated middle-class, who were still only a small proportion of the total female population at work by 1911. The proportion of working women who were employed in the professions went up by almost 5% between 1841 and 1911, though they were engaged in a narrow range of activities (Mitchell, 1998). This increase reflects both women's increased desire to move into these new avenues of work (Zimmeck, 1986) and the need to find a means of subsistence for the surplus of middle class women. It is interesting to note that where women replaced men in these sectors, the pay and conditions associated with the job often became devalued as more women entered. Zimmeck (Ibid) emphasised this, explaining how individual clerical tasks became associated solely with either men or women, with women often consigned to the more repetitive, mechanical parts of the job, leaving the more creative aspects to men. This is a useful example of the crucial influence that ideology and social norms had in Victorian society.

Many myths have been dispelled in this project. Contrary to popular belief, women had been an integral part of the workforce before industrialisation and continued to play a significant role afterwards, despite the frictions in the market, which led to a movement among the male unions and even officials, to sideline their function. Official participation rates of married women fell dramatically from 1851 to 1911, by 15% in fact (Hunt, 1981). In many cases, continuing in employment could only be done in the homework sector, so that records of the level of female employment are in fact understated. As the 19th century came to a close, the plight of women in home working received much public attention and led to an improved organisation of the female labour force, together with governmental enquiries into the problem. That women were more impoverished than men was well documented by Charles Booth in the 1880s and later by Bowley and Rowntree (Ibid.) The fact that women's wages were consistently a half to two thirds less than those of men up to 1914 cannot have helped (Pennington and Westover, 1989). By the outbreak of the First World War, social reformers were calling on male and female workers to unite and use their joint bargaining power to effect changes in conditions and pay for all. During the war itself, a revolutionary change came over the labour market out of the sheer necessity of wartime conditions. Constrained supply of adult males led to an unprecedented rise in the formal employment of women, due to processes that had been resisted for almost a century, such as dilution.[8] In some areas, such as clerical work, women had asserted their dominance before the cessation of war, while in most manufacturing jobs, women were ousted, as one worker has described, "On the eleventh hour of the eleventh day of the eleventh month the Armistice was

[8] For a full explanation see Horne,

announced. We were all stopped, just like that, no redundancy, nothing" (Griffiths, 1991). While the position of working women during the nineteenth century can be contrasted to that which prevailed during the 1914-1918 War, there is obviously much continuity. It would take another 40 years of the reform of attitudes and another world war, before married women were actively welcomed back into the workforce.

Bibliography

Blaug, M. (1958), "The Classical Economists and the Factory Acts A Re-Examination." *The Quarterly Journal of Economics* 72 (2).

Bythell, D. (1978), *The Sweated Trades: Outwork in Nineteenth Century Britain*. London: Batford.

Engels, F. (1993/1845), *The Condition of the Working Class in England*. Oxford: Oxford University Press.

Golby, J.M. (ed) (1986), *Culture and Society in Britain: 1850- 189: A Source Book of Contemporary Writing*. Oxford: Oxford University Press.

Griffith, G. (1991), *Women's Factory Work in World War I*. Sutton: Stroud.

Higgs, E. (1986), "Domestic Service and Household Production." In John (ed.) *Unequal Opportunities: Women's Employment in England 1800-1918*. Oxford: Basil Blackwell.

Hobsbawm, E.J. (1967/1960), "Custom, Wages and Work-Load in Nineteenth Century Industry." In Briggs, and Saville (eds) *Essays in Labour History*. London: Macmillan.

Hollis, P. (1979), *Women in Public: The Women's Movement 1850-1900*. London: Allen and Unwin.

Honeyman, K. (2000), *Women, Gender and Industrialisation in England, 1700-1870*. Basingstoke: Macmillan.

Humphries, J. (1987), "…The Most Free from Objection…The Sexual Division of Labor and Women's Work in Nineteenth-Century England." *The Journal of Economic History*. Vol. 47(4).

Hunt, E.H. (1981), *British Labour History 1815-1914*. London: Weidenfeld and Nicolson.

Jones, G. (1971), *Outcast London. A Study in the Relationship Between Classes in Victorian Society*. Oxford: Clarendon.

Lown, J. (1990), *Women and Industrialization: Gender at Work in Nineteenth-Century England*. Cambridge: Polity.

Marcus, J. (ed) (1982), *The Young Rebecca: Writings of Rebecca West 1911- 1917*. New York: Macmillan.

Mitchell, B.R. (1988), *British Historical Statistics*. Cambridge: Cambridge University Press.

Morris, J. (1986), "The Characteristics of Sweating: The Late Nineteenth-Century London and Leeds Tailoring Trade." In John (ed.), *Unequal Opportunities: Women's Employment in England 1800-1918*. Oxford: Basil Blackwell.

Nardelli, C. (1980), "Child Labor and the Factory Acts." *The Journal of Economic History.* Vol. 40(4).

Pennington, S. and Westover, B. (1989), *A Hidden Workforce: Homeworkers in England 1850-1985*. Basingstoke: Macmillan.

Pinchbeck, I. (1969/1930), *Women Workers and the Industrial Revolution 1750-1850* . London: Virago.

Roberts, E. (1988), *Women's Work, 1840-1940*. Cambridge: Cambridge Univeristy Press.

Thompson, E.P. and Yeo, E. (eds) (1971), *The Unknown Mayhew: Selections from the Morning Chronicle 1849-1850.* Harmondsworth: Penguin.

Young, G.M. and W.D. Handcock (eds) (1956), *English Historical Documents XII 1833-1874.* London: Eyre and Spottiswoode.

Zimmeck, M. (1986), "Jobs for the Girls: The Expansion of Clerical Work for Women 1850-1914." In: John (ed.) *Unequal Opportunities: Women's Employment in England 1800-1918* . Oxford: Basil Blackwell.

TRADE AND DIPLOMACY: THE ENGLISH MERCHANTS IN SIXTEENTH CENTURY RUSSIA

BY ALEXANDER GOROKHOVSKY

Senior Sophister

It is widely believed that it was Peter I (Peter the Great) who first 'cut a window into Europe' for Russia. However, as this account by Alexander Gorokhovsky demonstrates, Russia has been engaged in active international trade, particularly with England, at a much earlier period. Through the eyes of 'Marchant Adventurers of England' Alexander lets us see how the economics of 16th century differed from the discipline we study and practice today.

Introduction

This essay explores the nature of trade between England and Russia in the 16th century. While the main actor on the English side is believed to be the emerging class of capitalist merchants represented by the Muscovy Company, Russia's position in this relationship was determined solely by the Tsar and his political interests. I examine how the principal differences in the aims of the two sides eventually led to the annulment of the English monopoly on tax-free trade with Russia, arguing that one cannot observe any clear distinction between 16th century economics and politics.

The Setting

A period of extravagance that followed the Reformation left England with a large debt (£148,526 at 14% in 1555) (Scott, 1912). Since most of the loans were due abroad, the payment of the interest constituted a serious drain on the country's commerce. This resulted in greater attention to the development of foreign trade as a source of revenue for the English Crown.[1]

[1] An economic explanation for the increased interest of the English merchants in overseas trade in the third quarter of the 16th century given by Scott is that by 1550 the capital of the

By the middle of the 16th century, semi-finished woollen cloth made up 90% of English exports. After 1550, this trade started to contract: the London export of short-cloths, for example, fell from 132,767 in 1550 to 84,969 in 1552 (Fisher, 1940). These fluctuations marked the beginning of a period in which cloth exports were to be some 25% lower than they had been during the previous half-century (Ibid).

Fear of unemployment in the country's largest industry meant that the availability of the export market for cloth was a prevailing issue. At that time, English commerce was overwhelmingly dependent upon the Netherlands. The Dutch were England's major client for cloth exports, and Antwerp was the main commercial centre for both imports and exports. The Netherlands, however, were then a Spanish province and this meant that the stability of the European market for English cloth, and the availability of imports in return, were affected by the state of relations between Spain, England and the Netherlands. The dependence on a single product and a sole client had to be addressed.

The demand for other English goods was falling as well. A contemporary, Clement Adams, wrote that:

> "at that time, our Marchants perceived the commodities and wares of England to be in small request with the Countreis and people about us, and neere unto us, and that those Marchandizes which strangers in the time and memorie of our auncestors did earnestly seeke and desire, were now neglected, and the price thereof abated, although by us carried to their owne portes, and all forreine marchandizes in great accompt, and their prises wonderfully raised'. Because of this, 'certaine grave citizens of London… thereupon resolved upon a newe and strange navigation" (Hakluyt, 1589/1965).

The Voyage

> " 'Certaine grave citizens of London', who were said to number 240, formed a venture under the name of *The Mysterie and Compagnie of the Merchant Adventurers for the discoverie of regions, dominions, islands and places unknown* to finance an expedition to discover the north-east passage to the promised land, Cathay (the ancient name of China). The Spanish and Portuguese monopolies based on discovery and papal grant did not apply in the north; therefore, as Robert Thorne had pointed out in 1527 in an address intended for Henry VIII, this was the only area where the English were free to go without fear of political complications" (Hakluyt, 1589/1965).

---

country was depleted, with national production being less than it had been in the first half of the century. Under such circumstances, attempts would be made to secure a higher return on capital. Thus, the more economically rational agents would be more inclined to become involved in potentially highly profitable foreign trade (Ibid).

The journey was expected to be long and hazardous, and to provide funds for such a voyage would require considerable capital. This pointed to the "joint-stock company as the essential device for the conduct of … trade" (Willan, 1956). Such form of organisation allowed the promoters of the enterprise to spread the risks of an uncertain undertaking such as this difficult and possibly dangerous trip along previously unexplored route.

The promoters raised a capital of £6,000 in shares of £25 each from "every man willing to be of the societie" (Hakluyt, 1589/1965), devoting this sum to the purchase of some goods and three ships, the *Bona Confidentia* of 90 tons, the *Bona Speranza* of 120 tons, and the *Edward Bonaventure* of 160 tons, which were dispatched to the northern seas under the command of Sir Hugh Willoughby. Early in August 1553, the *Bona Speranza* and the *Bona Confidentia* lost touch with the *Edward Bonaventure*, and sailed on eastwards until they reached Lapland, where they wintered. During the winter Willoughby, who had sailed in the *Bona Speranza*, and everyone else on board the two ships "for lacke of knowledge have frozen to deathe". The *Edward Bonaveture*, with Richard Chancellor as captain, was more fortunate, reaching the White Sea and anchoring at the mouth of the Northern Dvina River. The region had just recently been added to Russia, and when its ruler, Ivan IV (Ivan the Terrible), heard of Chancellor's arrival, he invited him to come over to Moscow.

Chancellor arrived at a crucial point in the history of the Russian state, which was emerging on the eastern frontiers of Europe as a mighty and centralised Christian power under the rule of the grand duke of Moscow, who had just assumed the title of the Tsar (Caesar). The Russians considered it necessary to establish communications with Western Europe and to this end were making tremendous efforts to get control of the Baltic shore.

Chancellor made the voyage of over 1,000 kilometres through a country covered in snow and ice, and arrived at Moscow. He found Moscow both large (much larger than London) and of a primitive build, most houses being wooden. However, the palace of the Tsar was very luxurious, and Ivan's appearance when Chancellor was finally summoned to see him (the Tsar received the English guests seated on his throne with a gold crown on his head, wearing a gold cloak and holding a costly sceptre) certainly suited his high position. The ceremonial dinner was also quite impressive, with "all the furniture of dishes, and drinking vessels, which were then for the use of a hundred ghests, was all of pure gold, and the tables were so laden with vessels of gold, that there was no roome for some to stand upon them" (Hakluyt, 1589/1965).

The Establishment of the Muscovy Company

Ivan expressed his willingness to authorise the free passage of English ships to Russia "with good assurance on our part to see them harmlesse" and to allow "free marte with all free liberties" to English merchants throughout his dominions (Hakluyt, 1589/1965).

It was to the advantage of Russia to open a maritime trade as it did not yet have a port on the Baltic, while the Hanseatic League had a monopoly on the trade between Muscovy and Central and Western Europe.[2] The opening up of the northern route gave the Russians an opportunity for contact with Europe, without interference by those Baltic States with whom they were frequently at war. The English were not less optimistic, having found a vast market for the English wool, and receiving furs, hides, flax, hemp, train oil, fats, tallow, cordage, timber, wax, grain, and other goods in return (Riha, 1969).

On his return to England, Chancellor assured the shareholders that there were good prospects of a profitable trade with Russia. In order to secure the benefits of the newly opened trade to the discoverers of it, a charter was signed on 6th of February 1555, giving the company with the tiresome title of *Marchants Adventurers of England, for the discovery of lands, territories, isles, dominions and seignories unknowne, and not before that late adventure or enterprise by sea or navigation, commonly frequented*[3] exclusive rights to trade with Russia, or with any other countries that might be opened up by the adventurers in the future.

The promptitude, with which these privileges were granted, clearly demonstrates the importance given to the branch of trade now made available. The Political situation during this period required rapid arming of the country. The Muscovy Company had the advantage for equipping the navy, while the munitions obtained in Flanders had to be smuggled into England, naval requisites bought from Russia were to be brought in with the goodwill of the Tsar. Therefore, not only would a new market be found for English commodities with the opening of the Russian trade, but England would also obtain direct access to materials of the greatest importance to her navy, namely cordage and timber (Scott, 1912).

---

[2] The Hanseatic League was an association of German towns, which from major trading posts in Bruges, London, Bergen, and Novgorod, developed trade in much of Northern Europe and acted as an intermediary between east and west in this region, as the Venetians did in the Mediterranean. The Hansa co-ordinated and distributed capital, goods and skills throughout its domain and was engaged in many activities such as sheep rearing in England, iron production in Sweden, and agriculture in Poland.

[3] This title was very rarely used, for the company at once became known conventionally as the Russia Company, or the Muscovy Company.

Privileges Granted to the Company by the Tsar

Chancellor reappeared in Moscow as an ambassador in 1555, this time accompanied by the two agents of the newly set up Company, Richard Grey and George Killingsworth. The Company's agents were received in audience by the Tsar, who seems to have been most impressed by Killingsworth's beard "in length five foot two inches of assize" (Hakluyt, 1589/1965).

Negotiations ended with the English receiving the following privileges and concessions. The members, agents, and servants of the Company were granted the right to trade anywhere in Russia without paying any duties or tolls. Neither the merchants nor goods could be detained "for anie debt, duetie or other thing, for which they be not principall debters or sureties' or 'for any offence or trespasse committed … but only for such as they or any of them shall actually commit" (Hakluyt, 1589/1965) and such cases would be judged by the Tsar himself. The Company could hire directly workers of any sort. Swift settlements were promised in cases between the English and the Russians. Similarly, if any Englishman were wounded or killed, every effort was to be made to conduct a thorough investigation and immediately punish the offender as an example for others. None of the English merchants was to be imprisoned for debt. Also, if any Englishman disobeyed the Company's chief factor in Russia, the Tsar promised to give his full assistance in bringing such offender to justice, and lend prisons and instruments of punishment for that purpose when necessary.[4]

Chancellor sailed back to England in July 1556, this time accompanied by the Russian envoy, Osip Nepeia. On 10th of November 1556, their ship was caught in a storm and shipwrecked just off the Scottish coast. Chancellor drowned, but Nepeia survived. Although, he was taken hostage by the Scots for a few months, the first Russian ambassador to England was eventually rescued and after much trouble reached London, where he was received by the King and Queen on 25th of March 1557.

Philip and Mary, in gratitude for the privileges granted to the English by the Muscovite state, granted Russian merchants the right to trade both wholesale and retail in all parts of their realm, promised to place them and their property under special protection, and to assign them suitable warehouses in London and any other English cities. However, Russian merchants were not to be exempt from the

[4] Among the privileges granted to the Muscovy Company by Ivan IV was also the right to mint coins free of tax at Russian mints. However, as noted above, during Elizabeth's reign England was constantly short of money and the export of silver and gold was strictly banned, so in order to support their trading transactions English merchants were forced to acquire thalers from the continent (Zverev, 2003).

payment of customs duties, for they were to "be as free from paying of any customs or taxes here as the subjects of other Christian princes traffiquing within our said realme be" (Willan, 1956). These concessions were not very fair, given the Muscovy Company's exemption from customs duties in Russia, and they were in fact of no use since the Russians had no ships capable of navigation to England.

Trade

Nepeia returned to Russia with a new representative of the Muscovy Company - Anthony Jenkinson, who was an experienced trader and geographer, "a well educated and observant man of great abilities for commercial and diplomatic negotiations" (Vernadsky, 1969). The Tsar liked him and allowed him to proceed down the Volga River to Astrakhan, on the Caspian Sea, and from there to Bukhara.

Jenkinson succeeded in reaching Bukhara in 1558. From there, he planned to travel to Cathay by land, but constant wars in the region and robbers made the passage to China impossible, compelling Jenkinson to give up his original intention. A series of expeditions to Persia followed, though, with the Persian trade yielding large profits to the Company up until 1581 when it was given up (Scott, 1912).

There are no figures available as to the results of the trade with Russia itself at that period, but judging by the enthusiasm with which English traders who were not members of the Company took on the Russian trade after Ivan's army captured Narva - the Baltic port which could serve as a substitute for the White Sea route - it must have been very profitable. Since Narva had not been part of Russia when the Muscovy Company received its charter, English and other 'interlopers' rushed to exploit the new route which was not only shorter, but much safer than the one discovered in 1553, claiming that it did not come within the scope of the Company's monopoly.

The Muscovy Company brought the question of the English interlopers at Narva to the notice of the Parliament in 1566 and managed to obtain an act confirming the privileges granted to the Company by the royal charter of 1555. The act stated specifically that the right of the Muscovy Company to trade exclusively with Russia extended to all territories subject to the Russian sovereign, thereby including Narva in the sphere of its monopoly (Willan, 1956).

The act of 1566 could have impeded the unauthorised trade from England, but it did not stop merchants from other countries to take advantage of the new route to Russia. The Dutch, in particular, began to find their way into the country, and by 1569 the trade of the Company is generally believed to be less profitable than it had been prior to that. Partly that was due to the competition from the interlopers. But an even greater threat to the Company's business in Russia came from the Tsar, who

himself was encouraging the development of the Narva trade and by that time started to show his disillusionment with the English.

Diplomacy

Apart from securing armaments from England, Ivan hoped also to get skilled workers and professionals of all kinds, especially those who could teach the Russians skills useful in war. In 1547, a German adventurer, Hans Schlitte, offered Ivan his assistance in engaging German specialists of various sorts for service to the Tsar. He eventually hired 123 technicians and brought them to Lübeck, from where they were to proceed to Moscow. This became known to the Hanseatic League, which demanded from the Lübeck authorities that the technicians be forbidden to go to Russia. The reason for this was the fear that the spread of technical knowledge in Russia, as well as the import of military supplies, would strengthen the country both economically and militarily to such an extent that it will threaten the Hansa's business in Eastern Europe. Schlitte was arrested, and the men he had hired dispersed (Vernadsky, 1969).

It is interesting to note that, when Chancellor visited Moscow, he heard from the Livonians and Poles he met there of their fears of the military danger a 'civilized' Russia might present to her neighbours. He also wrote of Ivan's army:

> "Now what might be made of these men if they were trained and broken to order and knowledge of civill wars? If this Prince [the Tsar] had within his countreys such men as could make them to understand ye things aforesaid, I do believe that 2 of the best or greatest princes in Christendome were not able to match with him, considering the greatnes of his power and the hardnes of his people" (Hakluyt, 1589/1965)

Similarly, the grand-duke of Poland-Lithuania Sigismund Augustus in an attempt to stop the trade with Narva wrote to Queen Elizabeth:

> "The Muscovite sovereign daily grows stronger by acquiring the objects that are brought to Narva, not only goods but arms the like of which he never saw before. Not only are crafted items brought there, but craftsmen themselves are also coming there. By these devices he is acquiring the means to defeat everyone... Hitherto we have been able to defeat him only because he was ignorant and did not know the arts. If navigation to Narva is continued, what will remain unknown to him?" (Soloviev, 1995).

The English government paid little attention to the apprehension of Russia's neighbours and continued commercial relations with Moscow. However, the English Crown was concerned only to win more trading advantages for its merchants, while Ivan IV was interested in more than just trade or even the export of skilled labour to Russia. He saw England as his only potential ally in Europe. Queen

Elizabeth, on the other hand, always emphasised that the purpose of her relations with Moscow was solely commercial. It was impossible for her to enter into a political alliance sought by Ivan, for she had no intention of incurring hostility of the Tsar's enemies, especially Poland and Sweden. The obvious reluctance of the English side to become involved in the Baltic conflict led to disappointment and increasing resentment on the part of the Russian sovereign.

The following excerpt from Ivan's letter to Elizabeth best of all reflects a basic difference in their aims:

> "We had thought that you had been ruler of over your lande and had sought honor to your self and profitt to your country, and therefore wee did pretend those weightie affaires betweene you and us; but now we perceive that there be other men that doe rule, and not men but bowers and merchaunts the which seeke not the wealth and honour of our maiesties, but they seeke their owne profitt of marchauntdize: and you flowe in your maydenlie estate like a (common) maide" (Tolstoy, 1875).[5]

Ivan withheld all the privileges previously granted to the Muscovy Company, imprisoned the English merchants in Moscow, and seized all their goods. Although, he restored their franchise in 1572, a 50% duty was now imposed. Moreover, the Tsar persisted in encouraging the trade of Narva, so long as he was able to retain this Baltic port. When Narva was lost to Sweden in 1581, Ivan decided to encourage the trade of England's competitors (especially the Dutch) in the Russian North and ordered the transfer of all foreign trading establishments to the newly built port of Archangelsk (Clarkson, 1962). Finally, a new charter to the English withheld even from them the right – never accorded to their competitors - to engage in retail trade. Foreigners were forbidden to operate at specified points, and the commodities they were most interested in were declared state monopolies.

When Politics Interfere with Economics

Many of the decisions of moody and unpredictable Russian monarch did not have any rational motives.[6] The reasons behind this very one, however, may well be traced.

---

[5] Elizabeth made an effort to placate Ivan and sent her special envoy, Daniel Silvester, to Russia. The latter never reached the Tsar, though, for in Kholmogory, while Silvester was trying on "a newe yeolow satten jackett…, a thunderbolt came and stroeke him dead" (Bond, 1856).

[6] The widely accepted view among the historians of the period is that Ivan IV suffered from schizophrenia, which "projected itself in his psychology in the way of paranoia and mania of persecution, also associating itself with a substantial degree of sadism" (Chirovsky, 1973).

Ivan IV granted the extensive privileges to the Company in the hope that favourable treatment would produce a political alliance with England. When it became clear that his demands were not going to be satisfied, it was quite natural for the Tsar to reconsider the Company's privileged position in Russia, especially when negotiations with Sweden made the need for a military alliance with England less important.

All of the above indicates that there was no clear distinction between economics and politics in international relations at that time, particularly in the case of Anglo-Russian trade. This, in turn, had its roots in the nature of the Russian state system of that period. Although, the foundations of absolute monarchy in Russia were laid almost simultaneously with major European countries, it rested on a totally different socio-economic basis. In Western Europe, the rising power of the centralised monarchies had been rooted in the rise of trade, and the growth of state power was accompanied by the growing wealth and influence of the bourgeoisie (Clarkson, 1962). This economically independent new class was able to set limits to the power of even the most absolute monarchs. In Russia, on the other hand, there was virtually no bourgeoisie - the Tsar himself often performed the functions of an upper merchant class.[7]

The contacts with the English merchants might benefit the Tsar in his own commercial operations. However, as stated above, Ivan in his relations with England had in mind something besides trade. He saw England as a possible source of supply for arms and expected Elizabeth to form an offensive-defensive alliance with him. Thus, the Company 'could not expect a privileged position without giving, or

[7] As noted by John Hasse, Ivan IV was "a great marchant himselfe" (Hakluyt, 1589/1965). This arose partly from the Russian taxation system, which involved the collection of some of the state revenues in kind. Thus, the expansion of fur trade into Siberia, which led to exploration and colonisation as it did later in Canada, was accompanied by the collection of a fur tribute from the natives and the Russian traders. This made the state the biggest fur trader of all and enabled the Tsar to barter furs for western European goods (Willan, 1956). A number of other branches of trade were monopolised for the benefit of his treasury. The Tsar's monopoly in the middle of the 16th century included grain, hemp, rhubarb, raw silk, potash, tar, and caviar. Occasionally it was extended to include other goods as well. In addition to these 'forbidden' goods, foreign goods in general coming into Moscow were brought before the Tsar himself soon after their customs inspection and appraisal, he then selected what he liked, and only the remainder was released for sale. Taking these goods, in the words of Giles Fletcher, from the foreign merchants at a low price, the Tsar later sold them at higher prices (Fletcher, 1591/1966). Even goods in which anyone was permitted to trade (fur skins, honey, wax, lard, and others) were frequently bought for royal commerce at arbitrarily fixed and low prices, and then resold at higher prices in the internal market and abroad. Furthermore, merchants were sometimes prohibited from selling their goods until the stocks of these goods held at the Tsar's warehouses were sold out (Lyashchenko, 1970).

pretending to give, something in return … and could not reasonably complain if the Tsar expected his privileges to be paid for in a political coin' (Willan, 1956). Given the absence of institutional arrangements capable of limiting the Tsar's power, it became possible for the Russian monarch to operate against the free development of commercial relations with England, when it became clear that his interests are at odds with those of the Muscovy Company.

Conclusions

By the sixteenth century, the rising economic power of the bourgeoisie in Western Europe, although not yet expressed in acquisition of political power, had enabled the merchant class to act on the international arena with certain support from the Crown. However, a company engaged in foreign trade needed not only a charter at home, but also privileges from the government of the country with which it traded. Such privileges were difficult to acquire and maintain, especially in a country where trading activities depended solely on the whims of a despotic ruler.

Trade might and did need diplomacy for its support, but such diplomacy could only work when both sides were able to deliver what was sought by their counterpart. When this was not the case, there seemed to be little or no room for negotiation. The nature of imports and exports in Anglo-Russian trade suggests that it was more valuable to England than to Russia, thus enabling the Tsar to impede, rather than promote, the relationships with the English when it became obvious that the latter were reluctant to get involved into anything more than just a commercial partnership.

It can still be assumed, though, that economic and cultural contacts with England during the reign of Ivan the Terrible played an important role in Russia's major turn towards Western Europe under Peter the Great some 150 years later, and contributed significantly to the fact that Russian society was better prepared for Peter's reforms than is sometimes thought.

Acknowledgements

The author would like to thank Ann Carlos for valuable suggestions during the preliminary stage of this research paper.

Bibliography

Bond, E.A. (ed.) (1856), *Russia at the Close of the Sixteenth Century*. London: Hakluyt Society.

Chirovsky, L. (1973), *A History of the Russian Empire.* Vol. I, London: Peter Owen.

Clarkson, J. (1962), *A History of Russia from the Ninth Century.* Edinburgh: Longmans.

Fisher, M. (1940), "Commercial Trends and Policy in Sixteenth-century England" in *Economic History Review*. Vol. 10 (2).

Fletcher, G. (1591/1966), *Of the Russe Commonwealth: Facsimile edition with variants with an introduction by Richard Pipes*. Cambridge, Mass: Harvard University Press.

Hakluyt, R. (1589/1965), *The Principall Navigations, Voiages and Discoveries of the English Nation. Vol. I*. Cambridge: Cambridge University Press.

Lyashchenko, P. I. (1970), *History of the National Economy of Russia to the 1917 Revolution.* New York: Octagon Books.

Riha, T. (1969), *Readings in Russian Civilization. Vol. I,* 2nd ed. Chicago and London: The University of Chicago Press.

Scott, W. R. (1912), *The Constitution and Finance of English, Scottish and Irish Joint-Stock Companies to 1720. Vol. I*. Cambridge: Cambridge University Press.

Soloviev, S. (1995), *History of Russia from Earliest Times. Vol. X*. Gulf Breeze: Academic International Press.

Tolstoy, I. (1875), *Pervye sorok letsnoshenii mezhdu Rossiei I Angliei, 1553-1593*. St. Petersburg.

Vernadsky, G. (1969), *A History of Russia. Vol. V, Part I*. New Haven and London: Yale University Press.

Willan, T. S. (1956), *The Early History of the Russia Company.* Manchester: Manchester University Press.

Zverev, S.V. (2003), "The English Muscovy Company and the Supply of Silver Coin to Russia" in *Russia-Britain*, Moscow: Izd. dom Maksima Svetlanova.

BEING ECONOMICAL WITH THE TRUTH: ECONOMIC REDUCTIONISM FROM THE MACHINE AGE

BY STEVE DALEY

Junior Sophister

The discussion about the scientific status of economics is an argument without an end in sight spanning the pages of many academic journals, and the SER is no exception. In this issue, Steve Daley revisits the problem. Drawing on influential authors in the field of both economics and philosophy, he criticises existing methodology and suggests a new way forward for economic science.

"If your experiment needs statistics, you ought to have done a better experiment."

Ernest Rutherford (Bailey, 1967).

The prize of scientific prestige has animated a stream of allegation and counter allegation between economic theorists and econometricians. In spite of over half a century of practical work and theorising, John Maynard Keynes' deep mistrust of Jan Tinbergen's early work in econometrics still provokes nods of approval. David Hendry (1980) has repeated Keynes' likening of econometrics to black magic and alchemy, picking up exactly where Keynes left off a generation ago. It is easy to forget the defensiveness of economic theorists in the midst of social instability in the inter-war period: the Bolshevik Revolution propelled the Marxist critique of capitalism as the synthesis of theory and practice, so elusive to bourgeois economics; while the economic crisis of the Great Depression had forced economists to fight a rearguard defence of economic theory. It is not difficult to appreciate the personal anxieties of men like Keynes with the flight into mathematics, but it is less obvious why this dispute smoulders on into the 21st century. This brief essay begins with an introduction to the philosophy of the scientific method and its origins within the physical sciences, and then outlines the inadequate methodology of econometric inference. In particular, I wish to draw attention to the evasion of the experimental philosophy of science. In contrast to standard reflections, I will proceed to

interrogate the inability of economic theorists to modify the mechanistic axioms of natural science, in order to highlight the handicap of economic reductionism, so dominant in the discipline. I hope this emphasis on the scientific method will avoid the tedious conflation of the methodological limitations of economics/econometrics with the necessary constraints of rational enquiry. An apology is warranted, due to essay constraints, in view of the inevitable omission of the crucial elaboration of a holistic approach of social science.

The prosperity of inductive reason.

The Scientific Revolution declared a new confidence in the study of knowledge, demonstrated by a string of breakthroughs in astronomy, physics and other natural sciences from the 16th century onwards. In concert with René Descartes, Francis Bacon declared that knowledge ought not aim merely "of causes, and secret motions of things", but should serve "the enlarging of the bounds of human empire, to the effecting of all things possible" (Gillot and Kumar, 1995). Bacon and Descartes shared an appreciation of the human potential of science, but their conception of the *origins* of knowledge were at odds with each other. Descartes favoured the method of abstract deduction, in which knowledge derives from certain logical truths; in contrast, Bacon emphasised learning from empirical observation, a continual perfecting of non-conclusive reasoning (inductivism).

Gillot and Kumar (Ibid) sketch an ABC of the scientific method in the following simple steps: develop a novel theory; design the theory to yield experimentally testable predictions; and, finally, subject the theory to experimental evidence. This approach emphasises Bacon's inductive methodology, a synthesis of observation and abstract reasoning to induce 'experimentally testable' theories. The philosophy of induction paradoxically soothes and upsets the scientific hunger of economic theory, yet it need not be so. An honest appraisal of the scientific veracity of economics and econometrics must acknowledge the primacy of human reason and its retreat in contemporary scientific enquiry. Sadly, a diminished scientific method has become the *cri de coeur* of economists and econometricians in their perennial methodological spats (Hendry, 1980; Blaug, 1980; Hutchinson, 1981; Darnell & Evans, 1990).

Of fundamental interest to philosophers of science is the 'demarcation problem', a dilemma as old as Plato's partition of *nous* and *psyche*, i.e. rational knowledge *contra* instrumental learning. How do the explanations and theoretical constructions of science amputate foul conjecture from the healthy body of scientific learning? Keuzenkamp (2000) traces scientific suspicion of econometric methodology to 'Humean scepticism' outlined in the *Treatise on Human Nature.*

Hume incredulously asks: "why should the sun rise tomorrow morning – apart from the fact that we have never observed it do otherwise?" [1] Karl Popper (1983) repeated Hume's criticism of inductivism to deny the validity of any generalisation or probabilistic inference from empirical observation, explaining his "demarcation principle [as] a purely logical affair" of falsification. As Keuzenkamp (2000) remarks, "such deductive logic is of no use to scientists, economists or econometricians who seek real-world application." Regrettably, however, the methodology of econometrics (model specification, regression analysis and hypothesis testing) accommodates all too easily to the merry-go-round of falsifiability.

The Surrender Of Scientific Certainty.

Popper's "obnoxious rejection of induction" (Ibid) explicitly denied "an objective reality that human beings can understand" – not the total abandonment of the objectivity of nature *per se*, but more so that "the possession of truth is not recognisable." A blunder that, as W.H. Newton-Smith (1990) makes clear dismisses the notion "that there is a growth of scientific knowledge and that science is a rational activity." Sadly, this restriction of human intelligence is a common-sense truth in statistical reasoning and in the econometrics literature (Keuzenkamp, 2000). Laplace's Demon, sketched by the Marquis de Laplace (1749-1827) in *A Philosophical Essay on Probabilities*, presents the stumbling block of scientific advance as the want of "an intelligence sufficiently vast to submit these data to analysis" (Ibid.) However, as Gillot and Kumar (1995) caution: "for the classical physicists… probability was not an intrinsic feature of the world. Rather, it was a consequence of human ignorance." Isaac Newton summed up the materialist sentiment in his *Principia*:

> "Hitherto I have not been able to discover the causes of those properties of gravity from phenomena, and I feign no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy" (Ibid).

[1] Hume's light-hearted violation of intuitivism had minimal mileage - the accomplishments of the scientific project were measure enough to preserve faith in reason. Ironically, though, econometrics was established just as scientists' conviction in the method of induction was fading: the world of Newtonian science had just degenerated into turmoil with the discovery of quantum mechanics; and Karl Popper's caustic attack on science, *The Logic of Scientific Discovery*, was completed within a year of the first issue of *Econometrica* (1933).

The study of econometrics, on the other hand, is littered with sophisticated 'hypotheses' justified only by elaborate mathematics deduced in terms of the model specified. The contrast between the exceptional scepticism of medical science towards statistical inference in epidemiological studies, and its regime of randomised clinical trial clearly illustrates the narrow limits of probabilistic inference and mathematical modelling. Spanos' frank admission that: "no economic theory was ever abandoned because it was rejected by some empirical econometric test" testifies to the theoretical insignificance of statistical hypothesis (Keuzenkamp, 2000). Because mathematics is part 'invention', which is logically 'consistent', a mathematical model of a real-world system – physical or social – can never capture the system itself (Gillot and Kumar, 1995). This inconsistency undermines the foundations of econometrics, but does not explain why economic theory is unable to escape from the mire of mathematical modelling.

Positivism: Sciences Sans Raison.

Bacon's inductive method is founded on the belief that nature is governed by underlying mechanical causal interactions. Similarly, Descartes pioneered a purely mechanistic view of all non-human processes that eliminated the vexing tradition of asking 'why', reducing organic and inorganic entities into inert, senseless objects. Animals and plants are studied like machines, not because they have anything in common with TVs or PCs, but because they are mere objects of natural laws to be discovered by science. The progressive feature of Cartesian mechanical philosophy was that it possessed a "conception of universal causality in nature…without it, nature is made unknowable to humanity" (Ibid.) The mechanist approach allowed scientific enquiry to progress from explanations of simple objects to more complex processes, much in the same way that a watch could be understood from the aggregated operations of its simple mechanisms.

The method of induction, in the natural sciences, exploited this mechanistic philosophy to discover natural laws by a theoretical process of *reduction*. Lamentably, economic theorists are all too aware of the complications of mechanistic philosophy in the social sciences yet persist in the instrumentalist efficacy of observation and measurement. Theory is merely "an essential ingredient to classify facts," as inspired by positivist social scientists (Keuzenkamp, 2000). Economists have responded to the inefficiency of reductionism in economic theory by rejecting causation and determinism. Social science is exceptional, because society is the creation of reasoning human subjects, and economics is weakened by this profound difference between the study of nature and the study of society. Keynes plainly differentiated between the physical sciences and the moral sciences

(including economics), but is less at ease with the capacity of inductive reasoning to discover the *causa essendi* (the objective cause) – electing the inferior, imperfect *causa cognoscendi* (Ibid). Meanwhile, the Austrian economist, Friedrich Hayek openly admits his desire to set boundaries to human potential in economics, and dismisses economic policy as mere "pretence of knowledge" that yield only unintended consequences. As Hayek's colleague, Israel Kirzner, explains:

> "Our dissatisfaction with empirical work and our suspicion of measurement rest on the conviction that empirical observations of past human choices will not yield any regularities or any consistent pattern that may be safely extrapolated beyond the existing data at hand to yield scientific theorems of universal applicability" (Ibid.)

In contradiction, I remain upbeat that economic laws can be discovered by inductive reasoning and that a precondition of such scientific enquiry is a holistic focus on the economic sphere as the outcome of social relations. The progress of science (natural and social) owed its vitality to the forward-looking optimism in human potential throughout the 'machine age'. Today human industry is belittled. Until the spineless fashion of economists to seek refuge from the malaise of modern-day life in the aesthetics of mathematics or the microcosm of micro-theory is reversed, economics will remain, the dismal science.

Bibliography

Bailey, N. (1967), *The Mathematical Approach to Biology and Medicine*. London: Wiley.

Blaug, M. (1980), *The Methodology of Economics or How Economists Explain*. Cambridge: Cambridge University Press.

Darnell, A. and Evans, J. L. (1990), *The Limits of Econometrics, Edward Elgar*. Hants: Aldershot.

Gillot, J. & Kumar, M. (1995), *Science and the Retreat from Reason*. London: Merlin Press.

Hendry, D. (1980), "Econometrics: Alchemy or Science?" *Economica*, Vol. 47.

Hutchinson, T. (1981), *The Politics and Philosophy of Economics*. New York: New York University Press.

THE GAME OF NUISANCE SUITS

BY LAURA FARRELL

Senior Sophister

Yet another interesting application of Game Theory is used in this essay by Laura Farrell. Within this framework of analysis, she tries to explain the reasons behind the so-called nuisance suits. Although the scope of the paper does not allow for a full in-depth analysis of the phenomenon, it gives important insights into the nature of these suits and suggests ways of reducing their incidence.

Introduction

The economic analysis of law, relying on formal models, is a relatively new area in economics. From its modest beginnings in the 1960s, the economic analysis of law became an intellectual fad in the 1970s. However, the continuing progress of the subject remains impressive (Cooter and Rubinfeld, 1989). Law is particularly well suited for analysis by game theory, because the legal process is so concerned with conflict and the provision of definite rules to regulate that conflict. In law a major objective is to avoid inefficient outcomes by restructuring the rules and nuisance suits are one of the inefficiencies that a good policy maker hopes to eliminate (Rasmusen, 1994). One of the chief problems of court reform in recent years has been the nuisance suit or frivolous suit. This essay models the game of nuisance suits and attempts to explain why these are brought. Some solutions to the problem will be suggested, the goal being to reduce the number of nuisance suits.

In his article in *The New Palgrave Dictionary of Economics and the Law* Rasmusen (1998) defines a nuisance suit as a lawsuit with a low probability of success at trial, brought even though the plaintiff knows that his probability of prevailing would not justify his costs, if the judicial process were to be completed instantly. A negative expected value (NEV) suit might be classified as a nuisance suit in certain circumstances. A NEV suit is one in which the plaintiff would obtain a negative expected return from pursuing the suit all the way to judgment i.e. one in which the plaintiff's expected total litigation costs would exceed the expected

judgment (Bebchuk, 1998).[1] It is generally believed that cases with NEV suits are abundant. NEV suits constitute a large fraction, if not the overwhelming majority of small stakes cases (Bebchuk, 1996). The question we seek to answer is: "Why are these suits brought?"

The literature

The early economic literature on litigation, beginning in the early 1970s with papers by Landes (1971) and Gould (1973), has largely avoided the puzzle of NEV suits by focusing on settlement decisions in positive expected value (PEV) suits. In the 1980s, papers written to model pre-trial negotiation in the presence of asymmetric information, all assumed that the plaintiff's suit has a positive expected value.[2]

Only in the late 1980s, did research begin to focus on NEV suits. Rosenberg and Shavell (1985) in their paper *A Model in which Nuisance Suits are brought for their Nuisance Value* put forward an explanation that applies to some NEV suits. This model will be examined later.

Bebchuk (1987) seeks to answer the question; "Will a plaintiff that does not intend to go to trial succeed in extracting a settlement offer and if so, how much will he get?". His article written in 1984, was concerned with the factors that determine the likelihood of settlement in cases involving NEV suits. The explanation that he gives in these papers focuses on the effects of imperfect information. In 1996 Bebchuk wrote; "A New Theory Concerning The Credibility and Success of Threats to Sue" in which he addressed the question: "What can make NEV suits credible?" The defendant would not agree to any positive settlement amount unless he found the plaintiff's threat to litigate credible. In this article, he pursues the idea that divisibility of the litigation process can provide a plaintiff with a credible threat and enable him to extract a settlement, even if the plaintiff is known by the defendant to have a NEV suit.

The model

To answer our key question as to why nuisance suits are brought let us begin by describing the model of the nuisance suit game. The model is a dynamic game with symmetric information.

---

[1] A meritorious suit - one in which the likelihood of a plaintiff victory is quite high – might be NEV, if the litigation costs involved are sufficiently large relative to the amount at stake (Bebchuk, 1998).

[2] For example, Ordover and Rubinstein (1983), P'ng (1983) and Salant and Rest (1982) offered bargaining models of settlement decisions in the presence of asymmetric information.

Nuisance Suits I [3] models the essentials of the situation:

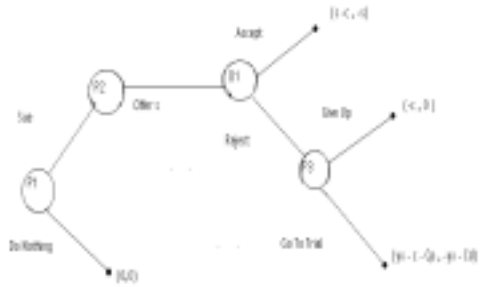
Nuisance Suits I: Simple Extortion

Players
A plaintiff and a defendant

The Order of Play

1. The plaintiff decides whether to bring suit against the defendant at cost *c* (an amount meant to include all initial expenses).
2. The plaintiff makes a take-it-or-leave-it settlement offer of $s > 0$.
3. The defendant accepts or rejects the settlement offer.
4. If the defendant rejects the offer, the plaintiff decides whether to give up or go to trial at a cost *Cp* to himself and *Cd* to the defendant.
5. If the case goes to trial, the plaintiff wins amount *x* with probability *y* and otherwise wins nothing (facts known to both sides and the court).

Payoffs
Figure 1 shows the payoffs. Let $yx < Cp$, so the plaintiff's expected winnings are less than his marginal cost of going to trial. This is what makes this a nuisance suit.



Payoffs: (Plaintiff, Defendant)

Figure 1: The extensive Form for Nuisance suits

[3] Box adapted from Rasmusen (1994).

To find a perfect equilibrium let us pursue the analysis by backward induction. The backward induction approach suggests that the analysis should start by examining what parties will decide in the final stage. We proceed backwards, identifying the decision that will be reached in each stage based on the understanding of the future consequence of each action. The analysis is complete once we reach the initial stage (Bebchuk, 1996).

At node P3, the plaintiff will choose 'Give Up', since by assumption:

$$yx - c - Cp < -c$$

This is because the suit is brought only in the hope of settlement, not in the hope of winning at trial. At node D1, the defendant, foreseeing that the plaintiff will give up, rejects any positive settlement offer. This makes the plaintiff's offer at P2 irrelevant, and, looking ahead to a payoff of $-c$ from choosing 'Sue' at P1, the plaintiff chooses 'Do Nothing' (Rasmusen, 1994).

The perfect equilibrium is:

- Plaintiff: 'Do Nothing', 'Offers', 'Give Up'
- Defendant: 'Reject'
- Outcome: The plaintiff does not bring a suit[4]

Thus, if nuisance suits are brought, it must be for some reason other than the obvious one, the plaintiff's hope of extracting a settlement offer from a defendant who wants to avoid trial costs. That hope is forlorn because the plaintiff himself bears trial costs and hence cannot credibly make the threat. It is forlorn even if the defendant's legal costs would be much higher than the plaintiff's ('Cd' much bigger than 'Cp'), because the relative size of the costs does not enter into the argument (Rasmusen, 1994). "The catch is that the plaintiff's threat to go to trial must be credible, but if his settlement offer is turned down, he will give up rather than pay 'Cp' for a probability 'y' of 'x'. Hence the size of 'Cd' is irrelevant" (Rasmusen, 1998).

Risk Aversion

How does risk aversion affect this conclusion? So far we have assumed both parties are risk neutral. Nuisance Suits I can be adapted to risk-averse players. Risk would enter at the trial stage, as a final move by Nature to decide who wins. In Nuisance Suits I, 'yx' represented the expected value of the award. If both the

[4] The equilibrium specifies actions at all four nodes of the game, even though only the first is reached in equilibrium.

defendant and the plaintiff are equally risk averse, 'yx' can still represent the expected payoff from the award – one simply interprets 'x' and '0' as the utility of the cash award and the utility of an award 0, rather than as the actual cash amounts (Rasmusen, 1994).

If the players have different degrees of risk aversion, the expected loss to the defendant is not the same as the expected gain to the plaintiff, and the payoffs must be adjusted. If the defendant is more risk averse, the payoffs from 'Go to Trial' would change to:

$$-c - Cp + yx, -yx - w - Cd$$

Where *w* represents the extra disutility of risk to the defendant. This, however, makes no difference to the equilibrium (Rasmusen, 1994). "The crux of the game is that the plaintiff is unwilling to go to trial because of the cost to himself, and the cost to the defendant, including the cost of bearing risk, is irrelevant" (Rasmusen, 1994). If nuisance suits are brought, it must therefore be for some more complicated reason.

The explanation

There are twists to the situation that can make the plaintiff's threat to go to trial credible, so suit-for-settlement becomes plausible. Two central explanations as to why nuisance suits are brought in practice are put forward namely sinking costs strategically and malicious emotions. Our earlier model is adapted to take account of each of these issues. Other explanations will also be briefly discussed.

Nuisance Suits II: Using Sunk Costs Strategically[5]

Suppose that the plaintiff can pay his lawyer the amount 'Cp' in advance, with no refund if the case settles. This inability to obtain a refund actually helps the plaintiff, by changing the payoffs from the game, so his payoff from 'Give Up' is:

$$-c - Cp$$

Compared to:

$$-c - Cp + yx$$

from 'Go to trial'.

Having sunk the legal costs, he will go to trial if:

$$yx > 0$$

[5] This model is based on Rosenberg and Shavell (1985).

i.e. if he has any chance of success at all (Rasmusen, 1994).

This, in turn, means the plaintiff would only prefer settlement to trial if:

$$s > yx$$

The defendant would prefer settlement to trial if:

$$s < yx + Cd$$

So there is a positive settlement range of:

$$(yx) \text{ to } (yx + Cd)$$

Within which both players are willing to settle. The exact amount of the settlement depends on the bargaining power of the parties. Here, allowing the plaintiff to make a take-it-or-leave-it offer means that:

$$s = yx + Cd$$

in equilibrium, and if:

$$yx + Cd > Cp + c$$

the nuisance suit will be brought even though:

$$yx < Cp + c.$$

Thus the plaintiff is bringing the suit only because he can extort 'Cd', the amount of the defendant's legal costs (Rasmusen, 1994).

Even though the plaintiff can now extort a settlement, he does it at some cost to himself, so equilibrium with nuisance suits will require that:

$$-c - Cp + yx + Cd \text{ > or = } 0 \text{ (equation 1)}$$

If equation 1 is false, then even if the plaintiff could extract the maximum possible settlement of:

$$s = yx + Cd$$

he would not do so, because he would then have to pay:

$$c + Cp$$

before reaching the settlement stage. This implies that a totally merit less suit (with $y = 0$) would not be brought unless the defendant had higher legal costs than the plaintiff ($Cd > Cp$). If equation 1 is satisfied, however, the following strategy combination is a perfect equilibrium:

Plaintiff: 'Sue', Offer: $s = yx + Cd$, Go to trial
Defendant: Accept $s <$ or $= yx + Cd$
Outcome: Plaintiff sues and offers to settle, to which the defendant agrees.

The Open Set Problem in Nuisance Suits II

Nuisance Suits II illustrates a technical point that arises in a great many games with continuous strategy spaces as Rasmusen (1994) points out to us. The equilibrium is only a weak Nash equilibrium. The plaintiff proposes,

$$s = yx + Cd$$

and the defendant has the same payoff from accepting or rejecting, but in equilibrium the defendant accepts the offer with probability one, despite his indifference. This seems arbitrary. Should not the plaintiff propose a slightly lower settlement to give the defendant a strong incentive to accept it and avoid the risk of going to trial? For example, if the parameters are such that,

$$s = yx + Cd = 60$$

why does the plaintiff risk holding out for 60 and possibly be rejected and receive 0 at trial, when he could offer 59 and give the defendant a strong incentive to accept? "One answer is that no other equilibrium exists besides $s = 60$. Offering 59 cannot be part of an equilibrium because it is dominated by offering 59.9; offering 59.9 is dominated by offering 59.99, and so forth" (Rasmusen, 1994). This is known as the open-set problem, because the set of offers that the defendant strongly wishes to accept is open and has no maximum. Therefore, a weak Nash equilibrium is still a Nash equilibrium (Rasmusen, 1994).

Rasmusen (1994) shows us that one can avoid the open set problem by specifying a more complicated bargaining game to avoid the issue of how exactly the settlement is determined. One could say that the settlement is not proposed by

the plaintiff, but simply emerges with a value halfway through the settlement range, so,

$$s = yx + Cd/2$$

This kind of modelling is called black boxing, because it is as of at some point in the game, variables with certain values go into a black box and come out the other side with values determined by an exogenous process (Rasmusen, 1994).

Nuisance Suits III: Malicious Emotions

Emotions are often important to lawsuits, and law professors tell their students that when the cases they study seem to involve disputes too trivial to be worth taking to court, they can guess that the real motivations are emotional. "Emotions could enter in a variety of distinct ways" (Rasmusen, 1994). The plaintiff might simply like going to trial, which can be expressed as a value of:

$$Cp < 0.$$

This would be true of many criminal cases, because prosecutors like news coverage and want credit with the public for prosecuting certain kinds of crime (Rasmusen, 1994).

Another emotional motivation for going to trial is the desire to inflict losses on the defendant, a motivation called malice. In this case, 'Cd' enters as a positive argument in the plaintiff's utility function. Rasmusen (1994) constructs a numerical example of Nuisance Suits III in which he assumes:

$y = 0.1$ $\qquad$ $Cd = 50$
$c = 3$ $\qquad$ $x = 100$
$Cp = 14$

and that the plaintiff receives additional utility of 0.1 times the defendant's disutility. He adopts the black boxing technique here and assumes that the settlement *s* is in the middle of the settlement range. The payoffs conditional on suit being brought are:

Plaintiff (Defendant accepts):

$$s - c + 0.1s = 1.1s - 3 \qquad \text{(Equation 2)}$$

Plaintiff (Go to trial):

$$yx - c - Cp + 0.1(Cd + yx) = 10 - 3 - 14 + 6 = -1 \qquad \text{(Equation 3)}$$

The plaintiff's payoff from 'Give up' is –3 therefore, he will go to trial if the defendant rejects the settlement offer. The overall payoff from bringing a suit that eventually goes to trial is still –1, which is worse than the payoff of 0 from not bringing suit in the first place. If s is high enough, the payoff from bringing suit and settling is higher still. If s is greater than 1.82, the plaintiff prefers settlement to trial and if s is greater than 2.73, he prefers settlement to not bringing the suit at all.[1] The plaintiff will settle for any:

$$s > \text{or} = 1.82$$

and the defendant will settle for any:

$$s < \text{or} = yx + Cd = 60$$

The settlement range is [1.82, 60], and s = 30.91. The equilibrium is as below:

- Plaintiff: Sue, Go to trial
- Defendant: Accept any s < or = 60
- Outcome: The plaintiff sues and offers s = 30.91, and the defendant accepts the settlement.

Incomplete Information

Bebchuk (1987) extended to NEV suits the model put forward by Bebchuk (1984) to analyse litigation decisions under asymmetric information, and it demonstrated that the presence of an informational asymmetry could explain the success of some NEV suits.

To see the role played by asymmetric information consider the situation where the defendant may not know whether the plaintiff's expected value of litigation is positive or negative. This uncertainty might result from private information that the plaintiff has about 'Cp', for example (Bebchuk, 1998). Consequently, the defendant cannot be certain that, in the absence of a settlement, the plaintiff would drop his suit. To avoid the risk of incurring litigation costs, the defendant therefore may elect to offer a positive settlement amount. This uncertainty model cannot explain, however, the success of suits that defendants know to be NEV suits (Bebchuk, 1996).

[1] The figure 1.82 = -1 + 3/ 1.1, rounded.
The figure 2.73 = 0 + 3/ 1.1, rounded.

Divisibility of Litigation Costs

As we have seen in the literature review, Bebchuk (1996) has shown that the divisibility of the litigation process can provide a plaintiff with a credible threat, and enable it to extract a settlement, even if the plaintiff is known by the defendant to have a NEV suit. What underlies the considered explanation is the recognition that litigation costs are generally not incurred all at once in a lump sum fashion but rather are spread over a period of time, with bargaining possibly taking place on numerous occasions throughout this period (Bebchuk, 1998). Suppose that the plaintiff cannot pay his legal expenses first, but both sides incur costs in 'T' stages, and at 'T–1', once most costs are sunk, the plaintiff finds it worthwhile to incur the costs of the last stage so as to have a chance at the trial judgment. He can then extract a settlement. Going back to 'T–2' the plaintiff would be willing to pay expenses of that stage, as the price of admission to the profitable settlement of 'T–1'. The reasoning continues back to stage one. The result is that the settlement is an agreement by both players to exit a costly rivalry that is profitable at each stage but ruinous overall (Rasmusen, 1998). Bebchuk (1997) proves that a finer division of the litigation process might sometimes improve, and can never worsen, the strategic position of the plaintiff and its ability to extract a settlement. However, as Bebchuk (1996) warns divisibility cannot always guarantee a NEV plaintiff with a credible threat. [2]

Solutions

Society seeks to induce plaintiffs to sue, if and only if, they believe they are entitled to prevail at trial. One needs to design rules that will ensure that optimal decisions to bring suit are made. This objective implies two goals: that a plaintiff bring a meritorious suit (which we define as a suit that deserves to win on the merits, as the plaintiff views the case) and that a plaintiff not bring a nuisance suit.

Fee-shifting rules

Bebchuk and Chang (1996) analyse how fee-shifting rules best serve the goals outlined above. They show that when plaintiffs cannot predict the outcome of litigation with certainty, neither the American rule (each litigant bears its own litigation expenses) nor the British rule (the losing litigant pays the attorney's fees of the winning litigant) would induce optimal decisions to bring suit.

---

[2] For an analysis of the factors that determine when divisibility can and cannot provide a credible threat see Bebchuk (1996) and Bebchuk (1997).

For example, under the American rule, plaintiffs will not bring all meritorious suits; the plaintiff will not sue if its litigation costs exceed the value of the relief that it expects the court to award. Also, the plaintiff will bring some frivolous suits; if the litigation costs are small enough, the plaintiff will find it worthwhile to gamble, either because the court might rule in the plaintiff's favour or because the case might prove to be better than it first appears.

Rasmusen (1998) notes that by shifting from the American rule to the British rule one can reduce the payoff at trial and so the British rule reduces the credibility of threats to go to trial with meritless suits.

Bebchuk and Chang (1996) also analyse the effects of more general fee-shifting rules that are based not only on the identity of the winning party, but also on how strong the court perceives the case to be at the end of the trial – that is, the 'margin of victory'. They explore how and when one can design such a rule to induce plaintiffs to sue, if and only if, they believe their cases are sufficiently strong. This analysis of fee-shifting rules based on the margin of victory is not only of theoretical interest but also of practical significance. Courts have interpreted Federal Rule of Civil Procedure 11 as an example of such a rule. In order to deter parties from filing frivolous papers in court, Rule 11 authorises courts to impose sanctions on those who file such papers.

Procedural Rules

These rules delegate the deterrence of nuisance suits to the judge in the original proceedings. First, the rules on how suit is brought can make nuisance suits difficult. These include rules on standing, limiting who can bring suit; on forum shopping, preventing plaintiffs from going to sympathetic or corrupt judges; on pleadings, requiring a specific-enough filing that its lack of merit is apparent; on evidence, barring hearsay and limiting expert testimony. Given these rules the judge has the ability to dismiss suits by summary judgment. Second, the judge can punish as well as dismiss, the purpose being deterrence (Rasmusen, 1998).

Conclusion

From our model of the game of Nuisance Suits I, we have seen that if nuisance suits are brought, it must be for some reason other than the obvious one, the plaintiff's hope of extracting a settlement offer from a defendant who wants to avoid trial costs. Our analysis has illustrated some of these reasons including sinking costs strategically and malicious emotions. Some rules, that one might have designed in order to induce optimal decisions to bring suit, have been examined. However this analysis is only the tip of the iceberg, many more explanations exist

for nuisance suits. These include defendant's ignorance, plaintiff reputation, plaintiff mistake, court error and unjust law. Also, other rules and policy instruments exist that aim to reduce the number of nuisance suits, for example, tort rules.

As we have seen law (and) economics research in the area of nuisance suits has been active since the late 1980s, but though it has illuminated extortionary settlements and the effects of fee-shifting, it has neglected many other aspects of the problem. Perhaps the biggest blanks, as Rasmusen (1998) tells us, are in the areas of public choice theory and theory-driven empirical study.

Bibliography

Bebchuk, L. (1984), "Litigation and settlement under imperfect information." *Journal of Economics,*Vol. 15.

Bebchuk, L. (1987), "Suing solely to extract a settlement offer" *Journal of Legal Studies* Vol. 17.

Bebchuk, L. (1996), "A New Theory Concerning the Credibility And Success Of Threats To Sue" *Journal of Legal Studies,* Vol. 25.

Bebchuk, L. (1997), "On divisibility and credibility: the effects of the distribution of litigation costs over time on the credibility of threats to sue" [Working Paper]. Cambridge; Harvard Law School.

Bebchuk, L. (1998), "Suits with negative expected value" In: *The New Palgrave Dictionary of Economics and the Law*. London: Macmillan Press.

Bebchuk, L. and Chang, H. (1996), "An analysis of fee-shifting based on the margin of victory: on frivolous suits, meritorious suits, and the role of Rule 11" *Journal of Legal Studies,* Vol 25.

Cooter, R. and Rubinfeld, D. (1989), "Economic analysis of legal disputes and their resolution" *Journal of Economic Literature,* Vol. 27.

Rasmusen, E. (1994), *Games and Information.* Oxford: Basil Blackwell.

Rasmusen, E. (1998), "Nuisance suits." In: *The New Palgrave Dictionary of Economics and the Law.* London: Macmillan Press.

Rosenberg, D. and Shavell, S. (1985), "A model in which lawsuits are brought for their nuisance value." *International Review of Law and Economics* Vol. 5.

Shavell, S. (1982), "Suit, settlement, and trial: a theoretical analysis under alternative methods for the allocation of legal costs." *Journal of Legal Studies,* Vol. 11.

THE VALIDITY OF MARSHALLIAN CONSUMERS' SURPLUS

BY BARRY JOHN RAFFERTY

Senior Sophister

The concept of the Marshallian consumers' surplus is probably familiar to every economics student. However, not everyone is aware of the pitfalls associated with this seemingly 'harmless' tool of economic analysis. Building on strong theoretical grounds, this essay by Barry John Rafferty warns about the limitations of consumers' surplus, highlights situations when its use is justified and gives an overview of alternative measures of welfare change.

Introduction

Marshallian consumers' surplus (hereafter referred to as consumers' surplus) is one of the most controversial concepts in economics. It is the most widely used tool of welfare analysis, but it is based on debatable theoretical foundations. Consumers' surplus seeks to provide a cardinal measure of the surplus utility a group of individuals get, cumulatively, from consuming a quantity of a good at a given price (Currie et al, 1971). This is the absolute net consumers' surplus. In this essay, I will be primarily concerned with the change in consumers' surplus, for a group of individuals, as a result of a move from a base economic state (seen here as a set of prices and incomes) to another economic state, and the use of this measure to rank different states relative to the base state.

The change in the individual consumer's surplus from a base state to another state, seeks to provide a cardinal ranking of the two states for the individual. Aggregating across individuals, the change in consumers' surplus seeks to provide a cardinal ranking for many individuals together of different states. Aggregating across goods seeks to provide a cardinal ranking of states when multiple markets are affected by different states (i.e. if the prices of many goods change). Therefore, used this way consumers' surplus seeks to be a powerful tool of welfare analysis, enabling us to cardinally rank various states for policy purposes when many individuals and many markets are affected. In these circumstances, it seeks to indicate how much society is better off (in terms of utility) in one state relative to another.

In light of the controversy surrounding consumers' surplus, I will analyse the problems associated with its use for the role outlined above. I will seek to answer a number of pertinent questions. Can the change in consumers' surplus from the base state to alternative states be used to cardinally rank alternative states relative to the base state? If not, then can it be used as the correct money measure of the change in welfare relative to the base state? By this I mean that the difference in consumers' surplus between alternative states and the base state would be a correct measure of the amount of money that the group of individuals are better off/worse off in one state relative to the base state. We cannot evaluate the difference in utility between states, since we do not know how much each individual subjectively values the monetary figure of the change in individual consumer's surplus and therefore we do not know what the sum of their valuations is. The correct money measure of the change in welfare (of the group of individuals) relative to the base state correctly ranks states with higher values corresponding to more preferred states (for the group together). If the change in consumers' surplus relative to the base state is not a correct money measure of welfare change, then what are the alternatives? Furthermore, do we need a money measure of the change in welfare, or can we just use ordinal preferences to evaluate whether one state is preferred to another? If we do see fit to use a money measure of welfare change that correctly ranks states, then could and should consumers' surplus be used as a good approximation?

What is Marshallian Consumers' Surplus?

Marshall quoted in Hicks (1941), referred to individual consumer's surplus as ''the excess of the price which he would pay rather than go without the thing, over that which he actually does pay''. This is absolute net consumer's surplus. Aggregating across individuals gives the consumers' surplus for the good. When applied to a single good, we can obtain total consumers' surplus for the good from the area below the ordinary market demand curve (Marshallian) and above the price line. This area can be expressed formally, as the definite integral below, where x is the quantity demanded of the good in question and P^0 is its price.

$$\int_{p^\circ}^{\infty} x(t)dt$$

The change in consumers' surplus from one state to another (where only the price of the good in question is changing between states) can be derived from the difference in total consumers' surplus, for the good, between the two states. That is

consumers' surplus in the new state minus consumer's surplus in the base state. This can be derived from the definite integral below, where P' is the price in the new state.

$$\int_{p^\circ}^{p'} x(t)dt$$

For Marshall, the above definite integral would be a cardinal measure of the change in utility for the group of individuals affected, of a move from the state with P^0 to the state with P'. Aggregating across goods when multiple prices are affected involves summing the changes in consumers' surpluses in the various markets affected.

Marshallian Consumers' Surplus as a cardinal measure of utility and use of changes in Consumers' Surplus from a base state to alternative states to cardinally rank states:

The main problem with consumers' surplus is that it purports to be a cardinal measure of utility. Changes in consumers' surplus from the base state are used to cardinally rank different states. This means that we can say by how much more a state is preferred to another state. We can get the difference in aggregate utility between two states. Utility theory nowadays, however, is purely ordinal. It is claimed that we cannot get a unique cardinal measure of utility or changes in utility. Samuelson claims that without ''introspective information'', obtained under interrogation, we cannot know about the intensity of an individual's preferences (Morey, 1984). Hence, we can know nothing of the intensity of aggregate preferences between states. We can only ordinally rank states. We can say that one state is preferred to another state, but not by how much. In light of this, the ability of consumers' surplus to measure cardinal utility and changes in cardinal utility becomes extremely tenuous. I will now examine the special case where consumers' surplus can be used to measure cardinal utility and changes in cardinal utility.

For consumers' surplus to be an exact cardinal measure of utility and for the change in consumers' surplus of a move from the base state to another state, to be used to construct a meaningful cardinal preference ordering between different states[1] then two things must hold.

Firstly, the marginal utility of money must be constant. This must be the case since there must be a constant conversion factor between changes in money and changes in utility. Therefore, the marginal utility of money must be constant

[1] i.e. we can say by how much utility the group of individuals are better off in one state relative to another

regardless of how much income the group of individuals have to spend, the amount of other goods they have, and the size of consumers' surplus. Every single monetary unit of consumers' surplus must be converted to utility at the same marginal rate.

Secondly, preferences for each good x, affected by a change in price must be represented by a quasilinear utility function. This would mean that all individuals face this utility function and therefore, we would be able to use it to represent aggregate preferences. Such a function would be of the form:

$$U(x, y) = v(x) + y$$

in the case where y refers to a composite good comprising all goods other than x (Dooley, 1983). Such a utility function is linear in y and (possibly) non-linear in x. For Marshall, such a function implied decreasing marginal utility in the first good and constant marginal utility in the second good (Ibid). Where y was money to spend on other goods, this implied a constant marginal utility of money. Such a utility function has the implication that the demand for x is independent of the level of income, implying that there is no income effect and that x has a vertical Engel curve. It is important to note that in this instance the ordinary (Marshallian) market demand curve is identical to the Hicksian compensated market demand curve.

The use of consumers' surplus as a cardinal measure of utility and therefore, changes in consumers' surplus to cardinally rank states, has faced a lot of criticism. The problems associated with the use of consumers' surplus as a cardinal measure of utility and utility change stems mainly from the validity (or lack of validity) of the assumptions required for consumers' surplus to be a cardinal measure. Dooley (1983) describes four main criticisms of consumer's surplus, and by extension consumers' surplus (since consumers' surplus is the sum of individual consumer's surpluses), which I will now outline.

Firstly, whether an additive utility function (with independent utilities for each good) adequately explains consumer behaviour. Marshall uses an additive utility function (the quasilinear utility function) to justify the use of consumer's surplus as a correct cardinal measure of utility. However, this was criticised heavily by Patten (Ibid) who argued that the utility an individual derived from consuming a quantity of a good was not independent of the quantity of other goods, which he could also consume. He argued that the utility of the good also depended on the amount of other goods that the individual could consume.

Secondly, whether the marginal utility of income could be treated as a constant. This is undoubtedly the most controversial assumption used by Marshall to allow consumer's surplus (and therefore consumers' surplus) to be a cardinal measure of utility. However, rigorous analysis has been conducted to show the implications of the constancy assumption. The best-known analysis of the constancy

assumption was by Samuelson (1966). He proved that the marginal utility of income could not be constant with respect to both price changes and to changes in income. Analysing the implications first of the marginal utility of money being constant with respect to price changes, Samuelson showed that the income elasticity of all goods must be unity. Adding the assumption of independent and additive utility, the price elasticity of demand would also have to be unity to ensure constancy of the marginal utility of income. Analysing the implications of the marginal utility of money being constant with respect to changes in income; Samuelson showed that all income would be spent on a single good only, with the income elasticity of demand for all other goods being zero, which implies all other goods having vertical Engel curves. All of these implications are completely at variance with empirical evidence. It is clearly the case that the constancy of the marginal utility of income is not a valid assumption.

Thirdly, whether the quantity demanded of a good can be treated as a function of its price alone. Walras criticised Marshall's assumption of partial equilibrium analysis claiming that one could not vary the price of a good, while holding constant the prices of all other goods and productive services (Dooley, 1983). In particular, he argued that ‘‘the selling prices of products and the prices of productive services are mutually interrelated’’(Ibid).

Fourthly, whether it is possible to make interpersonal comparisons. To get an aggregate consumers’ surplus, it is necessary to make interpersonal comparisons. Marshall assumed that ‘‘a shilling’s worth of gratification to one Englishman might be taken as equivalent with a shilling’s worth to another’’ (Ibid). This assumption would enable us to use the market demand curve to get the aggregate consumers’ surplus. This assumption however implies that each consumer has the same utility function and the same level of income. Also with regard to changes in economic state, Marshall claimed that ‘‘it happens that by far the greater number of events with which economics deals, affect in about equal proportions all the different classes of society’’ (Ibid). These two interpersonal assumptions do not hold since we do not know how each individual subjectively values their respective consumer’s surpluses. We do not know how a change in aggregate consumers’ surplus will affect individuals. Therefore, there is no way therefore that we can arrive at a precise cardinal value for the change in utility for a group of individuals, as a result of a change in state.

It is evident therefore that consumers’ surplus cannot be used to obtain a cardinal measure of utility and that use of the change in consumers’ surplus from the base state to cardinally rank states, for the group of individuals is not permissible.

Circumstances under which the change in Consumers' Surplus is the correct money measure of the change in welfare between two states:

Having established that the change in consumers' surplus from the base state cannot be used to cardinally rank and compare states; can the change in consumers' surplus from the base state be used to provide the correct money measure of the change in welfare? The correct money measure takes into account and compensates for income effects. It correctly ranks states, in terms of ordinal preferences, with higher values of the money measure representing more preferred states. The correct money measure is therefore a cardinally scaled monotonic transformation of the ordinal utility function (Morey, 1984). It also provides an indication in monetary terms of how much the group of individuals is better off/worse off, relative to the base state. In addition, the correct money measure gives the same money figure when multiple prices change, regardless of the sequence of the price changes (Johansson, 1991).

For the change in consumers' surplus to provide the correct money measure of the change in welfare, we may drop the assumption of the marginal utility of money being constant. However, aggregate preferences must still be represented by a quasilinear utility function. There must still be no income effect, with the ordinary market demand curves being identical to the Hicksian market demand curves. However this is extremely restrictive, with the income effect for most goods rarely if ever being zero. Therefore, the change in the consumers' surplus will be inaccurate as a money measure of welfare change the higher the income effect. It will fail to give the correct money measure of welfare change and may rank states incorrectly. Also, when multiple prices change, consumers' surplus may be affected by the path dependency problem, i.e. the sequence in which prices change.

Alternatives to Marshallian Consumers' Surplus:

Seeing as consumers' surplus and the change in consumers' surplus from a base state may fail to provide a money index that correctly ranks different states, but what are the alternatives? There are two options available. We can get a correct money measure of the welfare change or we could simply use the ordinal preference field described by Samuelson (Bergson, 1975) to ordinally rank states.

If we choose to use the correct money measure of welfare change relative to a base state to rank states, there are two alternatives. These are the compensating and equivalent variations developed by Hicks. According to Johansson (1991), the

compensating variation for a household[2] gives ‘‘the maximum (minimum) amount of money that can be taken from (must be given to) a household to make it just as well off as it was before a fall (rise) in prices”. Meanwhile, the equivalent variation for a household gives ‘‘the minimum (maximum) amount of money that must be given to (taken away from) a household to make it as well off as it would have been after a fall (rise) in price’’ (Ibid). The compensating variation returns the agent to the original indifference curve and is measured at the new prices. The equivalent variation is measured at the original prices and takes the agent to the new indifference curve. Summing both of these measures over individuals will yield the aggregate compensating and equivalent variations (compensating and equivalent variations will refer hereafter to the aggregated variations, unless otherwise stated). The two money measures will differ in all cases except quasilinear utility[3], since they are measured at different prices, with the value of a dollar depending on what the relevant prices are (Varian, 1992). Unlike the case of consumers’ surplus, the compensating and equivalent variations will give the correct money figure of the change in welfare when multiple prices change regardless of the order of the price changes.[4] Both measures will correctly rank any two states. Whilst the compensating variation cannot rank more than two states (since we would not be using the same prices), the equivalent variation can rank all states since it is measured relative to the base prices (Johansson, 1991; Morey, 1984).Therefore, all states will be ranked correctly in terms of ordinal preferences if the equivalent variation is used. Compensating and equivalent variations serve distinct purposes with compensating variation more useful for arranging compensation at the new prices (Varian, 1992) and the equivalent variation more useful for ranking states at the base prices.

The other alternative is to use the ordinal preference field spoken about by Samuelson to ordinally rank states, (Bergson, 1975) obviating the need for a money measure of welfare change between states. Bergson (Ibid) indicates how knowledge of the market demand equations[5] is sufficient information to determine the ordinal preference ordering and to derive an ordinal indifference curve map. Using this map, we can evaluate the effect on the group of individuals of the different states. We can only indicate a preference ordering, as we have no capacity to know the intensity of these preferences. In line with this view, all that is needed for decision-making is that different options or policies are ranked correctly, in so far as the different states corresponding to different policies are ranked correctly.

---

[2] Households here are equivalent to individual agents

[3] Where they will also be the same as the consumers' surplus measure of welfare change

[4] The compensating and equivalent variations are path independent.

[5] Relative prices indicate the marginal rates of substitution between goods.

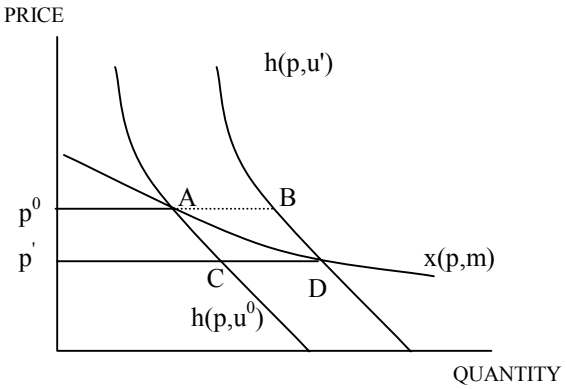
Which option is chosen will depend on whether one views correct money measures of welfare change to be useful for more than just their ordering of states. That is whether one sees any worth in using the value of money measure as some rough indication of the intensity of preferences. Bergson (Ibid) argues that often politicians need a money figure of the benefit or cost of a proposal, against which they can compare political costs or benefits and other factors unaccounted for in the economic evaluation process, e.g. externalities. A simple ordinal ranking may not suffice. As Morey interprets Bergson:

> ''The politician needs cardinally scaled measures of the economic benefits of each alternative to compare them with the political costs of each alternative so as to determine his ordinal ranking of the projects'' (Morey, 1984).

As Bhagwati points out ''whether we like it or not, this is what the policy makers do want'' (Bergson, 1975).

Having established that a correct money measure of welfare change is desirable for decision-makers, how can we calculate the compensating and equivalent variations of changes in states? This will depend on whether we can derive the Hicksian market demand curve, or not. If we can solve the integrability problem, then knowledge of demand will enable us to calculate a money metric utility function that is a monotonic transformation of the ordinal utility function. This in turn will enable us to calculate an expenditure function and consequently a Hicksian market demand curve. We will be able to get a Hicksian market demand curve for the original utility level, and a Hicksian market demand curve for the new utility level. Seen below, we have Hicksian market demand curves in the case where x is a normal good.

Figure 1: Hicksian market demand curves in the case where x is a normal good



For a normal good, the Hicksian market demand curves are h (p,u^0) for the original utility level and h (p,u') for the new utility level. Both of these curves are steeper than the Marshallian market demand curve x (p, m) when x is a normal good.[6] This is due to the fact that the Hicksian demand curves are compensated demand curves along which real income is constant, as opposed to the Marshallian demand curve which holds money income constant but fails to take account of the change in real income as price changes. The compensating variation (CV) between two states, with P' being the new state and P^0 being the original state, can be measured by the definite integral:

$$\int_{p'}^{p^\circ} h(p,u^\circ)dp$$

The equivalent variation (EV) between two states, with P' being the new state and P^0 being the original state, can be measured by the definite integral:

$$\int_{p'}^{p^\circ} h(p,u^\circ)dp$$

In the diagram above we can see that for a price fall from p^0 to p' the compensating variation is given by the area P^0ACP'. The equivalent variation is given by the area P^0BDP'. The change in consumers' surplus (CS) meanwhile is given by P^0ADP'. In this case the identity CV<CS<EV holds. This will be reversed if we are increasing price with the identity CV>CS>EV holding.

Therefore, we can see how to solve for the compensating and equivalent variations, when we can derive the Hicksian market demand curves. However, the problem is that we usually can not derive the Hicksian demand curves. Therefore, we have to estimate the two measures. How can this be achieved? We could use a method developed by McKenzie and Pearce to estimate the equivalent variation. This involves a Taylor series expansion from the base state (P^0, M^0) using the first and higher-order partial derivatives of the demand equations evaluated at (P^0, M^0) to determine the coefficients (Morey, 1984).

However, could we alternatively use the change in consumers' surplus to approximate the compensating and equivalent variations?

[6] If x is an inferior good, the Hicksian demand curves will be flatter than the Marshallian demand curve.

Use of Marshallian consumers' surplus to approximate the unknown compensating and equivalent variations:

Willig (1976) showed that the use of individual consumer's surplus gives a good approximation of the unknown compensating and equivalent variations for individuals when the income elasticity of demand is low or expenditure on the good is a small share of an individual's total expenditure. In his 1976 paper, he established precise bounds for the errors of estimating an individual's compensating and equivalent variations using the consumer's surplus measure of welfare change (i.e. the change in consumer's surplus between two states). From Johansson (1991) these error bounds are as follows:

$$(CV - CS) / CS \approx -\eta (CS / 2y)$$
$$(CV - CS) / CS \approx -\eta (CS / 2y)$$

Where η refers to the income elasticity of demand for the good, and y refers to total income available to spend on goods. According to Willig, if the absolute value of the terms on the right hand side of the equations above is less than 0.05, then these equations accurately reflect the errors of using consumers' surplus to measure the individual's compensating and equivalent variations. Willig (1976) concludes by saying:

> ''...at the level of the individual consumer, cost-benefit welfare analysis can be performed rigorously and unapologetically by means of consumer's surplus''.

This can be extended to justify using consumers' surplus to approximate the compensating and equivalent variations. The necessary criteria are now a low income elasticity of demand for all the goods affected, or that total expenditure on each of the goods affected represents only a small share of the total expenditure by the group of individuals.

Conclusion:

Therefore, consumers' surplus cannot be used as a cardinal measure of utility and the change in consumers' surplus cannot be used to cardinally rank alternative states. However, the change in consumers' surplus does provide a useful approximation to the unknown compensating and equivalent variations. This is providing that income elasticity of demand is low for the goods whose markets are affected by the change, or total expenditure on the goods affected constitutes a small

proportion of total expenditure by the group. However, as we consider changes in multiple prices and the path dependency problem[7] and also as we aggregate over more and more goods (the more and more prices change), errors begin to mount and accumulate. The change in consumers' surplus becomes less valid as a measure of welfare change. However, the money measure will give a rough indication of the benefit of a policy against, which policy makers can compare political and other unaccounted costs. In particularly with large projects, it may be more appropriate to use the McKenzie Pearse technique to estimate the equivalent variations. However, in the case of purely partial welfare analysis, where there are not that many prices changing, and such price changes do not unduly affect other prices, then consumers' surplus would appear to be justified as a useful money measure approximation of welfare change.

Bibliography

Bergson, A. (1975), "A Note on Consumer's Surplus." *Journal of Economic Literature*, Vol. 13.

Chipman, J.S. (1990), "Marshall's consumer's surplus in modern perspective" In: J.K. Whitaker (ed.) *Centenary essays on Alfred Marshall.* Cambridge: Cambridge University Press.

Currie, J.M.; Murphy, J.A. and Schmitz, A. (1971), "The Concept of Economic Surplus and its use in economic analysis" *Economic Journal*, Vol. 8.

Friedman, M. (1949), "The Marshallian Demand Curve" *Journal of Political Economy*, Vol. 57.

Hicks, J.R. (1941), "The Rehabilitation of Consumers' Surplus" *Review of Economic Studies*, Vol. 8.

Hicks, J.R. (1943), "The Four Consumer's Surpluses" *Review of Economic Studies*, Vol. 11.

Johansson, P.O. (1991), *An Introduction to Modern Welfare Economics.* Cambridge: Cambridge University Press.

[7] The overall change in consumer's surplus depends on the sequence of price changes.

Morey, E.R. (1984), "Consumer Surplus." *American Economic Review*, Vol. 74(1).

Samuelson, P.A. (1966), "Constancy of the marginal utility of income" In: J.E. Stiglitz (ed.) *The Collected scientific papers of Paul Samuelson.* Cambridge: The MIT Press. Vol. 1.

Varian, H.R. (1992), *Microeconomic Analysis.* 3rd ed. New York: Norton.

Varian, H.R. (1999), *Intermediate Microeconomics.* 5th ed, New York: Norton.

Willig, R.D. (1976), "Consumer's Surplus without Apology." *American Economic Review*, Vol. 66.

ANALYSIS OF UK INFLATION DYNAMICS USING ARCH AND ALLOWING FOR SEASONAL EFFECTS

BY DAVID MORRISSEY

Senior Sophister

Inflation is the one of the economic phenomena that is in the spotlight of forecasting. In this paper David Morrissey evaluates the dependency of UK inflation on its previous values, extending the analysis to include seasonality later in the text. The econometric analysis is conducted using an ARCH model with various extensions that appears to be superior to its OLS counterpart. Finally, the author concludes that his model has sound explanatory and forecasting powers and suggests further extensions for analysis.

Introduction

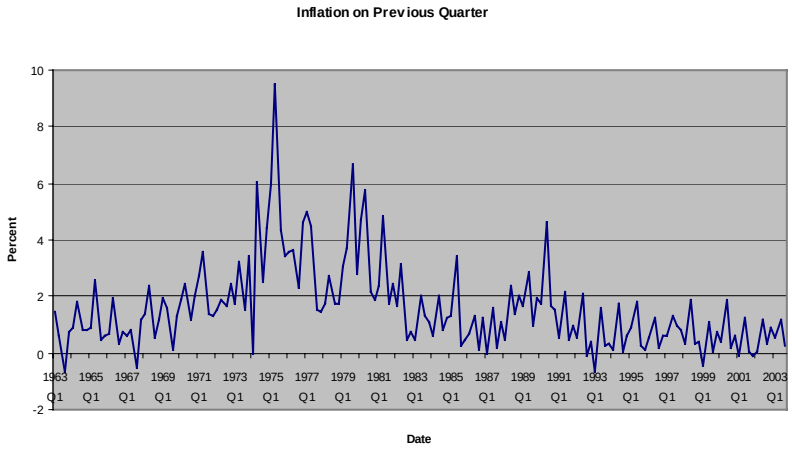
This paper intends to explain how the evolution of inflation over time depends on its previous values. Inflation is known to exhibit time varying volatility, hence traditional regression models, which assume homoscedastistic variance are inadequate. I will attempt to overcome this problem by using an ARCH model. It is instructive to begin by looking at a plot of inflation against time, Figure 1. I used UK data, a series of 168 quarterly CPI observations from 1962 to 2003, found on the UK's National Statistics website.

It is quite apparent from inspection that there are periods in which the variance of inflation is high (the 1970s) and others when it is comparatively low (the late 1990s). Hopefully an ARCH model will help to explain these dynamics.

The Model

I am going to run an autoregression; regressing inflation on its lagged values. Seasonal effects are often encountered in quarterly data, to allow for this I will also include seasonal dummy variables. In this project, I am using dummies to allow only for differential intercepts in each quarter, not for differential slope coefficients

Figure 1: UK Inflation 1963 - 2003



I will commence my analysis by regressing Inflation (Inf) on 5 lagged values of itself and 4 quarterly dummies. Note I do not include an intercept, so as to avoid the dummy variable trap.[1]

$$\pi_t = \alpha_1\pi_{t-1} + \alpha_2\pi_{t-2} + \alpha_3\pi_{t-3} + \alpha_4\pi_{t-4} + \alpha_5\pi_{t-5} + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \mu_t$$

I will first use the traditional Ordinary Least Squares (OLS) method of estimation. The results are given in the table below.

Table 1: UK inflation regression

Regressor	Coefficient	Standard Error	T-Ratio [Prob]
INF(-1)	.46734	.079784	5.8576[.000]
INF(-2)	.28542	.087220	3.2725[.001]
INF(3)	.019862	.090287	.21998[.826]
INF(-4)	.13337	.087221	1.5291[.128]
INF(-5)	-.079349	.079622	-.99657[.321]
Q1	.25114	.21435	1.1716[.243]
Q2	1.3392	.21438	6.2468[.000]
Q3	-.56275	.23954	-2.3493[.020]
Q4	.054633	.21159	.25820[.797]

[1] If I used 4 dummies and an intercept there would be perfect multicollinearity, the data matrix would be non-singular and estimation impossible.

Table 2: Diagnostic Tests

Test Statistics	LM Version	F Version
A:Serial Correlation	CHSQ(4)= 17.2499[.002]	F(4, 150)= 4.4382[.002]
B:Functional Form	CHSQ(1)= .40350[.525]	F(1, 153)= .37969[.539]
C:Normality	CHSQ(2)= 281.6049[.000]	Not applicable
D:Heteroscedasticity	CHSQ(1)= 19.1078[.000]	F(1, 161)= 21.3796[.000]

A:Lagrange multiplier test of residual serial correlation
B:Ramsey's RESET test using the square of the fitted values
C:Based on a test of skewness and kurtosis of residuals
D:Based on the regression of squared residuals on squared fitted values

The model explains the data quite well with an $R^2 = 0.64813$. The F-Test for the joint significance of all the explanatory variables is also quite encouraging an F value of 35.4582 clearly rejecting the null that they have no explanatory power.

Non White Noise Errors

There are clearly quite severe problems with the residuals in this model with the diagnostics pointing towards serial correlation, non-normality and heteroscedasticity. Heteroscedasticity is however an encouraging sign for the presence of ARCH effects as we obviously expect non-constant variance.

With reference to the problems discussed above we must treat the t-statistics quoted with caution. However it is interesting to note that only the first 2 lags of inflation appear to have explanatory power. For the next 3, we fail to reject the null that their coefficients are in fact 0.

Testing Linear Restrictions using Wald Tests

To investigate further the lack of explanatory power of the 3rd, 4th and 5th lags of inflation, I conducted a Wald test with the null hypothesis that, $\alpha_3 = \alpha_4 = \alpha_5 = 0$. The Wald test is especially useful in this case as even in the absence of normally distributed residuals, it is asymptotically Chi Squared distributed. The Wald test returned a value $W = 2.7806$ with an associated p value of 0.427. Hence, we fail to reject H_0, therefore in subsequent versions of the model I only consider the 1st and 2nd lags.

I conducted another Wald test to check the validity of including the seasonal dummies. In this the null hypothesis was that $\beta_1 = \beta_2 = \beta_3 = \beta_4$. This would indicate the absence of an intercept differential between any of the quarters. The

statistic returned was W = 43.1079 with an associated p value of 0.000. Hence we reject H_0 and conclude that we were right to include the seasonal dummies.

Testing for ARCH

The arch model was first proposed by Engle in 1982. He suggested modelling conditional variance as a function of past squared residuals.

$$\text{ARCH(1) model: } h_t^2 = E(\sigma_t^2 | \Omega_{t-1}) = \gamma_0 + \gamma_1 u_{t-1}^2$$

in general an ARCH(p) model includes p lags of squared residuals.

The testing procedure for ARCH is hence to run OLS, as we have above then save the residuals and run an AR(p) regression on their squares. A Lagrange Multiplier (LM) test is then conducted with null hypothesis that the coefficients on each of the p lagged squared residuals is 0. It is shown by Engel 1982 that this test boils down to obtaining R^2 from the AR(p) regression and then testing TR^2 as χ_p^2.

Microfit runs this test automatically see the output below,

Table 3: Autoregressive Conditional Heteroscedasticity Test of Residuals (OLS Case)

```
**************************************************************************
 Dependent variable is INF
 List of the variables in the regression:
 INF(-1)         INF(-2)         INF(-3)         INF(-4)         INF(-5)
 Q1              Q2              Q3              Q4
 163 observations used for estimation from 1963Q2 to 2003Q4

P = 1
**************************************************************************
 Lagrange Multiplier Statistic    CHSQ( 1)=   9.4400[.002]
 F Statistic                   F(  1, 153)=   9.4055[.003]

P = 12
**************************************************************************
Lagrange Multiplier Statistic    CHSQ(12)=  22.0917[.037]
 F Statistic                   F( 12, 142)=   1.8552[.045]
**************************************************************************
```

As evident above we reject the null hypothesis (at the 5% level of significance) of no arch effects for lag lengths of u_t^2, 1 through 12.

However, I was still unsure what lag length to use in my conditional variance function. So, I carried out a regression of u_t^2 on 8 lags of itself to investigate the t-statistics.

Table 4: Ordinary Least Squares Estimation

```
*******************************************************************************
 Dependent variable is USQ
 155 observations used for estimation from 1965Q2 to 2003Q4
*******************************************************************************
```

Regressor	Coefficient	Standard Error	T-Ratio[Prob]	
C		.40073	.21735	1.8437[.067]
USQ(-1)		.20324	.082675	2.4583[.015]
USQ(-2)		.028079	.084211	.33343[.739]
USQ(-3)		.052507	.084203	.62358[.534]
USQ(-4)		.23428	.084148	2.7841[.006]
USQ(-5)		-.054013	.084181	-.64162[.522]
USQ(-6)		.036358	.084196	.43183[.666]
USQ(-7)		.063798	.084197	.75771[.450]
USQ(-8)		-.043619	.082383	-.52946[.597]

```
*******************************************************************************
```

With an $R^2 = 0.12852$

The only significant t-statistics are the 1st and 4th lags of u_t^2 hence it appears that only they should be included in the conditional variance function.

ARCH Estimation

We are now trying to estimate the following model,

$$\pi_t = \alpha_1\pi_{t-1} + \alpha_2\pi_{t-2} + \beta_1Q_1 + \beta_2Q_2 + \beta_3Q_3 + \beta_4Q_4 + \mu_t$$

and we are going to assume the variance of μ_t follows the ARCH process

$$\text{var}(\mu_t) = h_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-4}^2$$

When variances are estimated using an arch process, they will not be independent. This implies that Maximum Likelihood is a more efficient estimation procedure than Least Squares. For a proof of this see Engel 1982.

For MLE it is necessary to make assumptions about the distributions of the stochastic disturbances μ_t. In this case, it is assumed they are normally distributed with mean 0 and variance h_t^2.

Results from Arch estimation

Table 5: GARCH(0,4) assuming a normal distribution converged after 91 iterations

```
********************************************************************************
 Dependent variable is INF
 166 observations used for estimation from 1962Q3 to 2003Q4
********************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 INF(-1)                    .60157             .16043             3.7498[.000]
 INF(-2)                    .27662             .10116             2.7346[.007]
 Q1                        .087743             .19189              .45725[.648]
 Q2                         1.3519             .15607             8.6622[.000]
 Q3                        -.85778             .26458            -3.2420[.001]
 Q4                         .13304             .21543              .61756[.538]
********************************************************************************
```

The estimated model is hence:

$$\pi_t = (0.602)\pi_{t-1} + (0.277)\pi_{t-2} + (0.088)Q_1 + (1.352)Q_2 - (0.858)Q_3 + (0.133)Q_4$$

with conditional variance function:

$$h_t^2 = 0.45 + (0.283)u_{t-1}^2 + (0.110)u_{t-4}^2$$

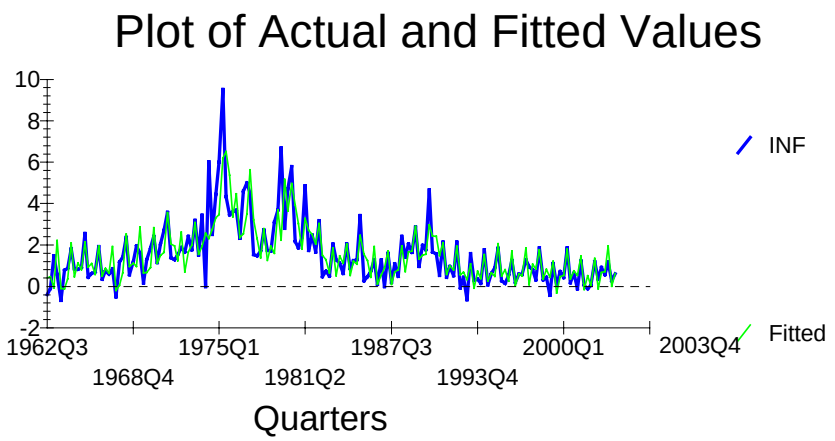
All coefficients are significant at the 1% level, except for those on Q1 and Q4. The 1st and 2nd lags of inflation have a positive impact on a given period's inflation, the first lag having a larger influence. It is also possible to isolate the equation for a given quarter with the dummy variables. For example the estimated equation for quarter 2 is:

$$\pi_t = 1.352 + (0.602)\pi_{t-1} + (0.277)\pi_{t-2}$$

The model has similar explanatory power to the one we estimated under OLS, the value of $R^2 = 0.62917$ see plot below. However, we can have more faith in the estimates as we have addressed the severe heteroscedasticity problem evident in the earlier model.

I now want to find some way to compare the OLS and the MLE ARCH models. I was thinking of comparing the dispersion of residuals, perhaps looking at outliers as in the Engel paper. However, I do not think this would be very instructive as we have a pre-defined the functional form for the variance of the ARCH residuals. Hence their distribution is endogenous to the model and does not give insight to the model's adequacy. So I am going to compare the models in terms of their forecasting ability.

Figure 2: Plot of actual and fitted values



Forecast Comparison

In comparing the models, I am using the version with 2 lags of inflation and the 4 dummies. I then estimate it using the observations from 1962–1999 and use this estimated model to forecast inflation for 2000-2003.

OLS Forecast

Table 6: Dependent variable is INF. 150 observations used for estimation from 1962Q3 to 1999Q4

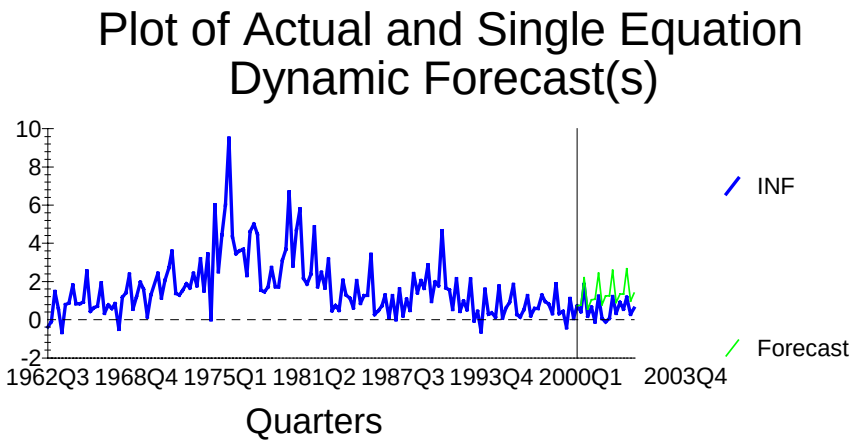
Regressor	Coefficient	Standard Error	T-Ratio[Prob]
INF(-1)	.47034	.078577	5.9858[.000]
INF(-2)	.33125	.078439	4.2230[.000]
Q1	.40141	.17928	2.2391[.027]
Q2	1.5898	.18210	8.7305[.000]
Q3	-.72121	.22320	-3.2313[.002]
Q4	.058426	.23008	.25394[.800]

R-Squared	.62818	R-Bar-Squared	.61527
S.E. of Regression	.96904	F-stat. F(5, 144)	48.6578[.000]

Summary statistics for single equation dynamic forecasts

Based on 16 observations from 2000Q1 to 2003Q4			
Mean Prediction Errors	-.82998	Mean Sum Abs Pred Errors	.82998
Sum Squares Pred Errors	.84766	Root Mean Sumsq Pred Errors	.92069

Figure 3: Plot of actual and single equation dynamic forecast(s)



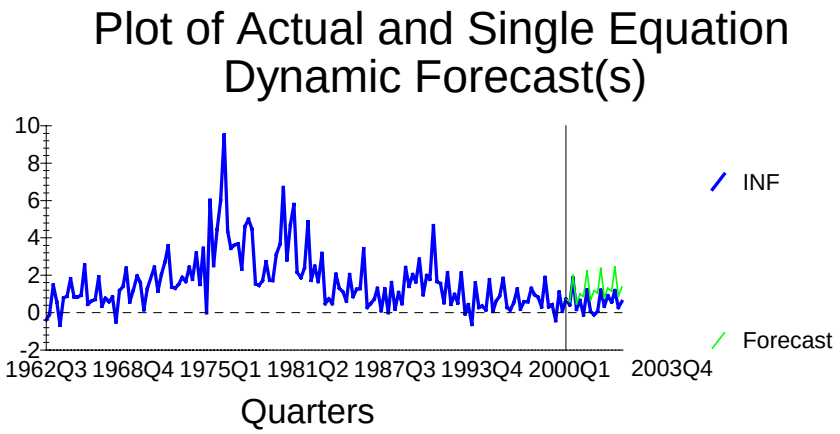
ARCH forecast

Table 7: GARCH(0,4) assuming a normal distribution converged after 146 iterations

```
********************************************************************************
 Dependent variable is INF
 150 observations used for estimation from 1962Q3 to 1999Q4
********************************************************************************
 Regressor             Coefficient       Standard Error         T-Ratio[Prob]
 INF(-1)                    .60962             .13884             4.3907[.000]
 INF(-2)                    .25576             .10418             2.4550[.015]
 Q1                         .14987             .20087             .74610[.457]
 Q2                        1.4290              .17158             8.3286[.000]
 Q3                       -.89312              .25442            -3.5104[.001]
 Q4                         .19358             .24292             .79690[.427]
********************************************************************************
 R-Squared                    .61586   R-Bar-Squared                  .59692

          Summary statistics for single equation dynamic forecasts
********************************************************************************
 Based on 16 observations from 2000Q1 to 2003Q4
 Mean Prediction Errors       -.69473   Mean Sum Abs Pred Errors       .69473
 Sum Squares Pred Errors       .62009   Root Mean Sumsq Pred Errors    .78746
********************************************************************************
```

Figure 4: Plot of actual and single equation dynamic forecast(s)



It can be seen from the summary statistics that the ARCH model is definitely more accurate in forecasting inflation. Whilst both models tend to overpredict inflation, every one of the four summary statistics is in displays smaller prediction error with the ARCH model.

Conclusion

This paper has shown that past values of inflation have a considerable capacity to explain its current value. This paper also outlines the procedure involved in using arch estimation, and shows that it improves the forecasting ability of the model. It has also been seen that seasonal dummies are appropriate for modelling inflation and can be combined to good effect with ARCH. It might have been interesting to try including seasonal dummies in the conditional variance equation to allow for seasonal patterns in volatility. Another interesting expansion of the model would be to use the conditional variances of inflation computed here in a macroeconomic growth model. This would allow investigation of the impact of inflation uncertainty on growth.

Bibliography

Gujarati, D.N. (2003), *Basic Econometrics.* 5th ed. Singapore: McGraw-Hill

Johnston, J & J. DiNardo. (1997), *Econometric Method*, Singapore: McGraw-Hill

Engle, R. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of United Kingdom Inflation' *Econometrica*, 987-1007.

National Statistics Online (2004), Available from: <*http://www.statistics.gov.uk*>

AGENDA 2000 MID-TERM REVIEW DECOUPLING OF SUBSIDIES

BY CIARA COMERFORD

Junior Sophister

Some argue that CAP has been more of a curse than a blessing for the EU economy, and the need for reform is evident. Ciara Comerford examines the changes to CAP proposed under Agenda 2000, focusing specifically on the decoupling of farm subsidies. After investigating in detail what decoupling entails for farmers' incomes, environment and world trade, she concludes that although it will involve adjustment costs for farmers, decoupling is a way forward for European agriculture.

Introduction

I propose commencing this discussion by explaining the reasons for reform, specifically why the Common Agricultural Policy (CAP) was not meeting the needs of farmers and consumers, how it was damaging the environment and distorting world trade. The Luxembourg Agreement outlined key horizontal measures to counteract the disrupting effects of the Current Agricultural Policy. To cut the link between production and direct payments was perhaps the most radical proposal, "the core element of the reforms is decoupling" (Fischler, 2003b). At this juncture, I concede that the confines of this essay would not allow me to explore decoupling and its impact on Europe and the world to the extent that it has consequences. Therefore, I can only hope to elucidate this concept to some extent, to illustrate what it will achieve in relation to farm incomes, in WTO terms, for the environment and in relation to bureaucracy. It is my contention that the proposed measures will have a definite positive impact on what farmers produce, the corresponding impact on our rural environment and will aid in the WTO negotiations.

Background

The Agenda 2000 Berlin Agreement of March 1999, set out EU agriculture policy to 2006 subject to a review operation in 2003. This Mid-Term Review was subsequently brought forward to 2002. As part of the Mid Term-Review, an initial communication from the Commission to the EU Council of Ministers was published

in July 2002. This was followed by detailed legislative proposals for agriculture policy reform in January 2003. Finally on the 26th of June 2003 an agreement was reached by the Council of Ministers in Luxembourg, which outlined a framework for a fully decoupled or partially decoupled single premium area based payment system.[1] Ireland has opted for a fully decoupled system, which hopefully heralds an era of profound and exciting change for farmers, with a move away from subsidy-based production to market driven production. What brought about this reform?

Soon after the advent of the CAP, budgetary issues coupled with the problem of growing surpluses made evident the need for reform. In previous years reform has come in the form of 'milk quotas', 'stabilisers' and 'set-aside'. Market imbalances and the fear that '80% of support was going to 20% of the farmers' led to the MacSharry Plan (agreed in June 1992). Shortly afterwards, the EU was able to conclude the Uruguay Round of world trade negotiations – subjecting agriculture to international trade rules for the first time (Wolfe, 2002). A series of factors including BSE, the discovery of salmonella in eggs, dioxin in chickens and foot and mouth disease increased public unease about food quality and reduced farmers' incomes. Policy makers soon recognised the need to improve environmental practices and the quality of food, while preserving the rural environment[2] (Tracy, 1997).

The Call For Reform …

CAP, viewed by some as a "crazy and distorted waste of resources" (Cadogan, 2003), was no longer suited to consumers or small farmers, it damaged the environment and it hurt the world's poorer economies. The EU would not be able to meet the challenges of enlargement and world trade liberalisation unless it reformed the CAP. It was of paramount importance to show that our policy could provide real value for money and real services to citizens throughout an enlarged EU (The European Commission in Ireland, 2003). The notion that reformation would promote a substantial simplification in the CAP, facilitated the enlargement process and helped to better defend the CAP in the WTO, coupled with the guarantee of income stability for farmers and the removal of the environmentally negative incentives of the current policy that illustrated a dire need for reformation. The Commission decided to move agricultural policy forwards; "indeed to break a series

[1] Due to the confines of this essay, it would be impossible to assess the impact of decoupling on all Member States. Ireland is the only country that has chosen to fully decouple livestock premia. The discussion is in reference to full decoupling and the economic analysis refers specifically to Ireland.

[2] Multifunctionality – a term coined to describe the view that agriculture is about more than just food production.

of anodyne measures that in the end turn out to be a poisoned chalice" (Fischler, 2002).

Achievements under the Mid-Term Review

The Commission outlined a set of proposals to ensure that these goals and objectives were achieved. To cut the link between production and direct payments, linking decoupled payments to environmental, food-safety, animal welfare, health and occupational safety standards (cross-compliance); to increase EU support for rural development by a modulation[3] of direct payments; to introduce a new farm advisory system; new rural development measures to boost quality production, food safety, animal welfare and to cover the costs of the farm advisory system; revisions to the market policy of the CAP, including arable crops, milk and rice (Fischler, 2003b).

> "Decoupling implies that payments will not be linked to production, farmers will receive direct support without having to produce, in return however, they are expected to provide, to satisfactory standards, the public goods demanded by society – the services the market has no way of rewarding" (Fischler, 2002).
>
> "They are also expected to comply with specific requests on environment, food safety and animal welfare, the Commission proposes to employ a farm audit to verify compliance and cut direct income payments to farmers if they fail to fulfil these requirements" (Fischler, 2003c).

Modulation serves the purpose of compensating farmers for the additional services they perform. The goal is to withhold an extra 3% of direct payments every year. An exemption will be granted for the first €5,000 of each direct payment, plus an additional €3,000 for every worker above two on the farm, ultimately, the expectation is that the expansion of the second pillar will result in the transfer of 20% of the funding from direct payments into rural development over seven years (Fischler, 2002). This implies a huge increase in EU funding for rural development. The Commission is adamant that financial support for rural development is not the sole purpose and new proposals have been put in place to improve product quality. These will include certification schemes for products and measures to promote quality products, products with geographical designations of origin and organic products.

[3] Modulation – the Commission's proposal to reduce farmers' direct payments and use these funds for rural development measures, such as environmental programmes, organic farming, promotion of quality or animal welfare (Fischler, 2003c).

Decoupling …

Decoupling means providing support payments to farmers in a way that is not linked to what they produce. The Commission proposes to introduce a single decoupled income payment per farm, that is, from 1st of January 2005, the link between EU subsidy payments and production will be broken. Farmers will receive an annual payment based on the amount of subsidies claimed in the 2000 to 2002 reference period. "To qualify for his/her full level of payments… the farmer must manage the same area as that farmed during the reference years" (Mooney, 2003).

The European Commission predicts that decoupling will help to achieve certain objectives that our previous policy was unable to do; decoupling will stabilise incomes of farmers; give farmers back their entrepreneurial freedom; improve market orientation of agriculture; enable farmers to supply products that consumers want; allow EU support to focus on quality as opposed to quantity; a single payment means that the value of support is constant therefore the farmer is encouraged to farm for the market, not for the subsidy, it also implies a system that is easier to manage, one that is simpler and more transparent; it will allow more money to end up in farmers' pockets than other forms of support that we have (OECD, 2003); it will strengthen the EU's hand in the WTO negotiations, cutting the link between production and subsidy makes direct payments non-trade distorting (The European Commission in Ireland, 2003).

… And farm incomes

Six impact studies (two internal[4] and four external[5]) published by the European Commission concluded that "the reforms envisaged would significantly improve market balances". It is reported that on aggregate farm incomes will improve "by 11% by 2010" (Binfield et al, 2003), however, this statement must be explored to a greater extent. It is predicted that the revenue of beef producers would increase by 4% but "dairy farmers and calf and weanling producers … stand to lose out severely" (O'Rourke, 2003). There is a predicted 30% fall in dairy farming incomes following decoupling and an annual €100 million loss in incomes due to lower calf and weanling prices (O'Rourke, 2003). Dairy, calf and weanling farmers are not the only ones set to loss out when full decoupling is imposed. Joe Walsh insisted that "the reduction in the Irish suckler herd and sheep flock as a result of decoupling would have a devastating effect on the processing industry" (Mooney, 2003). The meat industry warned that the CAP reform proposals would put hundreds

[4] The internal studies have been realised by DG Agriculture.

[5] The external studies have been realised by Food and Agricultural Policy Research Institute (FAPRI), University of Missouri-EuroCARE, University of Bonn Centre for World Food Studies, Amsterdam and the Netherlands Bureau for Economic Policy Analysis, The Hague.

of jobs at risk and result in export losses to the economy of €500 million per annum. Matthews, however, denounces these ideas suggesting that live trade is the major factor affecting the buoyancy of calf and weanling prices, 'in all probability there would be an increase in jobs' in the processing industry. He cites two mitigating factors that contribute to this contention, one, scarcity in mainland Europe will give us a greater opportunity to supply real market needs and two, improved quality will increase value added and a greater journey along the value chain, this implies more consumer cuts and more jobs (Matthews, 2002).

… And the environment

> "Full decoupling will have the greatest positive impact on the environment by the reduction in the contribution by agriculture to green house gas emissions", coupled with the notion that the vast majority of farmers (are critical of the level of bureaucracy) they like to roll up their sleeves and go out on the farm" (Ryan, 2003).

This implies that decoupling has environmental and bureaucratic implications. The environmental benefit arises because decoupling should lead farmers to more extensive production[6] as they will now use inputs only up to the point where their use is justified by the market price return, and not by the combined market price and direct payment, thus reducing pressures on rural resources (Matthews, 2002). Reduced production however, could have a knock-on effect on the processing sector and it could lead to the abandonment of land in some marginal farming areas. The Commission posits that reduced production is a desirable outcome; they believe that more targeted policy instruments, such as less favoured area or agri-environment payments, could meet the threat of land abandonment (Matthews, 2002). Another positive outcome is the prospect of greater funding for REPS on the one hand and a reduction in the opportunity cost of complying with REPS conditions on the other. This will be evidenced by a greater influx of farmers into REPS, if decoupling has the effect of reducing the intensity of agricultural production. Overall, the Commission's emphasis on environmental, food safety and animal welfare standards, as well as rural development, is to be welcomed. If the Commission implements the audit and penalty system that it purposes, the expected results of a cleaner environment and higher quality food are more likely.

The system of decoupling will integrate most existing direct payments a producer receives from various schemes into one single payment, this implies less paper work and red-tape for farmers, whose priority is to work the farm. ICSA have

[6] A recent review of the impact of the MacSharry and Agenda 2000 reforms on extensification commissioned by the UK Ministry for the Environment, Food and Rural Affairs came to a somewhat ambiguous conclusion (Matthews, 2002)

reported that the cost of administration of the various EU schemes to farmers will reduce by 50% (OECD, 2003). It must also be noted however, that since a large proportion of land in Ireland is in conacre (seasonally rented land), there will be a significant administrative burden when the first transfers take place.

… And World Trade

Talks at the 5th WTO Ministerial Conference in Cancun included negotiations regarding investment, competition, trade facilitation, transparency in government procurement and how to free up trade in agriculture (The Economist, 2003). There is a desperate desire for movement toward free trade in agricultural products because markets are currently distorted by more than $200 billion in annual subsidies, a 50% reduction in agricultural trade protection would provide a $27 billion boost to the world economy (Francois, 2001), developing countries could achieve annual welfare gains in excess of $100 billion and trade facilitation could also yield some developing countries gains of up to 5% of their GDP (OECD, 2003). The two key players in the trade negotiations are the US and the EU. In 2001, the Bush administration adopted a farm bill that includes measures that distort trade. The EU, on the other hand, has adopted a policy of trade liberalisation, "precisely in the area of domestic subsidies, decoupling, together with our proposal to scrap the de minimis[7] rule for developed countries, would put us ahead of our great partner and competitor" (Fischler, 2003a). Decoupling should help to reduce the trade-distorting element in EU subsidies, this is because decoupled payments would be eligible for the green box,[8] on the grounds that they do not distort trade because they are not directly connected to production (Wolfe, 2002). These trade-distorting measures should diminish the need for intervention buying and export subsidies, by reducing both production incentives and some guaranteed prices. This would be good for food exporting countries, including developing nations, and export subsidies would allow Europe to capture more markets. Overall, it would aid the conclusion of international trade talks. However, despite initial optimism that a deal could be brokered at the WTO meeting in Cancun, the WTO talks process broke

[7] *De minimis:* Developed countries: provision allows exclusion of product-specific support less than 5% of output value of Aggregate Measures of Support (AMS). Developing countries: provision allows exclusion of product-specific support less than 10% of output value from AMS.

[8] 'Green box' are subsidies that do not distort trade and can be maintained. The amber box covers subsidies that are subject to overall limits and must be reduced because of their impact on trade. The blue box reflects a compromise over how to deal with subsidies that are not as directly linked to production as price support s e.g. subsidies such as the direct payments created under the MacSharry reforms, which were viewed as less trade distorting or temporary.

down on 14th of September 2003, when developing countries refused to accept a WTO proposal that would have allowed for cuts in agriculture subsidies in developed countries only if developing nations agreed to launch new talks on liberalising competition and investment policies. In the end, the developing world is the biggest loser because it had the most to gain. It now appears as though it will take extensive political effort to put the talks back on the rails. If trade talks are not resumed by spring 2004, "they are likely to drift rudderless until well after the new American administration and the EU Commission take office in 2005" (The Economist, 2003). The expansion of the EU will further complicate the decision-making process. The Commission's proposals should reduce the cost of expanding the CAP to cover more countries. In the longer term, farmers in the new member-states would be discouraged from overproducing, in addition, by lowering direct payments in the EU-15, modulation would allow new member states to attain the same level of this form of subsidy as existing members in a shorter period of time. The EU initially argued that direct payments should not be extended to new members, as their farmers did not need compensation for price cuts. However, under pressure from the applicant countries for immediate access to direct payments, the Commission relented and proposed phasing in these subsidies over a ten-year transitional period. However, applicant countries are likely to be hostile to the Commission's call for the direct payments to be phased in slowly.

Conclusion

Should Irish farmers, consumers, taxpayers, environmentalists and free trade advocates applaud the Commission's move towards the decoupled single farm payment? It is my contention that we should embrace this change, however, I do concede to being sceptical at the outset. There are particular issues, which I am concerned about, if decoupling has the desired effect of reducing output there could be a major contraction in agricultural activity leading to a decline in rural economic activity; could the countryside ideal, where man toils the land leading to a thriving rural economy and community disappear forever. It is also reported that there could be a significant reduction in suckler cows across the EU "with the largest proportionate decline expected to occur in Ireland", thus having adverse effects on meat plants, service industries and livestock marts (Binefield et al, 2003). Basing direct payments on past entitlements is not cutting the cord from subsidies in the way that we should, "it reduces the scope for change and prevents it from setting a long-term strategic view about the future shape of the direct payments system" (Ibid). I come from a dairy farm background thus certain issues are extremely prevalent, I am concerned about the threat of a 30% reduction in incomes and the

abolishment of milk quotas come 2014. Payments were proposed to offset the extent of the decline in milk prices, but these payments are smaller than suggested in the January 2003 legislative proposals. The incentive for marginal producers to exit the sector are reduced, thus making it difficult for competent producers to expand, be competitive and produce what the market wants using economies of scale.

However, this issue cannot be viewed with bias towards a particular sector, there is too much at stake and in review, I can see that the benefits far outweigh the cons. Farmers now have the ability to produce what the market wants, to act as entrepreneurs and to engage in competition. Consumers and taxpayers will not be contributing towards food mountains and milk lakes and they will have a greater guarantee of purchasing quality produce. Due to the horizontal measures proposed, including modulation and cross-compliance, farmers will be encouraged to develop, attain and sustain a healthy rural environment. Unfortunately, the world trade talks collapsed in September, but I feel that Europe made a concerted effort to reach a compromise that will promote trade with the developing world.

Decoupling heralds an era of profound and exciting change for farmers with a move away from subsidy-based production to market driven production. This proposal should be embraced as we are moving out of an era of trade distorting policies, over productive techniques and poor quality produce.

Bibliography

Binfield, J et al. (2003), "The Luxembourg CAP Reform Agreement: Implications for EU and Irish Agriculture" [Online.] *FAPRI Ireland*. Available from: <www.tnet.teagasc.ie/fapri>

Cadogan, S. (2003), "A dispassionate view of farming." *Irish Examiner*, 02/10/2003.

Fischler, F. (2002), *Towards sustainable farming – presentation of the CAP Mid-Term Review.* Brussels: European Parliament Speech. 10/07/02.

Fischler, F. (2003a), *CAP Reform: A long-term perspective for sustainable agriculture.* Brussels: COMAGRI Committee for Agriculture.

Fischler, F. (2003b), *Commission proposal will have positive effect on farm incomes, studies say.* Brussels: European Commission Speech. 15/01/2003.

Fischler, F. (2003c), *EU fundamentally reforms its farm policy to accomplish sustainable farming in Europe*. Brussels: European Commission.

Francois, J. (2001), *The Next WTO Round: North-South Stakes in New Market Access Negotiations.* Adelaide: Centre for International Economic Studies and Rotterdam: Tinbergen.

Matthews, A. (2002), "Decoupling: environmental implications." [Online]. *Irish Argicultural and Food Development Society*. Available from: <www.teagasc.ie>

Mooney, P. (2003), "Decoupling would cut suckler numbers 30%". [Online]. *Irish Farmers Journal Online*. Available from: <www.farmersjournal.ie>

OECD (2003), "Cancun Trade Talk Breakdown: What Now?" [Online]. *OECD*. Available from: <www.oecd.org>

O'Rourke, P. (2003), *Irish Farmers Journal*. Vol. 56 (43).

Ryan, R. (2003), "Consumers hold key to future of farming." *Irish Examiner*. 28/10/2003.

The Economist (2003), "Cancun's charming outcome." *The Economist.* Vol. 368 (8342).

The European Commission in Ireland (2003), "The Commission CAP Reform Proposals" [Online]. *The European Commission Ireland*. Available from: <www.euireland.ie>

Tracy, M. (1997), *Agricultural Policy in the EU and other Market Economies.* Belgium: La Hutte.

Wolf, J. (2002), *The future of European agriculture*. London: Centre for European Reform.

EU FISCAL POLICY AND THE STABILITY AND GROWTH PACT: NECESSARY EVIL OR POLICY MAKER'S NEMESIS?

BY MICHELLE DALTON

Junior Sophister

With monetary policy being controlled by ECB, fiscal policy is especially important for the EU countries. However, the extent of fiscal stimuli has been limited somewhat by the criteria laid down by the Stability and Growth Pact. While fiscal stability helps promote macroeconomic stability, one of the three pillars of the New Consensus policy, stringent limits on the level of external debt can be detrimental in certain situations. In this essay, Michelle Dalton critically examines the rationale behind the Pact as well as its merits and limitations, and suggests possible enhancing modifications.

Introduction

The role of fiscal policy within the EU has changed considerably over the past few decades. Damaged post-war economies and the subsequent oil crisis in the early 1970s demanded a thoroughly Keynesian response centred on expansionary fiscal policy to try and initiate economic recovery. However, the Maastricht criteria for EMU membership heralded a period of sharp fiscal consolidation on the part of member states in order to shape up for the arrival of the Euro. It is this 'new consensus' paradigm that the EU hopes to continue through the SGP. However, while some may view it as a necessary evil to constrain a country's spending capacity for the greater good, the issue of whether the agreement is sufficiently flexible to cope with fluctuations is difficult to get away from. In this essay, firstly a brief overview of the potency of fiscal policy is offered. Next the issues facing the role of fiscal policy in Europe are highlighted. Then the central components of the pact are outlined, and the necessity of the SGP is examined. Finally, the flexibility problem is addressed, as are possible alternatives, which may provide fiscal discipline without the same level of constraint.

The Role and Relevance of Fiscal Policy

Ever since Malthus' seminal elucidation of 'general gluts', or in today's parlance economic slumps, government expenditure has been used as a counter-cyclical tool to stimulate flagging economies. However, fiscal policy is by no means a panacea. There are serious difficulties in implementing a successful package, and just like any tool at a government's disposal, it is open to abuse. The Lucas Critique offers a warning that demand management could prove ineffective. The hypothesis of rational expectations dictates that individuals may revise their expectations when a new policy is implemented and so the effects may be very difficult to predict. In addition there is the problem of Ricardian Equivalence, whereby households and firms may believe that an increase in government spending will lead to tax increases in the future and so a fiscal expansion will not have the desired stimulatory effect. Furthermore, the crowding out problem of private spending and investment being forced out by higher interest rates is also an issue, as are what Friedman terms recognition lags, decision lags and effectiveness lags. But perhaps the most dangerous consequence of all, is that of political cycles, whereby governments reduce taxes or increase spending, even when it is clearly the most economically unsuitable thing to do, simply for political reasons.

What is perhaps most necessary in Europe's fiscal policy approach is balance. It intrinsically has two central functions. The role it assumes must be adequate enough to allow individual member states to compensate for the 'one size fits all' monetary policy. However, its approach must also seek to minimise the dangers of fiscal policy as outlined above, with particular regard to the issue of sustainability of government spending. The dangers of debt accumulation are alluded to by Miles and Scott (2002), who note that in OECD countries debt levels have risen sharply over time. In 1960, in developed countries government spending was around 30% of GDP. Now the figure is more like 50-60%. Because of this trend, which is particularly acute in Europe, interest payments have increased on average threefold.

The problem of the ageing European population also has serious repercussions for the EU's fiscal position. In many member states there is a predicted doubling of dependency ratios by 2040, which will lead to an escalation of pension expenditure as more people retire. Thus, current fiscal policy needs to make preparations to enable members to cope with this future burden.

The ECB (2000) also contends that: "fiscal sustainability will be a key condition for monetary stability in the future". Therefore, to ensure the success of the common monetary policy, it is clear that some sort of co-ordinated fiscal approach is required. Through the peer appraisal approach of the BEPG, the EU can keep an eye on fiscal issues, but a narrower framework is also needed. Intrinsically

the role of fiscal policy in Europe is very much a compromise between allowing individual states to have autonomy regarding fiscal matters, and the need for co-ordinated economic stability throughout the EU.

The Need for the SGP? – Defining the Role of Fiscal Policy

There are essentially two main elements to the EU Stability and Growth Pact. Firstly, countries are required to keep their current accounts close to balance or in surplus over the medium term. Secondly, Article III-76 states that no 'excessive deficits' are tolerated in the short run (i.e. on an annual basis), where excessive is deemed to be a negative balance greater than 3%. However, there are some safety valves inherent in the plan, namely that in the case of a steep recession where GDP falls by more than 2% in a year, the conditions of the pact can be broken. Similarly if the fall in output is in excess of 0.75% (but not greater than 2%), a country might be granted a reprieve.

With the necessary trimming of fiscal indulgence in order to meet the Maastricht Treaty criteria now fading in the minds of Finance Ministers, a German-led initiative believed that some countries may now let their fiscal stance loosen, causing severe problems for other member states in the Euro area. Thus in 1997, we had the birth of the SGP. And while the aim of the continuation of the fiscal prudence that dominated the 1990s was a central *raison d'être* for the pact, it was not the only one. By the very nature of EMU, member states had to relinquish control over monetary policy. As a result fiscal policy has grown in stature, becoming the only tool that national governments can use when faced with difficult economic circumstances. As Willett (1999) asserts, the 'one size fits all' approach of EMU, has meant that the idea of members surrendering their fiscal autonomy to the EU as well, was definitely a step too far in European economic integration. Politics aside, there are vital economic principles at stake also; a low ECB-set interest rate may not suit the economic circumstances in all member states. This, coupled with the concept of subsidiarity, is essentially the reason why fiscal policy remains at the national level. However, given the economic circumstances within the EU, there is certainly a need for significant coordination of some sort. The SGP therefore dictates the parameters through which fiscal policy can be effectual, and ultimately curtails its role as a stabilising force in the Euro area.

The rationale for fiscal coordination also stems from the existence of international spillovers; excessive deficits may damage the entire Euro area and not just the individual country in question. The threat of moral hazard also exists, as Beetsman (1999) mentions that some members may be less inclined to restrain debt in the knowledge that although it is strictly condemned by the EU, other member

states will be forced to bail them out, as this is a far less painful solution than the possible collapse of the EU banking system (which could arise as a result of a run on a country's debt). Furthermore, as the Mundell-Fleming model shows that the role of fiscal policy is heightened in a fixed exchange rate system (such as EMU), fiscal policy may be a more tempting option and one which governments may abuse, especially in the year preceding elections.

So the question remains, can countries afford to keep running deficits when they already have accumulated a wealth of debt? The answer is clearly no, and consequently fiscal policy in the Euro zone, through the SGP, recognises this role.

The Flexibility of the SGP

As previously stated, fiscal policy is not a cure for all ills, and indeed if it is used excessively the results can be distinctly non-Keynesian.[1] Therefore, perhaps the SGP has a valid point: If countries are going to run up relatively large deficits regularly, fiscal expansion will not have the desired effect, and it may become relatively ineffectual. The SGP will prevent this happening and it will allow some scope, albeit limited, for cushioning blows. If a country's structural budget is balanced, the SGP parameters will allow for the effects of automatic stabilisers, however is this scope enough?

De Grauwe (2003) claims the SGP "risks putting the clock back fifty years" by obtruding countries from using the built-in stabilising properties of government budgets to smooth out deep fluctuations. Consequently it is difficult to see the SGP as sufficiently flexible, and indeed Germany and France seem to agree. Both countries have breeched the pact in successive years and are now in danger of being fined anything up to 0.5% of GDP. However, it is difficult to accept the credibility of the pact when even the European Commission President Romano Prodi pronounced it as being "stupid". The Economist (2003) believes that the European Commission will allow France to break the limit in 2004, as long as they promise to begin to cut back in 2005. This however weakens the credibility of the pact. If certain countries can get away with breaking the ceilings, how are we to believe that the SGP is a serious agreement, which must be adhered to by all member states? It is certainly interesting to note that Germany, who was so active in initiating the pact, can now see its shortcomings.

[1] Hogan (2001) contends that after prolonged use fiscal policy no longer has a stimulatory effect on output as budget deficits lead to a fall in confidence levels and people anticipate increases in taxation.

Apart from the problems concerning the stabilisation function of fiscal policy, there are further implications of the pact. As De Grauwe contends (1997): "As countries will be hindered in their desire to use the automatic stabilisation in their budgets during recessions, they will increase their pressure on the ECB to relax monetary policy." He continues to say that a central goal of the pact's genesis was in fact to reduce this very threat, but in reality the agreement may actually "have increased the risk of such pressure" (Ibid).

Price stability seems to be the central tenet by which the ECB operates, and indeed if countries consistently run deficits, it will put great upward pressure on prices, reduce confidence in the economy, and possibly lead to an inflationary spiral. However, as we have seen since the inception of the pact, in Germany particularly, deflation can also be a problem. In this case a fiscal injection in excess of the guidelines in the SGP could prove helpful in trying to avoid the plague of deflation. When the economic climate is favourable the SGP seems to be rational and appropriate. If Europe is greeted with a stronger recession than the current situation, the pact will probably lead to an overly painful recovery. The pact seems rather myopic in failing to acknowledge this fact.

Furthermore, the aforementioned ageing population problem will eat heavily into public funds through increased pensions costs. This threat of a burgeoning debt does certainly need to be tempered; and at least if governments pursue a prudent policy of fiscal correction over the coming years (what the ECB terms "pre-emptive" action) the Debt/GDP ratio should still have some room to absorb extra borrowing as the situation worsens after 2020. However, as the problem worsens over the coming years, keeping budgets close to balance or in surplus may simply not be a tenable option for governments.

De Grauwe (2003) continues to underline the unsuitability of the SGP for every member state. He argues that even if the SGP guidelines make sense as a "temporary strategy for highly indebted countries like Belgium, Italy and Greece", countries that do not have a debt problem, should not be made to adhere strictly to the plan.

Moreover, Hughes Hallett and McAdam (1999) also argue that the SGP may have long term effects which are even more far reaching than mere inflexibility:

> "Not only does constraining our ability to vary the policy mix constrain our ability to absorb shocks; it also destroys our ability to get precisely those longer term benefits (i.e. more investment and higher output potential) which these deficit reductions were designed to achieve" (Ibid).

By discouraging borrowing, governments may abandon very worthwhile investment projects which could, if initiated, significantly improve long term economic growth. The question of an appropriate debt level can surely only be

meaningful if we ask what the debt is used for. Just as firms borrow money when they see a profitable opportunity, member states should also have the capacity to build up debt if it is self-liquidating over a future period. It does appear that, as Canzaneri and Diba (2001) conclude, the SGP is indeed an "albatross" rather than a "delicate balance".

But while we are criticising the SGP for being too rigid, it is interesting to note that fiscal policy itself is quite inflexible also; the problem of time lags means that even if countries were to have total control over fiscal policy, it may not be a very effective weapon, *vis-à-vis* monetary policy, if faced with a major recession. In addition, there are serious sustainability questions attached to the use of fiscal policy, and at the very least we must praise the pact for recognising this. In fact the International Monetary Fund (IMF) (1997) raises this point, by noting that the SGP may indeed constrain fiscal activism, but in a sense a history of high budget deficits and prolonged government spending has already done so:

> "Concerns about the potentially constraining effects on countries' ability to pursue counter-cyclical fiscal policies need to be put into the perspective of the constraints imposed by large deficits in most EU countries over much of the past 25 years".

Alternatives and Improvements

Several modifications could be introduced to the existing pact to make it more flexible. There is the option of using the Debt/GDP ratio as a yardstick, however Beetsman (1999) suggests that the Deficit/GDP measure is a much more immediate check, and it helps to stop current governments being punished for the negligence of previous governments. On the other hand, it would allow governments to use fiscal policy as they see fit, given the economic circumstances, while still underlining the need to pay back any loans.

He also contends that the reference deficit level (which is essentially arbitrarily set at 3%) should be contingent on the state of the economic climate. A more generous deficit should be allowed when the deviation of actual from potential GDP is bigger and vice versa. However, measuring the accuracy potential GDP with any degree of certitude is at best problematic, at worst impossible.

We could potentially go a step further however. Fischer and Giudice (2001) concur that perhaps the actual size of the fiscal correction is not as important as the composition of expenditure. Should the EU be emphasising a shift away from current transfers and public wages expenses, rather than outright spending control? They maintain that it is the 'quality' of public finances that really should be kept in check, and perhaps the EU should address this dimension of the fiscal policy debate.

Conclusion

As we have seen, EU fiscal policy as embodied by the SGP is far from a Utopian solution. While many of the problems that pervade economic recovery in Europe may indeed be structural in nature, the option of using expansionary fiscal policy as a short run cushion neither should not nor cannot be dismissed altogether. For many countries the pact will act as a fiscal straitjacket that may prove impossible to adhere to. Yet, even in its present incarnation, the SGP should at the very least ensure that its ethos is applied consistently and fairly in order to uphold the credibility of the pact, and indeed the credibility of European economic policy in general. The success of the pact in its current form will only be seen in the fullness of time, and optimistic economic forecasts for late 2004/2005 may help to extend the longevity of the agreement. However, perhaps when faced with the onslaught of something more pronounced than a mere tumble in short-run demand, we will see some serious and altogether necessary amendments to its restrictive essence.

Bibliography

Beetsman, Roel. (2001), "Does EMU Need a Stability Pact?" In: Brunila, A. Buti, M. and Franco, D. (eds.) *The Stability and Growth Pact.* UK: Palgrave.

Canzaneri, M. and Diba, B. (2001), 'The SGP: Delicate Balance or Albatross?' In: Brunila, A. Buti, M. and Franco, D. (eds.) *The Stability and Growth Pact.* UK: Palgrave.

De Grauwe, P. (1997), *The Economics of Monetary Integration*. UK: OUP.

De Grauwe, P. (2003), "The Stability and Growth Pact in Need of Reform URL.":
Population Ageing and Fiscal Policy in the Euro Area. ECB Monthly Bulletin, July 2000.

The Economist (July 17 2003) *Chirac Storms the Stability Pact.* The Economist. 17/07/03.

Fischer, J. and Giudice, G. (2001) "The Stability and Convergence Programmes" In: Brunila, A. Buti, M. and Franco, D. (eds.) *The Stability and Growth Pact.* UK: Palgrave.

Hogan, Vincent. (2001), *Expansionary Fiscal Contractions*. UCD CER Working Paper. Available from: <http://www.ucd.ie/economic/staff/vhogan/panel11.pdf>

Hughes Hallett, A. and McAdam, P. (1999), "Implications of the SGP" In: Hughes Hallett, Hutchinson and Hongaard Jenson (eds.) *Fiscal Aspects of European Monetary Integration,* UK: Cambridge University Press.

IMF (1997), "EMU and the World Economy" In: *World Economic Outlook*. Washington DC: IMF.

Miles, D. and Scott, A (2002), *Macroeconomics – Understanding the Wealth of Nations*. Wiley: New York.

Willett, Thomas (1999), "A Political Economy Analysis of the Maastricht and Stability Pact Fiscal Criteria" In: Hughes Hallett, Hutchinson and Hongaard Jenson (eds.) *Fiscal Aspects of European Monetary Integration.* Cambridge University UK: Press.

DIVERGENT INFLATION RATES BETWEEN MEMBERS OF THE EURO AREA: CAUSES, IMPLICATIONS AND SUSTAINABILITY

BY MARK-RÜDIGER METZE

Junior Sophister

With monetary policy in the hands of ECB, one might expect to observe similar inflation rates among the EU member states. However, it is clearly not the case. This essay by Mark Metze identifies and explores the causes of inflation divergences within the EU, concluding that convergence of inflation rates is unlikely in the nearest future.

Introduction

Inflation[1] differentials within the Eurozone are an omnipresent phenomenon, one that has attracted widespread attention amongst European business and the general public. Above average inflation can lead to the reallocation of investment, bubbles in the property market as currently in Ireland, or a lasting nationwide consumption strike (as seen in Germany) raising unemployment and leaving the country at the verge of deflation in late 2002.

All of the above are short-run developments, which have occurred since the introduction of the Euro in 1999 and are likely to be dampened by the activation of automatic adjustment mechanisms such as lost competitiveness, an appreciation of real effective exchange rates, and drops in aggregate demand (Mortimer-Lee, 1998). Resulting in negative 'equity' situations, reductions/restraints in wage growth or even current account deficits inducing falls in government spending and wages, these developments then correct for national differentials.

The aforementioned short-run implications do, however, leave us alert to the fact that the European type of monetary union spans very differently developed

[1] Inflation is defined as "the persistent rise in the general level the of money prices" while price stability may exist at inflation rates in the range of 0-2 per cent. The Causes of this are composition (choice of items in basket), quality (qualitative improvement of items in basket) and substitution (consumers' purchase of cheaper substitutes than items in basket) biases in the CPI (McAleese, 1997) .

and regulated economies. The operation of almost entirely decentralised[2] fiscal policies and the extremely slow nature of effect characteristic of the above mechanisms of adjustment may bear significant hazards for the future of the currency system and its economies.

In the course of this essay we will take a look at the extent of inflation differentials across the European Monetary Union (EMU), discuss the most common causes of inflation differentials in a monetary union and compare our findings to the European Central Bank's (ECB) analyses of the problem in 1999 and 2003.

Part I: Preliminaries

Empirical Evidence

As pointed out numerously, and somewhat visible from Figure 1 below, the two decades preceding the establishment of the EMU were marked by a significant convergence of inflation rates across the later union states[3] starting from a ten percentage points difference between the highest and lowest national HICP[4] increase in 1980 and arriving at half a percentage point difference in 1997 subsequently increasing back to a two point difference in 1999 (ECB, 1999).

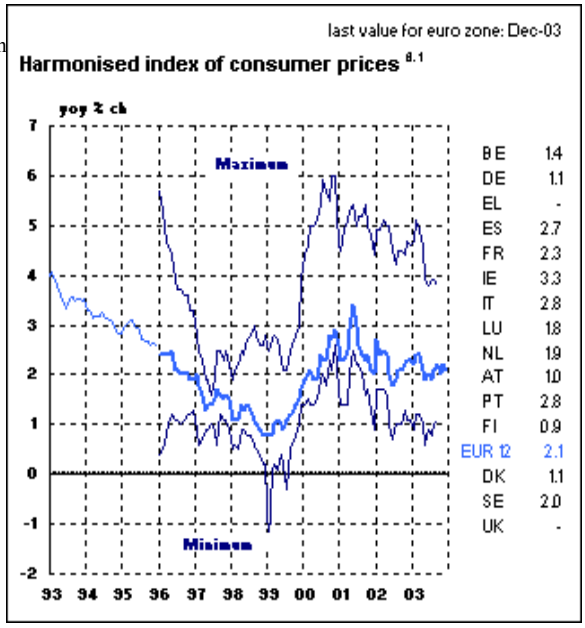
It is, however, also evident from Figure 1 that this convergence did not last and seems to have developed into a significant divergence of up to four percent in the aftermath of Stage Three of Economic and Monetary Union.

[2] The EU's purely re-distributive fiscal strength currently lies at a cap of 1.27% of EU GDP, 1.03% of which were used in the Budget of 2003 and six countries currently opting for a cap of 1% over the next budgetary cycle post 2007.

[3] For example Mortimer-Lee, 1998; ECB, 1999

[4] Using the Harmonised Index of Consumer Prices as a comparison of inflation rates, one should be aware of the fact that weights used in the construction of national HICP components differ taking account of dissimilar patterns of consumption. This could mechanically generate measured differentials across countries even if inflation rates for individual goods were equal (ECB, 37).

Figure 1: Harmon



General Causes of Inflation

Unspecific to monetary unions, a core of classic and strong explanations of inflation prevails. In assessing the causes of inflation differentials, we shall begin by emphasising a simple but very significant distinction: The causes of inflation - amongst which we also find structural factors - should not be confused with structural variables which cause differences in the transmission of the common monetary policy.[5]

Excessive growth of the money supply is clearly the most obvious cause of inflation. Money supply growth will result in inflation whenever it persistently exceeds the growth in real output, *ceteris paribus* (McAleese, 1997). While in a monetary union one may be tempted to expect money supply growth to be a symmetrical phenomenon, given the application of the same discount rate to all

[5] The latter only intensify or weaken inflationary or counter-inflationary impulses rather than cause these. While the former will only be significant and direct causes of inflation 'differentials' if they occur asymmetrically throughout union states, the catalyst mechanisms cause nothing but differentials.

banks throughout the union, we will later see how financial intermediaries[6] and different national propensities to use organised markets[7] cause differences in money supply growth throughout the Union.

Other common causes of inflation are demand and supply-side shocks to the economy, on the other hand, demand-side shocks can originate from bullish business and/or consumer confidence in the future of the private sector, monetary and fiscal impulses or changes in export demand due to increased competitiveness or booming trading-partner economies (Mortimer-Lee, 1998). On the other hand, inflation can originate from a 'demand-pull' process that describes the (almost) inevitable monetary expansion following from excessive government deficits (McAleese, 1997). The fate of the Stability and Growth Pact will determine whether such shocks might possibly cause inflation differentials in the future. Currently, however, the minimal French and German breeches cannot be attributed to having induced such effects.

Supply-side shocks occur if key materials, resources or labour become short. Such conditions tend to persistently raise costs ahead of the level of productivity, thereby resulting in what has been termed cost-push inflation (McAleese, 1997).

Finally, of course, rises in productivity will lead to wage increases over time, thereby raising prices and causing inflation. Inflation will also be created where wage increases are negotiated by strong labour unions that are successful in raising wages by more than productivity justifies.[8]

Part II: Causes of Divergent Inflation Rates in the EMU

Cyclical Factors

Different Stages of the Economic Cycle

Commenting on inflation differentials of only 2%, the ECB in 1999 interpreted that evidence suggested some dispersion of cross-country cyclical

[6] Financial intermediaries - banks and building societies that provide a channel for the transmission of funds between borrowers and lenders. The liabilities so created are non-marketable (Howell and Bain, 2000).

[7] Organised markets, such as stock exchanges, provide the legal and practical framework for the sale and purchase of tradable liabilities (Howell and Bain, 2000).

[8] See McAleese 1997 Such mechanisms are also amplified by the fact that economies consist of sectors which see little if any productivity increases (e.g. schools) and other sectors, which drive economic growth. Wage increases, however, are necessary in both sectors or employment would successively fall in the former and rise in the latter.

positions within EMU generating inflation differentials in the short term. It explained the bulk of such effects would come via impacts on non-traded goods that, in the short-run, depend on domestic rather than external demand (ECB, 1999).

To explain these findings briefly: Different cyclical positions in the monetary union refer most of all to demand-shocks that some countries experience relative to others. Given that some countries experience booms while others experience downturns or even recessions, there are two theoretical possibilities: the shocks are offsetting and thus perfectly asymmetrical, or less than offsetting and thus imperfectly asymmetrical. If shocks were perfectly symmetrical, stabilisation policy of the ECB[9] could be perfectly fair. But to the extent that shocks are asymmetric, the ECB stabilises too little from the respective points of view of the individual member states.[10]

Optimum Currency Area Theory (Mundell, 1961) posits that due to the impossibility of devaluing their currency in a monetary union, single fast-growing economies must pursue deflationary policies that in turn constrain their growth processes (De Grauwe, 2000). This view, however, has come under criticism and it has been pointed out by Krugman (1989) that due to the nature of income elasticities, characteristic of and faced by fast growing countries relative to slow growing countries,[11] the former can continue to grow faster without incurring trade balance problems. Should, however, current account deficits occur, then the higher productivity of capital in faster growing economies attracting investment flows from slow growers makes it possible to finance these (De Grauwe, 2000).

In deriving these conclusions, Krugman's criticism points at the possibility that the automatic adjustment mechanisms, lose their functions in constraining fast-growers due to their characteristic advantages in production and capital attraction. Given the fact that since the start of Stage Three of EMU, asymmetric shocks can no longer be corrected by changes in monetary policy or the exchange rate, such conditions would thus lead to persistent differentials in national inflation rates and

[9] Note that the only function the ECB is willing to accept is the maintenance of price stability 0-2% inflation. It can thus be described as a 'hard-nosed' monetary authority, indifferent about the levels of unemployment.

[10] This theoretical framework is an extension of optimal currency area theory by De Grauwe (2000), one that would evidently suggest countries with significant differences in cyclical positions to refrain from joining a monetary union.

[11] Economic growth implies amelioration of existing products and the development of new products leading to higher income elasticities of exports from fast growing countries relative to income elasticities of slow growing partners. Faster growing countries also tend to have greater income elasticities of exports relative to their income elasticities of imports (De Grauwe, 2000).

possibly to inflationary spirals in those countries for which current interest rates are already too low today.[12]

We should, however, note the ECB's conviction that cyclical movements in the Euro area have become more synchronised over time (ECB, 1999) and, while undoubtedly relevant, do not seem sufficient to explain the observed levels of differentials (ECB, 2003).

The Balassa-Samuelson Effect and the convergence of living standards

As we will see in later parts of this essay (sector structure) there are a number of reasons for which member economies might experience asymmetrical shocks. Clearly, one factor that stands in close conjunction with the hikes of the business cycle, is productivity growth. Growth in productivity, however, can also be the product of a catch-up phase of less-developed countries in a monetary union and thus occur in a longer-lasting and more consistent fashion than commonly resulting from the economic cycle.

The Balassa-Samuelson effect relates inflation differentials between countries in a monetary union to differentials in their productivity growth and can therefore be applied to explain inflation differentials stemming from both cyclical differences as well as catch-up phases. Distinguishing between traded and non-traded (assumed to be wage costs only) goods, the model postulates that competition between monetary union countries assures that price changes of tradable goods as opposed to non-traded goods are equalised. Since differentials in wage changes reflect differences in productivity growth[13] the latter must cause inflationary differentials, if it differs across the union (De Grauwe, 2000).

[12] Mundell emphasises the importance of either relative price and wage flexibility or sufficient mobility of labour in combination with a sufficiently centralised budget considering the reduced efficacy of monetary and exchange rate policy due to monetary union (De Grauwe, 2000). If we accept Krugman's criticism, we are prompted by the realisation that for some fast growers in the EMU neither of these conditions for an optimal currency may apply.

[13] A rise in productivity in the traded goods sector will tend to drive up wages in this sector, but since this increase in wages is matched by increased productivity, it will not give rise to higher traded goods prices. But since labour is assumed to be mobile across sectors, firms in the non-traded goods sector will have no option but to offer higher wages in order to retain their workers. In the non-traded goods sector the increase in wages will not be matched by a productivity increase, thereby raising costs. This increase in costs will lead to an increase in prices in the non-traded sector. (ECB, 1999)

Table 1. Implied inflation differentials to the euro area average due to the BS effect compared with actual HICP inflation differentials between 1995 and 2002

	DE	FR	NL	AT	FI	ES	IT	BE	PT	IE	GR
Estimated BS inflation differential	-0.6	0.1	0.1	0.2	0.5	0.5	0.5	0.6	0.7	1.3	1.6
Actual HICP inflation differential 1995-2002	-0.7	-0.4	0.6	-0.4	-0.3	1.1	0.9	-0.2	1.1	1.2	1.9

Source: ECB

The ECB in both 1999 and 2003 emphasised the importance of the Balassa-Samuelson effect in explaining inflation rate differentials but attaches greater weight to the catch-up nature of productivity, income and price level convergence as opposed to purely cyclical causes in its 2003 report (ECB, 2003). With regard to the implications of catching-up trends, the ECB emphasises its concern to avoid inflation synergies between such 'convergers' and normal performers. From the comparison of Balassa-Samuelson estimates of differentials with actual differentials between 1995 and 2002, it emerges that the effect is most significant for the EMU's formerly least developed countries and interestingly, Germany. We may, however, question whether Ireland's contribution to HICP inflation differentials can any longer be accounted for as the result of convergence. Entirely different to Spain, Portugal and Greece the country's price level currently stands at 12 percentage points above the EU average (ECB, , 2003), i.e. it may still be catching up in terms of productivity (B-S effect) and standard of living, but this process seems at last to have decoupled from convergence in the general price level. If Ireland continues to experience inflationary differentials above the EMU average, Krugman's criticism of Optimal Currency Area Theory and its implications would be a particularly good explanation of this unparalleled performance.

Institutional and Sector Asymmetries

Sector Structure

As one may expect from the principle of comparative advantage, different regions and different countries engage in the production of very different goods and services. If this is the case, asymmetric (country-specific) shocks are more likely to

occur and symmetric shocks across the union are more likely to have asymmetric effects.[14]

There exists, however, no agreement as to whether or not trade integration under the Single European act and furthered by EMU will lead to a converging structure in production resulting in the convergence of sector patterns, increased intra-industry trade[15] and shocks being predominantly symmetric resulting in symmetric effects. While these consequences of trade integration are predicted by, amongst others, the EU Commission, a more 'pessimistic view' is proposed by Krugman (1991). Contrary to the Commissions view, his analysis emphasised that trade integration will lead to the concentration of production (so as to allow for the reaping of advantages from economies of scale) resulting in sector-specific shocks becoming country-specific (De Grauwe, 2000). While this agglomeration effect is blind to national borders[16] and studies like that by Freudenberg et al (1995) have presented evidence of substantial increases in intra-industry trade in the EU between 1980 and 1994, both views are likely to remain significant.[17]

This is also proposed by Dornbusch et al (1998). They examine the transmission of a tightening in interest rates across the union countries. They find that those countries with a large share of Gross Domestic Product (GDP) in construction, capital goods and consumer durables will be more exposed to changes in the interest rate, *ceteris paribus*. These findings designate the countries Luxembourg (cluster of financial services) and Germany (producer of capital goods *par excellence*) as the extremes of an otherwise less country specific effect.

Degrees of Openness and the Transmission of Demand and Supply Shocks

More than sector structure *per se* the degree of openness plays a significant role in the transmission of demand and supply shocks. Generally, the more open an economy (i.e. the greater the value of its exports and imports over GDP) the greater the probability that it will diverge from a monetary union's average inflation. This phenomenon derives from three distinct mechanisms at work in open economies:

[14] See De Grauwe (2000) and Begg et al (1998). Mortimer-Lee finds that economies in monetary unions may react very differently to demand shocks in the way they translate demand impulses into growth and inflation (inflation-growth divide). Using correlation analysis, it is derived that the similarity between Germany and most other EU countries in their reaction to demand shock is much less than for supply shocks (1998).

[15] Intra-industry trade: the exchange of broadly similar goods.

[16] A good example of which being the regional business cluster around the production line of the Smart car at the French-German border.

[17] See European Economics lecture 30/01/2004

Firstly, small open economies (SOEs) - the most open type of economy - given their lack of sufficient economies of scale in crucial industries[18] are above average importers of inflation with the degree of the latter mechanism hinging crucially on the exchange rate *vis à vis* their trading partners[19] (McAleese, 1997).

We shall account for this importance of the exchange rate for open economies in general as the second mechanism driving inflation differentials. If, for example, the ECB tightens interest rates, then relatively more open union countries will experience more of a loss in competitiveness and more of a terms-of-trade[20] improvement (Dornbusch et al, 1998). This is due to the more pronounced effects of exchange rate changes on demand, supply and the price level in open economies.[21]

The reverse mechanism, interest rate reductions, as pursued by the ECB in the past years, will therefore leave more open economies at relatively higher prices and less open economies with lower price hikes. Given the fact that interest rate exposure[22] and exchange rate exposure coincide to greater or lesser degrees for the different union economies, the effects of a change in the interest rate vary throughout EMU.[23]

Thirdly, more open economies are often both more exposed to extra-union Foreign Direct Investment (FDI) and extra-union export demand, a relationship that can contribute to inflation differentials between countries in a monetary union.

---

[18] Let us take as an example Ireland and the absence of an auto-industry from the island economy.

[19] The ECB in 2003 also finds that "different levels of exposure to external shocks, such as the marked fluctuations of energy prices and exchange rates over the last four years, also appear to have contributed to the existence of inflation differentials across euro area countries. Due to national differences in the degree of openness concerning extra euro area trade and oil dependency, import prices and inflation have been affected differently across countries. The resulting impact on inflation dispersion should, however, be temporary." (ECB, 2003) .

[20] The terms of trade are the ratio of export to import prices.

[21] For an effective appreciation of its currency, profit margins of exported goods fall as foreign exchange buys less domestic currency, aggregate demand falls and output falls. But since imports prices fall and reduce the price of production, supply shifts outwards leaving the economy at previous output levels but lower prices. Less open economies also experience the re-equilibration to previous output levels through shifts of AD and AS, however, these movements are less pronounced resulting in relatively less marked changes in price levels. (See De Grauwe, 2000)

[22] sensitivity to credit, as described under 'Sector Structure' (III.2.1)

[23] See Dornbusch et al. 1998.

Differences in Financial Systems and the Transmission of Interest Rate Shocks

As we could observe in the context of sector differences and exchange rate exposure, the monetary mechanism[24] may differ significantly between the EMU members.

Dornbusch et al (1998) emphasise, however, that in Europe, differences in the monetary mechanism essentially derive from the role of financial markets and banks. The credit channel, they find, is relevant in the EMU because, especially in continental Europe, banks provide the bulk of firms' credit. So while the Anglo-Saxon financial systems experience large wealth effects as consequence of a rise in interest rates, the continental type may suffer from banks' credit rationing. Affecting national inflation through both real factors and relative money supply growth the systems may produce significant differentials in the former.[25]

Amongst the continental EMU countries that are broadly similar in their use of bank as opposed to capital market credit, significant disparities also lead to differences in monetary policy transmission. Since banks aim at cultivating long-term relationships with customers they are prepared to absorb, at least temporarily, some effects of an interest rate hike (Ibid).

Table 2: Differences in the response of bank lending rates (in basis points) to a 100 basis point rise in central bank interest rates

After:	One Month	One Quarter	Two Quarters	One Year
DE	0	36	53	74
NL	71	95	102	103
BE	63	95	93	93
FR	51	53	55	58
ES	0	100	104	105
IT	19	72	97	106
UK	100	100	100	100

Source: BIS (1995)/Dornbusch et al., 1998.

[24] The monetary (transmission) mechanism refers to the ways in which a change in the interest rates affects an economy. It works along the lines of three channels: The 'monetary channel'/'textbook channel' concerns the transmission of interest rates across financial markets by arbitrage along the yield curve and across financial products affecting the market value of wealth. Secondly, the 'bank lending channel'/'credit channel' operates through the supply of bank loans to borrowers without direct access to financial markets and the third, the "balance sheet channel"/"broad credit channel" operates through the effect of monetary policy on the value of collateral, and thus the availability of credit to those requiring collateral to obtain funds (see Dornbusch et al 1998).

[25] Ireland, once more, appears to be an outlier due to firms' greater use of capital markets as source of credit.

Since the degree of absorption depends crucially on the competitiveness prevalent in the national banking sectors, differences are still significant, but consolidation is ahead. So while bank responses to a rise in interest rates varied markedly in 1995 and Germany remains the EMU country with the greatest bank competition per square kilometre (The Economist, 2003) developments point at consolidation. While the G10 *Report on Consolidation in the Financial Sector* (2001) finds that, "the euro has accelerated the speed of financial market integration in Europe," joint ventures and takeovers currently seem imminent in European banking. According to the report and in line with current events, consolidation is likely to reduce competition across the union and lead to a fall in the disparities in the response of bank lending rates and inflation differentials.

A third cause of inflation differentials rooted in the structure of the financial systems stems from differences in legal frameworks. By granting a greater degree of protection to banks extending mortgages, some European legal systems make lending easier by reducing the percentage of collateral required for borrowing, leading to the very different transmission of shocks throughout the union and thus differentials in inflation records (De Grauwe, 2000; Dornbusch et al, 1998).

Procedures for Pay Negotiations and Productivity Differences

As introduced previously, strong labour unions can raise wages by more than productivity increases. Hence, if union degrees of centralisation and strengths differ between euro zone countries, inflation differentials are likely to result. According to De Grauwe (2000), countries with regional labour unions contribute more to inflation[26] while perfectly centralised and perfectly decentralised unions are less likely to do so.[27] Equally important, a correlation of growth in real wages between 1987 and 1997 and trade union membership by Mortimer-Lee (1998) produces a clear and positive relationship between both variables. It emerges that the Eurozone countries with the lowest average real wage growth are the Netherlands and France (0.5 and 0.9% per annum), while Finland is most unionised and experienced annual real wage increases of 2.4%. Contrary to such findings, the Solow-McDonald model, predicts a convergence of labour unions' wage demands and unemployment tolerance in a monetary union due to the introduction of a centralised monetary policy (De Grauwe, 2000). Theory here, however, is not (yet) verified by actual developments. The ECB in its 2003 report highlights that "the

[26] If unions bargain locally there exists a free-rider problem and thus unions do not exercise wage restraint.

[27] If unions are perfectly centralised, they will take into account the inflationary effect of wage increases. Perfectly decentralized unions (unions at the firm level) will exercise restraint since wages have direct impact on the employment situation of their members by affecting the competitiveness of the firm.

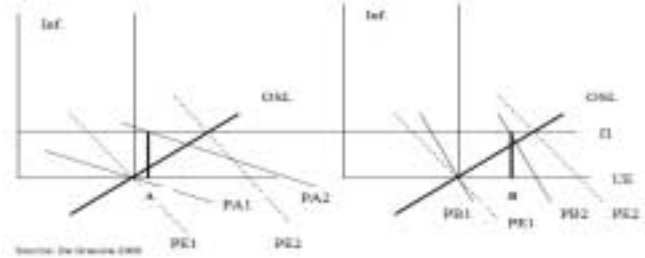
observed diversity in inflation rates since 1999 is mirrored by a considerable diversity in profit margin changes and unit labour cost (ULC) developments." It finds profit margins to be the dominant explanation for inflation differentials for Belgium, Spain, Greece, France, Ireland and Italy while ULCs are important for Portugal, Luxembourg, the Netherlands, Austria and Germany.[28]

Should the Solow-McDonald model be verified at later stages and inflation differentials continue, the outcome would be non-optimal, leaving countries with lower productivity growth and/or above average inflation without yet another instrument available for the correction competitive disadvantages (De Grauwe, 2000).

Labour Market Rigidities

Labour markets across Europe are differently regulated while labour mobility between EMU members is still very low. Where two union countries differ in their labour markets' degree of rigidity, the ECB's reaction (monetary policy to achieve optimal euro area inflation (Π) at the intersection of the optimal stabilisation line (OSL), with the post shock EMU-Phillips curve, PE2) to a symmetric shock (equal shifts in national Phillips curves PA and PB) will initially produce equal inflation rates but different national unemployment trade-offs (De Grauwe, 2000). Since unemployment reactions differ (A, B) this will cause asymmetrical changes in other real factors between the countries and ultimately translates into inflation differentials.[29]

Figure 2: Labour Market Rigidities



[28] We should note the fact, that this pattern does not mirror the characteristic influence of labour unions as described by De Grauwe (2000) and as found by Mortimer-Lee (1998).

[29] Slopes of 'National' Phillips curves (PA and PB) differ according to the degree of labour market rigidity: The more rigid a country's labour market the steeper its (short-run) Phillips curve since employers are relatively less capable to react to inflationary surprises by firing employees.

Policy

Fiscal Policies

Since under EMU, one monetary policy must now fit all, fiscal policy, from both Keynesian[30] and Monetarist[31] perspectives, has increased in importance. The fact that under EMU the crowding out effects - be they partial or complete - will be spread throughout the entire euro zone as opposed to only those countries pursuing expansionary policy alone may rise the efficacy of fiscal activism (Mortimer-Lee, 1998). If applied in accordance with the national economic cycles only, fiscal policy could in this way reduce inflation differentials and thereby contribute to more convergence.

The breach of the European Stability and Growth Pact by France and Germany should, nonetheless, have caused mixed feelings with Monetarists (seeing an increasing likelihood of monetary union failure ahead) and Keynesians (expecting some moderation of these dangers). Enthusiasm about fiscal policy is also dampened by its extremely slow nature of effect. The common two-year period it takes to affect the economy is also a strong indication as to why current inflation differentials cannot yet stem from the recent breaches.

Conclusion

In the course of this essay we have taken a look at cyclical, structural and policy-triggered causes of inflation differentials in a monetary union and have linked the first two of these three broad categories to the divergence in the Eurozone. We have seen that - depending on conditions - a variety of direct and catalyst factors can be the cause of such developments, an understanding that acknowledges the peculiar role of the ECB as a policy maker squeezed between its 'hard-nose' Bundesbank legacy and the apparent divergence of its members.

We found that the Balassa-Samuelson effect of normal and catch-up growth is a key source of inflation divergence, specifically in the case of the strongest

[30] Depicting the LM curve as positively sloped, J.M. Keynes suggested government spending as adequate reaction to demand deficiency and the resultant output gaps (difference between potential and actual GNP resulting in unemployment). By increasing expenditure, government would shift the IS schedule to the right and thus increase output and interest rates. For a detailed discussion see Blanchard 19 chapter 6 and SMG chapter 11

[31] Based on the assumption of a vertical LM curve, Monetarism depicts fiscal expansion under independent central banks, i.e. in the absence of accommodating monetary policy, as ineffective tool, raising interest rates rather than output. Fiscal restraint, however, is regarded as highly efficient as it reduces interest rates and thus stimulates consumption and investment. For a detailed discussion see Blanchard (2000).

outliers. Its explanatory power does, however, vary markedly across the Eurozone with 'cost chain' patterns being equally inconsistent.

Finally, we must conclude that as long as dissimilarities in the labour market regulations, labour power and financial systems persist, inflation rates are unlikely to converge given the significantly different structures and orientations of the EMU economies.

Bibliography

Blanchard, O. (2000), *Macroeconomics.* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Center for Economic Policy Research (2004), "Industrial Location, The Labour Cost–Product Quality Nexus." [Online]. *Centre for Economic Policy and Research*. Available from: <http://www.cepr.org/pubs/bulletin/meets/2273.htm>

De Grauwe, P. (2000), *Economics of Monetary Union*. 4th ed. Oxford: Oxford University Press.

Dornbusch, R. et al. (1998), "Immediate challenges for the European Central Bank" In: *EMU: Prospects and Challenges for the Euro*, Begg, D. et al. (eds.), Oxford: Blackwell Publishers.

European Central Bank (1999), *Monthly Bulletin: October 1999*. Frankfurt: ECB. Available from: <http://www.ecb.int>

European Central Bank (2003), *Inflation Differentials in the euro area: potential causes and policy implications*. Frankfurt: ECB. Available from: <http://www.ecb.int>

European Commission (2004), *European Economy.* Available from: <http://europa.eu.int>

Freudenberg, M. et al. (1995) "Intra-Industry Trade and the Single Market: Quality Matters." *Economic Policy Research.* Discussion Paper 1959.

Group of Ten (2001), *Report on the Consolidation of the Financial Sector: January 2001*. Paris: OECD. Available at <http://www.oecd.org>

Howells, P. and Bain, K. (1992), *The Economics of Money, Banking and Finance.* Addison: Wesley Longman.

Krugman, P. (1989), "Differences in Income Elasticities and Trends in Real Exchange Markets." *NBER Working Paper*, (2459).

Krugman, P. (1991), *Geography and Trade*. Cambridge:MIT Press.

McAleese, D. (1997), *Economics for Business*, London: Prentice Hall.

Mortimer-Lee, P. (1998), "Can Euroland work with one interest rate?' Mimeo.

Mundell, R. (1961), "A Theory of Optimal Currency Areas." *American Economic Review*, Vol. 51.

Temperton, P. (ed.) *The Euro.* Chichester: John Wiley and Sons.

Shaw, G. et al. (1997), *Macroeconomics: Theory and Practice in the UK.* 3rd ed. Oxford: Blackwell Publishers.

MERCANTILISM AND THE ASIAN DEVELOPMENT MODEL

BY EIMEAR SEXTON

Senior Sophister

In this essay Eimear Sexton discusses the mercantilist nature of the *Asian Tigers*' institutions and policies since the 1960s as well as examines their governments' position regarding balance of trade, balance of payments on current account and exchange rate regime. After looking at the Asian Crisis of 1997, she goes on to conclude that mercantilist policies of the Asian Development model are sound policy prescriptions for other developing economies.

Introduction

While the assertion by a Harvard expert on Asian Studies that "there are at least three models of East Asian development" (O'Hearn, 1998) is generally seen as understated, there are a number of common characteristics of economic development present in most of the newly industrialised countries of East and Southeast Asia. These characteristics together form what has been termed the 'Asian development model'.

There has been phenomenal growth in this region over the past number of decades. A glance at Asian tiger growth rates over the 30 year period from 1960 is staggering, and when put in the context of OECD growth rates of 2-3% over the same period, the significance of the statistics are highlighted. While growth in the Asian economies then slowed, annual growth rates of 5.2% to 8.3% in the first half of the 1990s are still quite impressive (Numazaki, 1998).

Table 1: Average annual GDP growth rate

	1960-70	1970-80	1980-89
South Korea	8.5	8.7	9.2
Taiwan	9.2b	9.7	8.0
Singapore	8.8	9.0	6.9
Hong Kong	10.0	9.5	6.3

Source: Chowdhury and Islam, 1993.

The four Southeast Asian nations grew at an average annual rate of 6-7% between 1965 and 1980 and, excluding the Phillipines. These economies grew at 5-7% per annum during the 1980s. Late industrialiser China grew at 6.9% on average for the period 1965-1980 and 9.7% for the 1980s.

While neither 'East-Asian' nor 'Southeast Asian' accurately reflects the regional distribution of the countries being considered in this paper, for the purpose of clarity, they will be collectively referred to as the 'Asian Economies'. They include: Japan, South Korea, Taiwan, Hong Kong, Singapore, Thailand, Malaysia, Indonesia, the Philippines and China. Similarities of both policies and structures are readily observed across the Asian Economies, although many are at different stages of development.

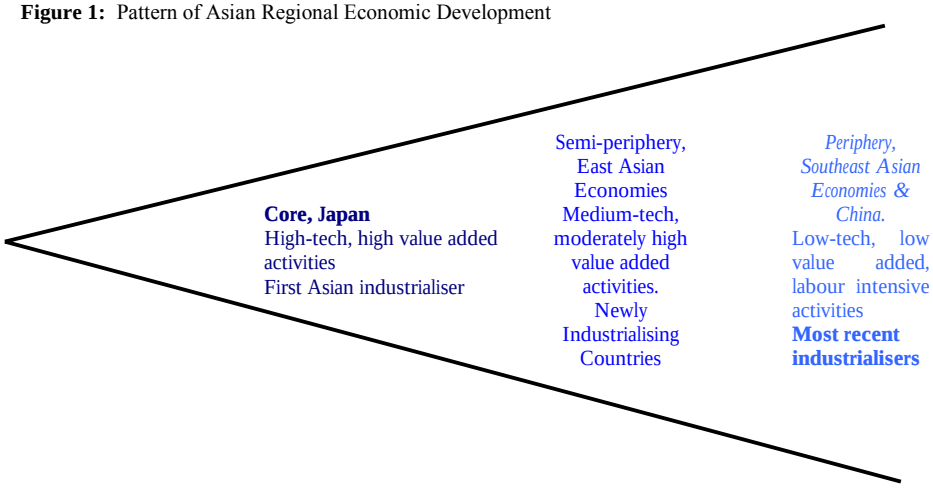
There have been three waves of industrialisation in the region. Japan was the first non-Western economy to industrialise, with economic growth taking off soon after the end of World War II. Following Japan, the East Asian tigers of South Korea, Taiwan, Singapore and Hong Kong emerged in the early 1960s.[1] In later years rapid economic growth occurred in Thailand, Indonesia, Malaysia, the Philippines,[2] and finally China.

The development of the region as a whole has largely been due the industrialisation of Japan. As Japan developed over the latter part of the 20th century, and moved its production up the value chain, in other words, producing more technically advanced and value added goods, it began first to import raw materials from its Asian neighbours and then to outsource lower value added elements of the production process. This process 'stimulated… economic growth in less developed nations along the western Pacific Rim' (Hill and Fujita, 1995). What emerged was an inverted 'V' or what Akamatsu Kaname (O'Hearn, 1998) calls the 'flying geese' pattern of regional development of the Asian economies with Japan at the forefront (See Figure 1).

These regional patterns also map out the balance of economic power, with core, technology rich Japan holding the most power in the region to attract high value industry.

[1] These four will be called the 'East Asian Economies'
[2] These four will be called the 'Southeast Asian Economies'

Figure 1: Pattern of Asian Regional Economic Development



Source: O'Hearn, 1998

It is my hypothesis that the policies and structures present in the Asian Economies since the 1960s are largely mercantilistic. The model of Asian development is undeniably similar in spirit, even if not by intent, to the policies and structures which dominated economic thought in Western Europe from the Middle Ages, nearing 1600, to the publishing of Adam Smith's *Wealth of Nations* (1776). In evaluating the weight of this proposal, I shall examine the prevalence of policies and structures present in the East Asian region, which mirror those present in the mercantile era, the notion of economic power as political power, the support of such policies by structures held in place by the 'strong' or 'developmental' nation state and policies towards trade including import-substituting industrialisation and export-led growth. I shall also discuss the desire for balance of trade and current account surpluses and foreign exchange reserves and the 1997 financial crisis in Asia. Based on my conclusions, I shall comment on the validity of mercantilist policies as policy prescriptions for emerging economies.

Economic Power as Political Power and Structures of the strong developmental nation state

In his book entitled *Mercantilism*, Professor Heckscher (1935) described mercantilism as a uniform body of doctrine built up on five different parts, in which "The Pursuit of Power" was the most important. In his view, economic measures were the primary weapon used to secure political unification and national power. This has been the defining feature of Asian economic and political policy.

Japan's imperial ambitions at the turn of the 20th century are well known. The Meji era, brought about by the death of the emperor in 1912, was marked by huge government domestic and overseas investments and defence programmes in an effort to increase its political power. During the following Taisho era, in an effort to expand its influence in China, Japan declared war on Germany[3] and occupied German-leased northern territories in China and its islands in the Pacific. Japan's power in Asia grew with the demise of the tsarist regime in Russia and the 1917 Bolshevik Revolution. The year 1919 saw Japan sitting among the 'Big Five' powers at the Versailles Peace Conference, a confirmation of its rise in political significance. The extent of Japan's ambitions were shown by their invasion of China in 1931 and the now infamous attack against US forces at Pearl Harbour in December 1941. However, following 14 years of war, their humiliating defeat and the annihilation of Hiroshima and Nagasaki in 1945, the death toll in Japan had risen to over 3 million. The country was in ruins and a huge occupation army of US and British soldiers remained in Japan as part of the surrender agreement.

This put paid to Japan's pursuit of power through militaristic means. After WWII Japan re-oriented its policy towards the attainment of power and influence in the world through the build up of economic might. Japan had effectively been thrust towards the mercantilist ideal of "state making and national economy making at the same time" (Schmoller, 1896). It was through economic growth that Japan's power would be realised. By the 1970s, Japan had achieved so much progress that it was predicted to overtake the US as the world's number one economy.[4]

Most pieces of mercantilist policy that were put forward identified the merchant's profit with the national good. This was based on the perceived link between trade and power – that 'trade' tended to 'follow power' (Magnusson, 1994). This perceived connection prompted Child to stress in 1693 "…that Profits and Power ought jointly to be considered" (Magnusson, 1994). With the pursuit of the power of the nation state at the forefront of public policy, and economic growth through trade the means to attain this power, state intervention was an essential part of mercantilist doctrine. The notion of the strong nation state has been at the heart of the Asian development model. Alvin So (1995) effectively sums up the regional mentality towards Asian economic governance when he says that "states have a strategic role to play in taming international and domestic forces and harnessing them to national ends."

While neo-classical economists argue that East Asian economic success has been based on its relative economic efficiency in world markets, those who take institutional approaches to economic change contend that Asian governments have

[3] In August 1914.
[4] See, for example, Vogel (1980) and Kahn (1970).

intervened, often in a market-distorting way, to ensure the prosperity of domestic industries. While the World Bank (1993) has used the East Asian 'miracle' economies to extol the virtues of export-led growth paired with openness to international markets, O'Hearn (1998) notes that a large number of experts have argued that Asian successes were largely the result of effective state interventions, which strategically protected key industries to build them up until they were able to compete in export markets. The developmental role of Asian states has been facilitated by their autonomy and independence from the pressures of social groups and classes.

States were involved in supporting some activities while discouraging others, or what Amsden (1989) refers to as the use of 'carrots' and 'sticks'. This involved, for example, the imposition of quantitative restrictions and tariffs on the imports of some sectors, while other industries were exempt from such restrictions. As development and growth took place and economies shifted towards export-led growth, support of indigenous and desirable industries was made by direct subsidies, enabling them to make a profit with higher than average costs and still sell at prices lower than international market equivalents.

Interventionist policies had become so rampant during the early stages of economic growth, that, for example, Ed Mason dubbed the Korean state "Korea Incorporated", arguing that the South Korean economy was like a huge corporation, with the state as its board of directors (O'Hearn, 1998). Although the heights of the economy were dominated by huge *Chaebol*, like Hyundai and Samsung, these in turn were strategically directed by the state.

Interventionist behaviour of the type practiced in this region is largely frowned upon in the global arena. However, Asian countries have remained largely camouflaged from the piercing lens of Western focus for the past number of decades. This has been the result of a long history of geopolitics that has been prevalent in this region since the beginning of the Cold War. As front-line states in this political war of Western capitalism versus Soviet-led communism, in addition to receiving financial aid for accepting a Western military presence, the East Asian states essentially exempted themselves from economic policy rules with no obvious repercussions (Cumings, 1984; So, 1995).

The Mercantile System and its Policies Towards Trade:

Import-Substituting Industrialisation

A key element of early mercantilist industrial policy was the practice of what is now termed Import-Substituting Industrialisation (ISI). ISI endorses the protection of domestic industry from foreign competition behind high tariff walls

and the provision of a range of incentives and subsidies for favoured industries. The creation of work and employment and the nursing of industries, both as ends in themselves and as means of strengthening the country, became the aims of state policy at the end of the 17th and throughout most of the 18th centuries. The methods used were tariffs and embargos on imports, prohibitions of the export of tools and skilled craftsmen, the encouragement of the import of raw materials or of their production at home, the supervision of the quality of products and subsidies to those who were developing new industries.

Import substitution was used extensively as a policy for growth in Asian economies. This type of inward-oriented strategy of economic development was adopted in Japan in the years post WWII, in the East Asian Economies in the 1960s and in South East Asian Economies in the 1960s and 1970s.

Import substitution policies are closely correlated with the early stages of industrialisation in these Asian economies. Import licensing and quantitative restrictions were the most far-reaching instruments of protectionist policy at that time, but they also led to most abuses, both in their administration and in permitting inefficient manufacturing enterprises to be established (ADB, 1971).

While import substitution has become almost completely redundant in the Asian economies since the 1970s, the encouragement of home consumption is a feature still found in many of the Asian economies, especially in the area of food production. Japan, for example, actively and overtly pursues a policy of agricultural protectionism. Despite the obvious cost disadvantage of maintaining agricultural production in a country with high labour and land costs,[5] this is considered subordinate to its desire to maintain self-sufficiency in the interest of its national security. In 2003, Japanese rice production was supplemented by only 7.5% of total consumption in imports in this area (of the 8,658,000 metric tons of rice consumed in Japan in 2003, only 650,000 tons came from abroad). Similarly, China's modernisation began in 1978, with reform of its agricultural sector with the goal of becoming fully self-sufficient in food production. This goal was largely reached by the 1990s and today China, the world's biggest consumer of rice imports less than 1% of its total consumption.

Desire for Balance of Trade and Current Account Surpluses

Mercantilists looked upon the economic process from the point of view of the primitive stage which capitalism had reached at that time and were thus led to identify money with capital, identifying "…precious metals as the sole constituents of the wealth of the nation" (Viner, 1930). Professor Heckscher has given an interesting account of the "fear of goods", the almost fanatically exclusive concern

[5] Heckscher-Olin theory, the textbook treatment of land and labour costs

with selling, which characterised mercantilist thought. In sharp contrast with the aim of securing an abundance of goods, which had characterised earlier state policy, the mercantilists thought, in the words of their greatest German representative, Johann Joachim Becher, "that it is always better to sell goods to others than to buy goods from others, for the former brings a certain advantage and the latter inevitable damage" (Heckscher, 1935).

It was particularly in the sphere of foreign trade that the 'fear of goods' showed itself, resulting in the mercantilist search for an export surplus, a relative surplus. Davenant, writing in 1697, argued that in domestic trade the nation in general did not grow richer, only a change in the relative amounts of wealth of individuals took place; but foreign trade made a net addition to a country's wealth.

The Asian Economies have vigorously pursued a current account surplus as a major part of its economic policy. Their generally impressive current account surpluses have been led by the shift in policy towards export-led growth. Over the 15 year period 1970-1985, manufactured exports grew annually by more than 25% in South Korea, Taiwan and Singapore (O'Hearn, 1998). The conventional wisdom is that these rapidly expanding exports drove higher economic growth rates in general. The exceptional growth over the period in question has been illustrated. Average export growth per annum over the period 1970-1979, a strong period in terms of economic growth in the region, is shown below:

Table 2: Average Export Growth per Annum, 1970-1979

East Asia	South Korea	37.9%
	Taiwan	30.8%
	Singapore	28.0%
	Hong Kong	22.1%
South East Asia	Indonesia	34.9%
	Malaysia	23.3%
	Philippines	17.6%
	Thailand	25.2%

Export-led Growth

By 1994, combined exports from the four Asian tigers, the four 'cubs' and China had grown steadily to 17.3% of total world exports from 8.3% of world exports in 1981 and 13.9% of exports in 1991. The composition of these exports has changed over the course of development of these economies, however, evolving from food and textile exportation in the earlier phases of development to more manufactured, more technology reliant goods as economies mature.

Pursuing a policy of export-led growth in Asia has not been without its difficulties. For many of the Asian Economies, natural resources in terms of fertile land, metal and mineral reserves have not been in large supply. For these countries, comparative advantages were not endowed but have been built or captured over time. However, one area in which they held a considerable advantage was in their abundant supply of cheap labour. The condition of high-density population in a country was strongly regarded by mercantilists as a propeller of economic growth. Indeed, with the rise of the Netherlands as an economic power in the 18th century, many mercantilist writers identified this as one of the sources of its increased affluence over this period. The general lack of substantial natural resources has had its consequences for production. The industrial strategy of export-led growth in Asian economies has generated a double incidence of dependence (or 'double dependency': Numazaki, 1998) on the United States as a substantial market for exports and a dependence on Japan for imports as a consequence of Japanese 'out-processing'[6].

Accumulation of Specie and Foreign Exchange Reserves

While during the mercantilist era countries accumulated stocks of gold over time as a result of their policies to encourage balance of trade surpluses, such stocks of wealth held by Asian financial institutions are held primarily in US dollars. Since the collapse of the Bretton Woods accord, the US dollar has assumed the role of the World's international reserve currency of choice. In effect, the dollar has replaced gold and other precious metals as the primary store of wealth for most countries.

The extent to which the Asian Economies hold reserves is well in excess of that, which is standard in Western economies. Immediately prior to the Asian financial crisis, in 1996, international reserves of these countries were as indicated in Table 3.

These quantities of reserves are even more substantial when one takes into account the huge volumes of imports such countries had, with many exporting firms reliant on raw materials from abroad and the cross-country trading of subcomponents in line with the vast networks of Original Equipment Manufacturers (OEM) established in the Asian region.

[6] The success of most Asian economies is based on the passing on of certain parts of the production process through Original Equipment Manufacturer (OEM). To this end, such economies are reliant on Japanese imports to fulfil their role in the production process

Table 3: International reserves immediately prior to the Asian financial crisis

Country	International Reserves (months of imports)
Japan	8.5
Taiwan	10.3
Hong Kong	10.9
Singapore	7.0
South Korea	2.8
Indonesia	6.1
Thailand	6.6
Malaysia	4.3
Philippines	4.4
China	9.1

Source: Hong Kong Monetary Authority, (2004); CEPD for Taiwan, (2004); APEC (2004).

While reserves of gold were built up in the mercantilist era in a somewhat passive way, reserves in these Asian economies were used to buttress their fixed exchange rate regime. Foreign exchange reserves also serve to hedge against the relatively massive amount of foreign-held debt in all Asian economies of this time against dips in confidence. Easy bank loans made available by 'relationship banking' and the government-guided 3B trilogy of bureaucrats, bankers and businessmen (Choi, 2000) facilitated the evolution of the huge debt to equity ratios of financial institutions. Large US dollar and other international reserves are also seen as a sign of economic power.

'All Fall Down' - The End of the Asian Miracle (Krugman, 2000)

Many commentators believed that the Asian financial crisis of 1997 signalled the end of the Asian miracle. The Asian crisis began on the 2nd July 1997, when the Thai government devalued the Baht in response to overwhelming speculative pressure. What followed was a spate of attacks on the currencies of the region. Indonesia, Korea, Malaysia and to a lesser extent the Philippines, Taiwan, Singapore and Hong Kong all fell victim to the so-called Asian flu. Thailand's currency depreciated by half, Korea's by 20%, Malaysia's by 30%. Only Hong Kong was able to maintain its exchange rate. The stock markets of the region also came under attack with investors pulling their money out *en masse*. The Thai stock exchange lost 90% of its value by the end of 1997, while Indonesia lost 85%, Korea 80% and Singapore 60%. The collapse was not just confined to the financial world;

real economic activity was hit hard. Indonesia's economy contracted by 13.7% in 1998, Thailand's by 10% and Hong Kong, Korea and Malaysia each saw contractions of between 5 and 7.5%. There was a massive increase in unemployment, the rate more than tripling for Thailand, Indonesia and Korea in the two years from 1996.

There are two competing views on where the fault of the Asian financial crisis lies. The first of these is that the model of Asian development itself was to blame for the crisis. Asian capitalism (or crony capitalism) is based on the symbiotic relationship between government, banks and industry. As Krugman (2000) suggests in his usual cutting fashion, "Many Westerners have turned the story of Asia's crash into a sort of morality play, in which the economies received their inevitable punishment for the sins of crony capitalism." Proponents of this view argue that many of the same factors, which underpinned Asia's extraordinary economic growth, were also the source of its weakness; policy directed lending, the close ties between corporations, governments and financial institutions, fixed exchange rates, and the general control of market forces.

The opposing view is the failure of Western capitalism or a market-failure view. This perspective, vehemently argued in Radelet and Sachs (1998), Wade (1998) and Wade and Veneroso (1998), places the Asian development model in the context of its own tradition. The Asian model of financial intermediation, characterised by close relationships between government, the banking sector and industry, had played a major role in the impressive growth of the Asian economies. While this system had operated effectively in an era of restricted capital movements, supporters of this latter view of the crisis hold that it was the unrestricted inflows of foreign capital following capital liberalisation in this region, prompted by the International Monetary Fund (IMF) and Western powers that upset this symbiotic system and caused it to unsettle the equilibrium in place (Dean, 2000)

While Asian officials proclaimed the soundness of the economic fundamentals on which their systems were based, they failed to recognise that they were assuming that the separation of economics and politics was straightforward. Investment, capital flows and growth are part of the economic sphere and Asia performed well in this sphere. However, regulation supervision, transparency and a sense of balance between dynamic growth and system stability can only be generated in the intertwined political sphere, in which Asian nations failed miserably. This view is reinforced by Sopiee (1996), who observes that, "The tiger economies of East Asia have been first rate in the management of the economic fundamentals in the real economy. They have been second rate in the management of market fundamentals."

Cerra and Saxena (2003) in an IMF survey said "the recovery phase (in Asia) is predominantly characterized by a return to the normal growth rate rather

than a higher-than-normal growth rate". If it is the case that such economies are growing once more, notwithstanding the fact that such growth is from a lower base, a case can be made in line with the market-failure view that it was not the mercantilistic-type policies used by Asian economies in the 1960s through the 1990s that caused the financial crisis of 1997 and consequently such policies have not yet been rendered redundant as effective tools for the facilitation of economic growth and development.

Conclusion

"…Mercantilism… was intended to promote production and commerce of private entrepreneurs who benefited from and contributed to the consolidation, prosperity and power of nation-states, with foreign trade being the most strategic variable."

(Palgrave, 1976)

Palgrave's definition, while it refers to a period of economic thought that 'ended' more than two hundred years ago, could easily describe the economic policies and motivations of twentieth century Asia. Yusuke, chief economist of the Institute of International Finance and former Asia expert at the IMF, noted a "deeply rooted mercantilist instinct" in Asia "with an almost religious attachment to trade and current account surpluses" as well as a tendency much greater than other countries to hold foreign reserves.

Together the Central Banks of Japan, China, Hong Kong, Taiwan and South Korea hold around $1.3 trillion in official reserves (or over half the global total). Their desire to accumulate such a stockpile of dollars is twofold. It is fuelled by their desire to hold enough reserves to insulate them from a financial crisis like that of 1997 and their wish to keep their currencies from appreciating. The dollar has depreciated significantly in 2002 and 2003. Correspondingly, Japan's reserves have increased by 36%, China's by 65% and Taiwan's by 49% over the period (The Economist, 2003).

Another effect of Asian intervention in soaking up excess dollars to support their exchange rates has been that their export prices have been kept down, arguably an orchestrated effect, leading to yet more of the export-led growth it has become renowned for. Exports now account for 64% of the region's GDP, having increased from 55% in the early 1990s, before the crisis (The Economist, 2003). This impressive export performance has concomitantly led to substantial current account surpluses. The combined current account surplus of Japan, China, Eastern and South Eastern economies of Asia was $133 billion in 2002.

It is my view based on the above evidence that mercantilist policies are still being used at present in Asian countries. I believe that the financial crisis in 1997 was a consequence of an amalgamation of financial regimes that were not in sync, a market failure rather than a government failure. Although economic fundamentals were somewhat to blame, the extent of the damage caused by the crisis was probably unjustified by the afflicted economies' crimes. Despite the crisis, the Asian economies have shown signs of a remarkable recovery, suggesting that the economic fundamentals on which their regimes were based were different to those of the Western model but not necessarily unsound. Moreover, their growth since the crisis has taken place within a much more sustainable market framework i.e. in a market where much capital liberalisation has taken place. While markets are not fully liberalised – China, the most notable proponent of market restrictions, still has its currency, the Yuan, fixed to the Dollar.[7] They are now a great deal closer to the norm of industrialised Western countries.

Despite the Asian financial crisis of 1997 and its aftermath, it is clear that many of the tenets of the Asian development model in particular, its mercantilistic policies of a strong nation state, export-led growth, trade and current account surpluses – are valid policy prescriptions for other emerging economies. What it less clear is how countries should proceed, regarding capital account liberalisation and their choice of exchange rate regime.

Bibliography

Amsden, A. (1989). *Asia's Next Giant: South Korea and Late Industrialisation*. Oxford: Oxford University Press.

Asia Pacific Economic Corporation (APEC) (2004). *Asia Pacific Economic Corporation.* Available from: <www.apec.org.sg>

Asian Development Bank (1971). *Southeast Asia's Economy in the 1970s*. London: Longman

Asian Development Bank (2003). *Key Indicators 2003.* [Online]. Asian Development Bank. Available from: <http://www.adb.org/Documents/Books/Key_Indicators/2003/pdf/prelims.pdf>

[7] China's yuan has been effectively fixed to the dollar since 1994

Cerra and Saxena (2003). "Did Output Recover from the Asian Crisis." *IMF Survey*. 30/10/03.

Choi, J. J. (2000). "The Asian Financial Crisis: Moral Hazard in More Ways than One" In: Choi, J. J. ed. *Asian Financial Crisis: Financial, Structural and International Dimensions*. New York: JAI

Council for Economic Planning and Development (CEPD) (2004). "Council for Economic Planning and Development." [Online]. *The Government of Taiwan*. Available from: <www.cepa.gov.tw/english>

Cumings, B. (1984). "The origins and development of the Northeast Asian political economy: industrial sectors, product cycles and political consequences." *International Organisation*, Volume 38 (1).

Dean, J. W. (2000). "Which Failed: Asian Capitalism or International Capital Markets?" In: Choi, J. J. *Asian Financial Crisis: Financial, Structural and International Dimensions*. New York: JAI.

Heckscher, E. F. (1935), *Mercantilism*. London: George Allen & Unwin Ltd.

Hill, R. C., and Fujita, K. (1995). "Product cycles and international divisions of labor: contrasts between the United States and Japan." In: Smith, D. and Borocz, J., *A New World Order? Global Transformations in the Late Twentieth Century*. Westport, CT: Praeger.

Hong Kong Monetary Authority (2004). "Hong Kong Monetary Authority." [Online]. *The Government of Hong Kong Special Administrative Region of the People's Republic of China.* Available from: <www.info.gov.hk/hkmn>

Islam, M. and Chowdhury, A. (1993). *The Newly Industrializing Economies of East Asia*. London: Routeledge.

Jomo, K. S. (1998). *Tigers in Trouble: Financial Governance, Liberalisation and Crises in East Asia*. London: Zed Books.

Khan, H. (1970). *The Emerging Japanese Superstate*. New York: Prentice Hall.

Krugman, P. (2000). *The Return of the Depression Economies*. London: Penguin.

Magnusson, L. (1994). *Mercantilism*. London: Routledge.

Numazaki, I. (1998). "The Export-Oriented Industrialization of Pacific Rim Nations and Their Presence in the Global Market" In: Kim, E. M. *The Four Asian Tigers*. London: Academic Press.

O'Hearn, D. (1998), *Inside the Celtic Tiger: The Irish Economy and the Asian Model*. Sterling, USA: Pluto Press.

Palgrave, R (1976). *Dictionary of Political Economy*. Gale Group: London.

Radelet, S. and Sachs, J. (2000), "The Onset of the East Asian Financial Crisis", in Krugman, P. *Currency Crises: A NBER Conference Report.* Chicago: The University of Chicago Press.

Schmoller, G. (1896/1967). *The Mercantile System and Its Historical Significance.* Augustus M. Kelley: New York.

Smith, A. (1776/2003). *Wealth of Nations.* London: Bantam.

So, A. and Chiu, S. (1995), *East Asia and the World Economy*. Thousand Oaks, CA: Sage.

Sopiee, N. (1996), Speech at the 29th Meeting of the Pacific Basin Economic Council, Washington, May 25th.

The Economist (2003). "Oriental Mercantilist". *The Economist*. 20/09/03.

Viner, J (1980). "English Theories of Foreign Trade Before Adam Smith". *The Journal of Political Economy.* Vol. 38.

Vogel, E. (1980). *Japan as Number One.* New York: Harper Collins.

Wade R. and Veneroso, F (1998). "The Asian Crisis: the High Debt vs. Wall Street- Treasury IMF Complex." *New Left Review*. 3(28).

Wade, R. (1998). "Debt and Development Crisis of 1997: Causes and Consequences." *World Development.* 26 (8).

World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*. Oxford: Oxford University Press.

DEVELOPMENT, INEQUALITY AND POVERTY IN BRAZIL

BY BRENDAN GRAHAM

Senior Sophister

Brazil, despite trying a wide range of economic policies over the years including import substitution and New Consensus, has failed to achieve the standards of living enjoyed by Western European countries. Brendan Graham tries to resolve this conundrum by examining the causes and the extent of poverty and inequality in Brazil. He finds that pronounced inequality and population growth are the key factors accounting for failed development of the country. Having deduced that, policy solutions are identified as, according to the author, 'Brazil has reached a crossroads in its development process'.

Introduction

Brazil is a vast country that occupies nearly half the area of South America, and borders all its countries but two (Chile and Ecuador). It has a population of 172.6 million people approximately (World Bank, 2004), making it the sixth most populated country in the world. In terms of natural terrain, Brazil is quite unfortunate in that a good deal of its land (83%) is not arable. This is due to the large amount of rugged hills and mountain plains, scrubland, and above all, the rain forests, which occupy about one half of the country's surface area. However, it has great potential in the industrial sector, which includes the manufacturing of steel, chemicals, consumer goods, and military hardware. As well as this, there is a range of natural resources such as cocoa, soy and iron ore. As one of the world's largest producers of hydroelectric power; energy supply does not pose a problem.

Brazil may be considered a sleeping giant in terms of its economy, in that it has all the potential to become one of the main players on the world's economic stage but has consistently failed to realise its promise. As indicated above, its industrial sector in particular has the possibility to become immense. Despite this, Brazil is still ranked only as a Newly Industrializing Country by the OECD and is commonly given the status of a Lesser Developed Country. Why is this? The two main underlying reasons are the inequality of income distribution and the poverty that exist in the country today. This is partly due to the policies pursued by the

Brazilian government, which has perhaps been guilty of pursuing economic strategies that focus on growth rather than development. It is the difference between growth and development that I shall be considering first.

Growth versus Development

The distinction between growth and development has come under much scrutiny for a number of years now. Perhaps the most significant era was in the 1960s and 1970s when the World Bank, one of the main proponents of growth strategies, shifted its focus to a more development-orientated approach. In academic literature, there has been no unanimous consensus. For example, Colman and Nixson (1986) state that "a number of writers have begun to question the generally accepted distinction between growth and development".

It is important to clarify first what growth and development are. Growth can be described as changes in "the total value of wealth produced in a nation's economy before any adjustments are made for indirect taxes, subsidies and depreciation" (McCarthy and Spencer, 1992), with GNP commonly used to measure these changes. Development has generally been more difficult to define. Put in its simplest terms, it can be described as the "development of people rather than things" (Owens, 1987). The question facing us is whether we can use measures of growth such as GNP and GNP per capita to measure changes in development.

It is impossible to find a definitive answer to this question. There are many who will point out the often-close correlation between the ranking of countries in terms of GNP per capita, and their ranking in terms of measurements of development, such as the HDI. For example, China is ranked 107th in the world in terms of its GNP, and is 106th in the world in terms of its development, as measured by the HDI, a difference of only one place. McAleese and Burke argue "GNP per head serves as a remarkably good proxy for more sophisticated measures of human welfare" (McAleese and Burke, 2000).

However, if we use Brazil as a case study, we can find many counter arguments to the theory that growth is a good indicator of development. Brazil has at some stages experienced rapid growth in its economy, in terms of GNP. However, as Todaro (2000) points out, "during these times of prosperity, most Brazilians have benefited very little". Perhaps Brazil's president between 1969 and 1974, Emilio Garrastazu, best summarized the situation existing in the country. When asked by a journalist what he thought of Brazil's 'economic miracle' (referring to growth rates of over 10% per annum between 1967 and 1971), the president replied, "The miracle is good for Brazil's economy, but bad for its people" (Goulet, 1983). Unless

economic growth results in an improvement for its people, as is clearly not the case for Brazil, the growth is somewhat meaningless.

As well as this, we must consider a number of the shortcomings of GNP as a measurement of growth, which will of course distort its accuracy if we use it to indicate development. A number of these failures are relevant to Brazil. For instance, GNP takes no account of non-monetary services. Americans, for example, may have a much higher tendency to hire cleaners for the house, painters for painting jobs, and so on. All the money generated is included in GNP. However, if Brazilians do not outsource for these types of jobs, their GNP will be lower. The development between the two nations is not actually different, but is made to appear so if we use GNP as an indicator for development. Conversely, GNP takes no account of environmental damage, which works against Brazil. Their pursuit of rapid growth has been disastrous for the environment. In 1995, for example, 110,000 square miles of rain forest were cleared, an all-time record (Todaro, 2000). While this type of environmental damage undoubtedly impacts upon the development of the country, in that the natural habitat is destroyed, GNP does not take this stunted development into account.

For one of the starkest examples of how widely growth and development can differ, one must look outside Brazil to a comparison between the economies of Sri Lanka and Guinea in 1992. At this time, the two countries had a similar GNP per capita, both at $500. However, the difference between the two in terms of development was remarkable. Sri Lanka had a HDI figure of 0.665, compared to a mere 0.191 in Guinea. Breaking down how these figures were derived, we find that Sri Lanka's life expectancy was 71.2 years against Guinea's 43.9, the adult literacy rate was 89% compared to 27% in Guinea, and there was an infant mortality rate of 24 out of 1,000 in Sri Lanka, with the figure at 135 in Guinea (Todaro, 2000). The GNP per capita figure gives no account of this huge gap in development. It is clear, therefore, that income per capita has a tendency to be a very poor indicator of development.

So is the solution to use the HDI alone as a measure of development? Unfortunately this cannot be the case, for the HDI itself has shortcomings. While the HDI includes longevity, knowledge, and income per capita, it omits many of the other factors that contribute to development such as the aforementioned environmental damage, population growth, and, significantly for Brazil, inequality. This inequality has resulted in widespread poverty. So while Brazil had a HDI figure of 0.81 in 1995 (Todaro, 2000), this may be overstated due to the extensive inequality of income and the widespread poverty that is not fully taken into account by the HDI. These will be discussed in the next section. As McAleese puts it, "the single HDI value for each country is a national average which can conceal inequalities at a regional and sectoral level" (McAleese, 2001).

Whatever the case, there would still seem to be more than sufficient grounds to continue to focus on development rather than growth, and to consistently improve on our measures of development in order to aid us in achieving this goal. We cannot undervalue the merits of growth itself, for it gives the potential to achieve a greater quality of life for its people. But we cannot presume it to be sufficient.

Inequality

Inequality of income has for a long time been a thorn in the side of Brazil's economy, hindering both its growth and its development (for simplicity's sake we will consider GNP and GNP per capita, and HDI to be the measures of each respectively, even though we know them not to be fully accurate). It would seem that many of the problems facing Brazil's economy today stem from this inequality, either directly or indirectly, with a number of these problems strongly interlinked.

The figures relating to income inequality are startling, with a study of income distribution showing the extent of the problem. Almost half the income (48%) generated by Brazil's economy is received by only 10% of the people, with the highest 20% accounting for 64% of the income. This compares with a figure of 39.3% in Sri Lanka, despite the fact that Brazil has a GNP approximately six times higher (World Bank, 1999). Further to this, the poorest 20% of the people own a mere 2% of the income (Todaro, 2000). To put it another way, the richest fifth enjoys income about 25 times that of the poorest fifth (Mankiw, 2001). This distribution of wealth in Brazil is one of the most unbalanced of any country in the world, surpassed only by three African nations (Reuters, 2003). We can also consider the inequality in other areas outside of income. In agriculture, for example, the richest 1% own 44% of all arable land in the rural areas, with 50% of the farmers toiling on just 3% of the land (Todaro, 2000).

It is impossible for us identify any single reason for this inequality - a variety of explanations are plausible. Inflation exacerbated the problem of income distribution according to the Human Development Report for Brazil in 1996, as well as contributing to other problems such as growth and employment levels (UNDP, 1996). Inflation has been a persistent problem for Brazil's economy. In 1993 for example, staggering levels of hyperinflation at over 2,700% were reached before it responded to policy changes in 1994 (Todaro, 2000). As has been previously mentioned, the pursuit of growth over development strategies may have contributed significantly. Whatever the cause, we know that inequality has been a persistent trend in Brazil's economy for a number of years, and one that has caused widespread poverty. In Brazil, poverty is prevalent in a variety of forms. I will

describe where it exists, and the role that inequality has played in causing these types of poverty.

Growth and income per capita

Our first concern rests with growth for the country. As has been noted, even when the economy has been successful in terms of growth a minority, who have been able to make huge profits, from low industrial wages, has largely reaped the rewards. In effect, the periods of rapid growth have created a situation whereby the richest have become even richer, and the poor have found their struggle as difficult as before. More recently, productivity levels for the Brazilian economy have suffered due to the disparity in income distribution. Between 1985 and 1994, it experienced one of the lowest real incomes per capita in all of Latin America, averaging at –0.4% (Todaro, 2000). It would seem that the problems of growth and income distribution have fed into each other, creating a vicious circle. In other words, the disparity has contributed to poor growth levels, which have in turn further exacerbated the imbalance in income distribution. The 1996 Human Development Report for Brazil supports the argument that a negative link exists between poor growth and income inequality: "the interruption of the economic growth curve in the 1980's caused the increase of inequity and poverty, accentuating the historical tendency towards the concentration of income" (UNDP, 1996).

With the stagnation of growth in Brazil, income per capita has dropped. Even recent changes have been significant, with GNP per capita at $4,270 in 1997 dropping to $3,580 recently (World Bank, 2004). There was a fall in the growth of real wages between 1980 and 1991 of –2.4% per annum (Mankiw, 2001). Those on lower incomes and middle incomes have been the worst hit. Therefore, this income inequality has had a direct impact on poverty. Forty million Brazilians live on $1 a day or less, a situation commonly called 'absolute poverty'. Seventy million people (40% of the population) live on $2 a day or less (Todaro, 2000).

Unemployment

It would be misleading to think that poverty can only be thought of in terms of income per capita. It can relate to other areas; we can think of poverty as a form of severe underdevelopment. In Brazil's case, pervasive inequality has caused poverty in the employment sector. The unemployment rate itself is a problem, with open unemployment currently at 7.1% (Banco Central do Brasil, 2004). Unemployment is particularly high in densely populated urban areas. However, this does not reveal the extent of the problem. With so many people earning so little of the country's income, the nature of the work itself shows a form of poverty. In the cities, there are not nearly enough vacancies to absorb the high proportion of the work-seeking population, thus many people resort to activities such as street

vending, drug peddling and prostitution. In Sao Paulo, 43% of all employment is in the 'informal' sector, meaning self-employed jobs such as hawking, junk collecting and personal servants (Todaro, 2000). Those in the industrial sector are often exploited, working on a very low wage, but with work in such short supply, they cannot afford to leave their jobs. In the rural areas, the working conditions are sometimes appalling. Alan Thomas describes those of one particular area; the open-cast Serra Pelada gold mines in Northern Brazil: "Labourers…have a regular though very low income, working a 12-hour day carrying 120 lb sacks up 200-foot ladders with considerable risk of injury or work-induced illness" (Thomas, 2000). Again, inequality feeds into the problem in the employment sector, which further widens the income gap between the rich and the poor.

Living Conditions

Poverty exists in Brazil with respect to living conditions also. This is particularly true in the large cities, where the urbanization process has resulted in the growth of huge slums and shanty-towns. For example, in Rio de Janeiro the 'favelas' (the colloquial name for these types of slums) account for 60% of the total urban population. These settlements generally lack sewage systems, electricity and clean water. Of course, not just the large cities are affected. In Rio Branca, a town that lies in the province of Acre; in 1991 40% of urban homes were without running water and 88% had no sewage pipes. The town also serves as a useful example to illustrate the problem of rapid urbanization. In 1970, the town's population stood at 36,000 people. In the following 30 years, this had increased to over a quarter of a million (Thomas, 2000). With such impoverished conditions prevalent, there has been little chance for the development of Brazil's people. Again, it must be emphasised that inequality has been the driving force behind the existence of such conditions.

Other causes of poverty

This is not to say that the cause of all poverty-related problems can be attributed to inequality. Population growth, for instance, can equally have an impact. The population growth rate may not seem like a problem in itself; in fact the figure of 1.4% in 1998 is quite respectable when compared to other lesser-developed countries. Nevertheless, it has meant a population increase of over 10 million people in less than five years; 162.1 million people in 1998 to 172.6 million people today. Such growth makes it increasingly difficult to maintain a steady growth of income per capita, and the challenges of development become even trickier. Similarly, the huge debt amassed, twice that of the United States, is a major problem, as Mark Weisbrot pointed out in a recent BBC news report. In the long run, the task of solving the problematic issues of poverty are made more difficult still, with a

constant outflow of money to pay off the debt. However, inequality is a central link to many of these inter-related problems, and provides the main obstacle to Brazil being able to achieve sustained development for its people.

Solutions

Solutions to problems of such magnitude are certainly not simple. It is important for us to recognise first that the problems must be dealt with in a way that is specific to Brazil, so that it takes its history, culture and general way of life into account. It is important for us to avoid the mistakes of other models, which "ignored history, lumped diverse nations and cultures into one homogenous category, assumed common goals and values…and generally treated undeveloped countries with lofty contempt" (Quinn, 1988). With this caveat in mind, we can divide our solutions into two categories: how to enhance the potential of Brazil, and how to alleviate the problems of poverty.

Brazil has enormous potential in a wide range of areas. Industrial development is particularly significant in Brazil's future. With many of the core components for a thriving industrial sector, including an energy source, natural resources, and a large potential workforce, the possibility for it to prosper is certainly there. However, other areas would need investment, for example in agriculture, so that raw materials could be available to industry, and in education. The benefits would be to provide employment for people, as well as increasing productivity. The government could possibly use tax incentives and the cheap cost of labour to encourage investment in industry from abroad. The Human Development Report of 1996 identifies other possible advantages Brazil has, such as favourable demographics, the business and technological training acquired, and the dimensions of the internal market (UNDP, 1996).

The development of industry would aid in alleviating the various problems of poverty. However, this alone would not be sufficient, especially with such a deeply rooted culture of inequality. The government has an important role to play in shifting the balance. The inequality would seem to be both the cause and result of poverty in Brazil - for living standards, the employment sector, and growth. Therefore, by tackling poverty itself, the problem of inequality may be partially solved. A decentralisation project, moving people from cities into rural areas may be appropriate in order to prevent over- crowding and squalid conditions in urban areas. Industries could be encouraged to base themselves in rural areas in order to create jobs. Investment would be required from the government to provide clean running water and sewage systems. Whatever specific action is taken, the aim should always be to improve the quality of life for Brazil's people. One important step would be to

ensure some sort of minimum wage. This would automatically take money away from those profiting most and give it to the poorer people. However, such steps must be taken with great caution, as they can lead to further unemployment.

Conclusion

In many ways, Brazil has reached a crossroads in its development process. With the recent inauguration of the first working-class president, Mr Luiz Inacio Da Silva, there may well be a further concentration of efforts on development. Indeed, the president stated that his goal upon his appointment was for every Brazilian to have 'three square meals a day'. With careful management of this development, including a head-on tackling of the problems of poverty, and a shift in the balance of the embedded culture of inequality, Brazil might finally be able to fulfil its recognized potential and become a dominant figure on the world's economic stage.

Bibliography

Banco Central do Brasil, (2004). *Special data Desmenation Standard*, [Online]. *Brazilian Government.* Available from: http://www.bcb.gov.br/sddsi/sddsi.htm

Colman, D and Nixson, F. (1986) 'The Concept and Measurement of Development' In: Farmar, A. (ed.) *The Developing World: An Introduction to Development Studies Through Selected Readings.* Dublin: Development Education Support Centre.

Goulet, D. (1983), "Obstacles to World Development: an Ethical Reflection" *World Development,* Vol 11(7).

Mankiw, N.G. (2001), *Principles of Microeconomics (2nd edition).* Texas: Harcourt College Publishers.

McAleese, D. (2001), *Economics for Business, Competition and Globalisation.* Pearson Education Limited: Essex

McAleese, D., and Burke., D. (2000), *Policy Objectives for a Regional Economy.* In *The Economy of Ireland: Policy and Performance of a European Region*. 7th ed. O' Hagan, J.W. (ed.) Dublin: Gill and MacMillan.

McCarthy, D. and Spencer, K. (1992), *Modern Economics for Irish Students.* Dublin: Mentor Publications.

Quinn, R., (1988), In: *Introduction* to *The Developing World: An Introduction to Development Studies Through Selected Readings.* Farmar A. (ed.) Dublin: Development Education Support Centre.

Reuters, (2003), "Brazil Gets First Working Class President". *Irish Times.* 02/01/2003.

Thomas, A., (2000), *Poverty and the End of Development*, in *Poverty and Development into the 21st Century.* Oxford University Press: Oxford

Todaro, M.P., (2000). *Economic Development.* 7th edition. Essex: Addison Wesley Longman.

UNDP, (1996), "*1996 Human Development Report for Brazil*" [Online]. *UNDP* Available from: <http://www.undp.org.br/HDR/Ddr96/hdr1-6.htm>

World Bank, (2004). "*Brazil Data Profile"*. [Online]. *World Bank.* Available from: <http://www.worldbank.org/data/countrydata/countrydata.html>

World Bank, (1999). "*World Development Report 1998/1999"*. [Online]. *World Bank*. Availible from: <www.worldbank.org/wdr/wdr98>

CHINA AND ITS DOLLAR PEG – THE TRUE SOURCE OF GROWTH?

BY MARIA BRIDGEMAN

Senior Sophister

In recent years, the Chinese economy has been growing rapidly and attracting attention of researchers worldwide. However, its policy of fixing Yuan to US Dollar has attracted criticisms from one of China's main competitors in the world market – the USA, demanding appreciation of Yuan. This essay by Maria Bridgeman makes an enquiry into the validity of these allegations. The author uses the Mundell-Fleming IS/LM/BP framework to analyse the situation, assesses the role played by the peg in this country's economic expansion, examines the theory of pegging in the context of other countries' experiences and presents her arguments for sustaining the peg, denying the American critique of this policy.

Introduction

US manufacturers have long called on the US government to put pressure on China to revalue its Yuan peg. They argue, due to high growth and rapid accumulation of foreign reserves (US$346.5 billion by the first half of 2003) that the Yuan should be appreciated from its current band of Renminbi (RMB) 8.276-8.28 per US dollar. They feel the artificially weaker Yuan is giving Chinese competitors an unfair advantage in the global markets, as well as flooding the US market with cheap Chinese imports. This has resulted in the US having a bilateral trade deficit with China of US$103 billion in 2002 (up US$20 billion from 2001). This has had an effect on employment with US unemployment figures up by about 2% on their 1990s averages. The dollar has weakened by 1.8% against the Japanese Yen and by 9.6% against the Euro in 2003 to September.

The Chinese, though realising the long run need to loosen the bands on its currency feel that for the moment the peg is justifiable. They argue the basis of China's competitive advantage is not its cheap currency, but factors such as labour costs, technology and infrastructure. I believe China has been used as a scapegoat for the global economic downturn. There is already fierce speculation about a policy change with likelihood of currency speculators betting on the appreciation. This

could result in a financial bubble that could prove detrimental. Monetary stability is another argument for changing China's currency policy. China's domestic credit rose by 38% of GDP last year, sparking off numerous credit scandals, many of which are still unfolding. To combat this China's Central Bank has declared its intention to develop a market for its short-term instruments, and has raised the level of reserves its banks must hold, which would act to stabilise monetary growth. The US market happens to determine global prices for tradable goods. Pegging its currency to the dollar minimises the risk to suppliers based in China. The absence of currency risk is a major factor in the rapid redistribution of global production capacity to China. Thus, China feels it should stick with the peg until inflation returns.

This debate involves many parties involved with conflicting interests, so there is no clear answer at this stage. I aim to look at the role of exchange rates, drawing on Mundell's work. International experience in currency pegs will also be analysed to see if there are any lessons to be learned. The Chinese situation must then be broken down to assess the true drivers of the blistering growth it has been experiencing of late. Is the US using China as a scapegoat and being somewhat neomercantilist in over emphasising trade deficits? Are the Chinese pushing their luck or is there real risk of crisis? I will endeavour to analyse the arguments on both sides and look to other pegging experiences to reach a conclusion on the best way forward for the global economy as a whole.

Mundell's exchange rate economics:

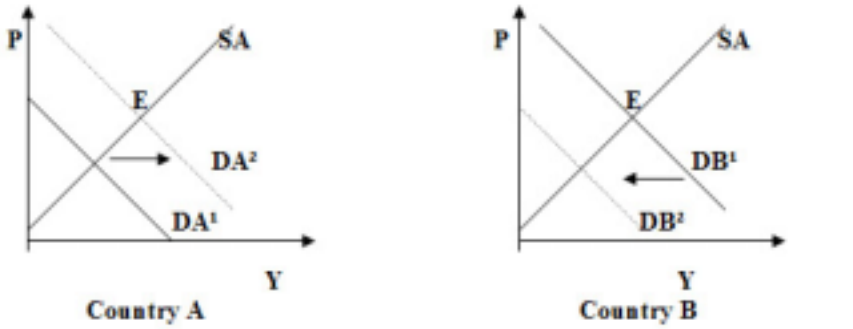
Nobel prize winner Robert Mundell (1961) is best known for his paper on optimum currency areas and has been dubbed 'the grandfather of the euro'. His thesis set out to decide whether it is preferable for countries to operate under fixed or floating exchange rate regimes. This developed into an advocacy of floating rates based not around national boundaries, but around regions known as optimum currency areas. These regions would have sufficient wage and price flexibility or labour mobility to replace the exchange rate as an adjusting mechanism.

This theory is illustrated with the simple model of two regions in which there is a shift in aggregate demand for goods from one to the other (in Figure 1 below from B to A). The demand curve shifts outwards for A and inwards for B. Both countries are moved from their initial full-employment equilibrium point E. At full-employment the increase in demand in A creates upward pressure on prices and wages. If A were to fully absorb the inflationary pressures of the increase in demand, B would quickly become more competitive causing an increase in aggregate demand and restoring equilibrium. However, the tendency is for A to resist a rise in the price level, resulting in a recessive tendency on B (as prices are generally inflexible downwards). The result of this will be a current account surplus in A coupled with

moderate inflation; while B on the other hand will experience a current account deficit and unemployment.

Mundell cites wage and price flexibility and labour mobility across countries as possible ways of restoring equilibrium, as well as a complex fiscal transfer system, where A would finance new domestic demand in B through a transfer of tax receipts.

FIGURE 1: SHIFT IN AGGREGATE DEMAND FOR GOODS FROM COUNTRY B TO A



Mundell (Ibid.) argues that demand shifts from the products of B to the products of A, a depreciation of country B or an appreciation of country A would correct the external imbalance and also relieve unemployment in country B and inflation in country A. This is the most favourable case for flexible exchange rates based on national currencies.

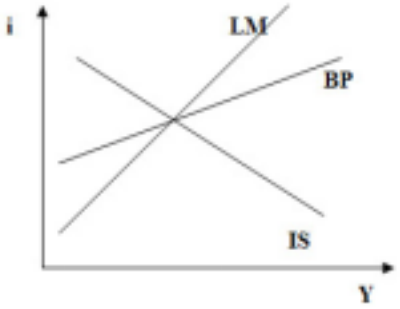
For countries to operate under fixed exchange rates or more extremely in an optimal currency area, the above adjusting mechanism would not be possible. Mundell (Ibid.) made two requirements for fixed exchange rates to work:

- High factor mobility between the countries or within a region.
- Similar experiences of economic shocks between countries or within a region.

The similar economic shocks would reduce the shifts in demand and the high factor mobility would play the adjusting role, when needed, in place of exchange rates. These points he used to pre-empt and later promote the European Monetary Union, which I will discuss in more detail later.

One of the most significant advances made by Robert Mundell was the extension of the standard workhorse of macroeconomics, the IS-LM model of the Hicks-Hansen synthesis to an open economy. This became known as the Mundell-Fleming model.

FIGURE 2: THE BASIC MUNDELL-FLEMING IS/LM/BP MODEL



The basic structure of this model is to divide the economy into three markets and study how they interact. The Goods market (IS), the Money market (LM) and the Balance of Payments (BP), which shows how an economy connects with the rest of the world.

The IS curve represents all points at which the goods market is at equilibrium given various combinations of interest rate (i) and national income (Y). It is derived from the investment function, which relates interest rate to capital stock in the economy, and the aggregate supply - aggregate demand framework for the closed economy. A decrease in the interest rate will cause investment to increase, which results in a higher level of national income, and a shift down the IS curve.

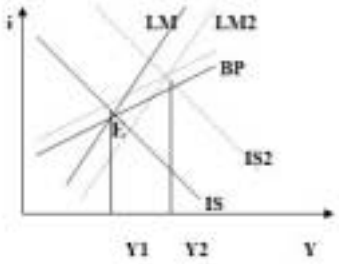
The LM curve represents all the points where demand for money is equal to the given and fixed supply in relation to interest rate and national income. The transaction demand for money consists of the active working balances held for the purpose of making business payments as they become due, and is positively related to national income. The speculative demand for money arises from the desire to hold money balances instead of interest bearing securities, the higher the interest rate, the smaller the speculative demand. The interaction between aggregate supply and aggregate demand, and the supply and demand for money is observed by an upward sloping LM curve.

The BP curve represents the various combinations of interest rates and national income at which the nation's balance of payments is in equilibrium at a given exchange rate; any trade deficit is matched by capital inflows or vice versa. The BP curve also depicts the elasticity of capital flows, as the flatter (more horizontal) the BP curve is, the more elastic the capital flows. The BP Curve is derived from the interaction of the goods market (imports and exports), capital flows (to neutralise trade imbalances) and foreign exchange rate (supply and demand for foreign currency). A depreciation of a nation's currency shifts the BP curve to the right; likewise an appreciation in currency would shift BP to the left.

In Figure 3 I depict a BP curve that is in equilibrium. If the IS and LM curves intersected to the left of BP, there would be a trade surplus, likewise if they intersected to the right there would be a BP deficit.

In a situation of flexible exchange rates at a point of unemployment, starting at point E where all three markets are in equilibrium with an external balance and unemployment, the nation could use easy monetary policy to shift the LM curve to the right. The nation has an external deficit, with flexible exchange rates, the nation's currency depreciates and this causes the IS curve to shift to the right and the BP curve to shift to the left until IS/LM/BP intersect. If this point is still below full employment, this procedure can be repeated until full employment is reached.

FIGURE 3: IS/LM/BP UNDER FLOATING EXCHANGE RATES



Under fixed exchange rates the depreciation shown above (or alternatively appreciation) could not happen and any expansionary (or contractory) monetary policy will not be effective to increase employment.

Mundell's work will form the background of my analysis of currency pegs and particularly, the role of exchange rates in promoting stability and competitiveness.

Pros and Cons of Currency Pegs

Having examined the theory of exchange rates and pegging; I will now examine the case for and against using a currency peg to promote stability in an economy. Although most developing countries' pegs have failed this could be because of timing, rather than the actual peg mechanism. I think time has shown that currency pegs can indeed serve a purpose, but have historically been implemented for longer than needed.

Honohan and Lane (1999) outline the pros and cons of the exchange rate stability provided by a currency peg or as they call it 'tracking.' They argue that although both anchor country and its trackers enjoy most benefits, there are at least two sources of asymmetry:

- If one zone is smaller than the other, exchange rate stability will be more important to the smaller, since it will be proportionally more heavily dependent on international trade.
- If one is a net creditor to the other, changes in the exchange rate have transfer effects between the two regions in the sense of adjusting the value of outstanding debts, posing risks to the financial system and/or debtors.

First, they outline the adverse effects of exchange rate uncertainty, namely discouragement of investment and the high price setting of risk-averse firms in an uncertain environment. This leads to a reduction in international trade as domestic transactions are less risky and this in turn leads to less growth.

They draw on Sapir, Sekkat and Weber's (1994) empirical work on the difference between the effects of short-term versus long-term variability in exchange rates. This can be quite substantial due to the use of hedging to combat short-term volatility and the power of longer-term misalignments to affect FDI and portfolio investments. Anchoring a currency gives less room for bubbles to develop on expectations of appreciation.

Honohan and Lane (1999) then list other positive effects such as pro-competitiveness through greater price clarity, reduced risk for foreign investors when buying domestic assets, creating a more liquid domestic capital market, reduced risk for domestic buyers of foreign assets, which would in turn improve competition in the credit market and provide greater opportunities for a wider range of financial contracts to be traded, which would improve risk management. These effects, as pointed out by Obstfeld (1994a) can contribute to a higher average, if more variable long-term growth rate, due to more high risk-high return projects being undertaken.

Honohan and Lane (1999) then examine the possible dangers of pegging and why most pegs eventually collapse. Inappropriate fiscal and monetary policies can cause misalignment. Unsustainable pegs can temporarily survive through adequate reserves, capital controls and excessively optimistic market expectations. International capital mobility also increases the probability of speculative attacks weakening the pegs sustainability. Argentina's pegging experience illustrates a peg that was kept long after its purpose was served. The Asian financial crisis is a blatant example of how international capital mobility can cause destruction.

The Role Of The Dollar Peg In China's Steaming Economy

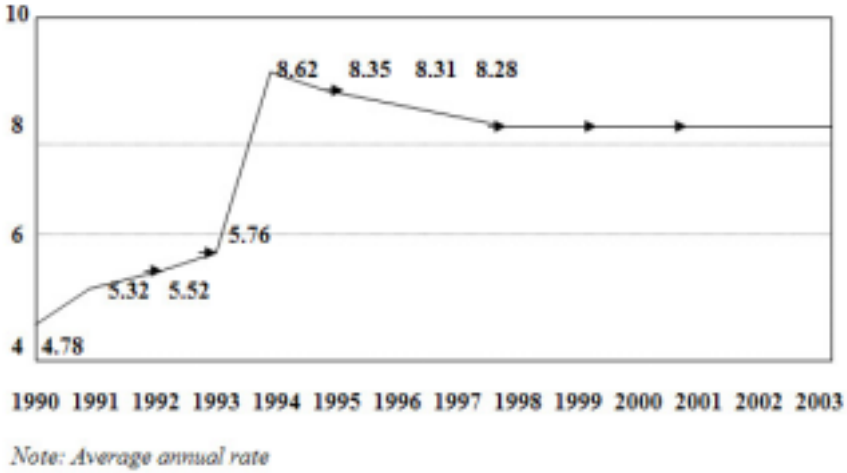
Prior to 1994, a dual rate system existed between China and the US, in which a 'market' and 'official' rate existed side by side, with most transactions, particularly for state-owned enterprises, being settled at an official rate. At

approximately RMB 5-6 to US$1, the official rate was an overvaluation, which benefited Chinese importers. This dual system was abolished in 1994, with the rate being decided at RMB 8.5 per US$1, appreciating to RMB 8.28 to US$1 in 1997, staying in the band of RMB 8.276-8.28 per US$1 since. Harner (1999) attributes the appreciation in 1996 and 1997 to:

- China's strong current account surpluses, driven by net exports;
- A strong inflow of FDI funds; and
- The requirement for Chinese companies to sell foreign exchange to the banks.

The Peoples Bank of China (PBOC) has tried to control this, as it does not want to see pronounced appreciation of the RMB.

FIGURE 4: TRENDS IN RMB/US$ EXCHANGE RATE:



Note: Average annual rate

Coupled with this firmer peg was a commitment to reform the banking sector, which had been characterised by lack of independence from local government and the influence of local and national officials over lending decisions and resulting bad loans particularly given to the weak SOE sector.

The Asian financial crisis of 1997 was caused by countries such as Thailand, Indonesia, Malaysia, the Philippines and South Korea that had their currencies pegged to the dollar, running current account and net external asset deficits. Speculators picked up on this and attacked the currencies and stock markets

of these countries. Subsequent devaluation caused these countries to seek financial assistance from the IMF.

China, although not in the same danger as these countries owing to its huge foreign currency reserves, positive international payments position and safer international debt levels, was worried by this and put even greater effort into accumulating reserves and cleaning up its banking system post-1997, its three main objectives being:

- Cleaning out bad debts and recapitalising the banks;
- Reorganising and downsizing bank organisations; and
- Modernising and improving management and accountability.

China's CPI growth, after months of being in negative territory, started to rise in October last year. The index's growth hit 1% in April, but subsided afterwards. It registered 0.5% in August.

The PBOC estimated that the 1% point hike in required reserves implemented in September will freeze 150 billion Yuan (US$18 billion) in base money, a term that includes cash in circulation and all reserves at commercial banks.

It is widely estimated that this number, multiplied with a currency multiplier, which stands at a little above 4 currently, may result in a potential total contraction in money supply of as much as 600 billion Yuan (US$72 billion) (People's Daily, 2003). This will stem worries about overheating of the economy due to a growing money supply fuelled by huge dollar reserves.

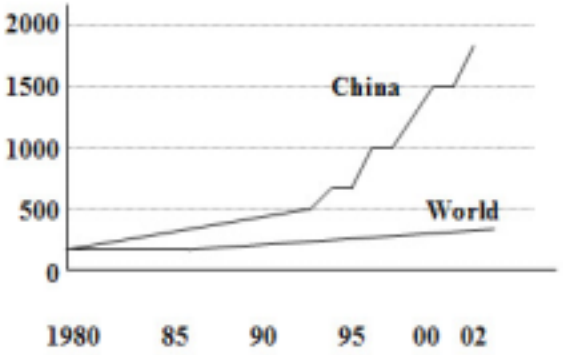
In the period 1992 to 2002, China's real GDP growth has averaged about 10%. This has been in part because commitment to these and many other reforms, in part because of the labour costs, technology, quality control, infrastructure and of late in part because of the downward slope of the dollar; giving it extra competitiveness with its parallel devaluation. In particular with the US, China has a competitive edge of cheaper manufacturing and labour costs and hence cheaper prices for its exports, which are replacing domestic US goods at a fast pace. There has also been significant relocation of manufacturing facilities out of the US and into China. The proportion this undervalued currency has played relative to the other instruments in China's blistering growth has caused much debate, as I will discuss in the next section.

US Woes:

The huge US current account deficit and subsequently weakening dollar has cast a disproving eye on China, whose bilateral trade surplus with the US is $125 billion (The Economist, 2003b), the biggest US bilateral trade deficit worldwide. It is claimed China's 'unfairly' low currency and 'unfair' trading strategy are the cause of the jobless recovery, the US is currently experiencing. The

price adjustment mechanism, which brings countries back into balance of payments equilibrium with competitive exports as a result of devaluation, has not been able to happen bilaterally with China due to the peg. The US debate over the Yuan peg has been much politicised and publicised. We have seen mass meetings of new anti-China trade groups such as Wisconsin's 'Save American Manufacturing' and 'Mad as Hell' in Connecticut, on their self-proclaimed Manufacturing Awareness Day on the 1st of August 2003. Banners waved 'Leave China in the sink' and 'Free trade is a myth' in such protests fuelled by anger at the activities of big firms buying from Chinese firms and/or moving manufacturing to China (The Economist, 2003b). The 2.7 million manufacturing jobs lost in the US (one in six) in the past three years lead to increasing political pressure from such grass roots groups for protectionism and tariffs, especially since China's recent WTO ascension. As China's importance in the world economy grows, (see Figure 5 below) these fears have grown.

FIGURE 5: MERCHANDISE EXPORTS, $ TERMS 1980=100



Part of China's terms of ascension to the WTO was 'safe guards' for other countries to have the option to impose tariffs on Chinese goods if they are deemed to be causing 'market disruption'. The US succumbed to the political pressure for protectionism, using one of these clauses to impose protection on $500 million worth of Chinese textile imports. There are also half a dozen bills lurking around in congress, with the intention of imposing more tariffs if China does not revalue.

The US critics argue that after years of fast growth and huge inflows of foreign investment, the Yuan should be stronger. Instead, as the Dollar has fallen steadily over the past 20 months, the Yuan has fallen and is clearly undervalued. According to economic theory, Mundell's being testament to this, exchange rates should be appreciating in countries with rapid productivity growth. China's economy has been growing much faster than the rest of the world and its current

account has been in surplus since 1994. It is argued for a fast growing, emerging economy with high levels of investment, a current account deficit is more normal. In 2002, China was the largest recipient of foreign direct investment, with inflows totalling $53 billion. This is also a factor that should theoretically push up a currency's value. China's huge build up of reserves are the final factor that suggests an appreciation would be the order of the day.

The Economist documents various estimates of the degree of overvaluation. Its own informal 'Big Mac Index' estimates the Yuan as being undervalued at a whopping 56%; Ernest Preeg, of America's Manufacturers alliance estimates the Yuan should rise, according to market forces by 40%; the bank UBS estimates the figure at 20%. Other estimates can be as low as 10-15%. The tight capital controls China enforces, as well as its state of restructuring makes it difficult to use traditional measures and hence the variation of estimates (The Economist, 2003a).

Arguments For Sustaining The Peg

Roach[1] (2003) speaks of the irony of the growing Worldwide support for a revaluation. Chinese subsidiaries of global multinationals and joint ventures with industrial-world partners have accounted for 65% of the total increases Chinese exports from 1994 to mid-2003. A high-cost industrial world has made a conscious decision that it needs a Chinese-based outsourcing platform to increase productive efficiencies. He argues that dismantling the Yuan peg would destabilise the very supply chain that has become so integral to new globalised production models. He then moves on to argue China does not compete on the basis of an undervalued currency. Instead competing, mainly in terms of labour costs, technology, quality control, infrastructure, and an unwavering commitment to reform. He estimated, if China were to revalue the Yuan upward by 10% or even 20% (a change he does not expect nor advise), its exports would suffer minimal loss of market share. A key reason for this is that China's export prowess is mainly in the role of an assembler and its exports have a high content of materials and products made elsewhere.

It is true that the end game of China's reform is to open up its capital account and make its currency fully convertible, however the Chinese argument is that it is too early. An ill-timed revaluation could cause deflation just after China has pulled itself out of that value destroying process. Financial bubbles are also a threat that comes with revaluation, particularly in the housing market. It is clear that the road of reform is a long one and all the hard work could be easily wasted with an early revaluation. The instability that a revaluation would bring could cause damage to world markets, not forgetting China's role in the upkeep of stability in the 1997 Asian crisis.

[1] Of Morgan Stanley

Standard & Poor's warned on the 15th of September 2003, that confidence in Chinese banks was already so low that a heightened debt crisis could lead to their collapse. Paul Coughlin, Managing Director of Standard & Poor's Asia-Pacific Corporate & Government Ratings, said:

> "We learned from the Asian currency crisis in the 1990s that the combination of a weak banking system, floating exchange rates and free flows of capital can be a very dangerous combination. China's banking system is insolvent, with problem loans estimated by Standard & Poor's at 45 percent of total loans, and its risk control systems are ill-prepared to deal a rapid liberalisation of the exchange rate and capital controls." (Chan, 2003)

Mundell himself has been quoted in support of keeping the peg for the time being, stating in a speech in Taipei on 19th of September 2003 that revaluation of Yuan will be a disaster for China. He declared:

> "It would not solve the problems of the US, Japan and Europe. It would cut off the brightest light for global expansion in China. China has said 'no' and rightly so, that is the right position to take (Ibid)"

In an earlier comment in Beijing, Mundell predicted that the 40% increase in the Yuan being called for in some circles could halve China's annual economic growth, aggravate deflation and produce a financial crisis similar to that experienced in Asia in 1997 (Ibid.).

This might sound strange when you look at Mundell's criteria for fixed exchange rates. However, his model was based on developed countries with full employment. China is a developing (albeit rapidly) country with less than full employment and the currency peg is needed to promote stability in the short run. He believes that the US should not be allowed to increase its own employment as per his IS/LM/BP analysis at the expense of China, whose revaluation would be at the cost of great opportunity. This peg and subsequent surplus is giving China a chance to play a stable, important part in the global economy.

Conclusions and Policy Recommendations:

Having examined the evidence, it seems that a theory cannot be objectively applied to any situation. Mundell himself agrees with the sustenance of a currency peg that goes against the tenets of his most famous theory on exchange rates. However, its advantage is clearly seen when applied to his IS/LM/BP. I think this comes back to the difference between developing and developed countries and giving the developing ones a chance to reform and become developed. This gives a

new texture to a playing field that would otherwise be level between two equal players.

Looking at the huge US current account deficit, I, along with many others would blame an insatiable consumer goods and debt market coupled with high investment rather than a competitive Chinese economy. America has a serious lack of domestic saving. George Bush's economic advisor, Gregory Mankiw has argued that most American jobs have been lost in industries (machinery, transport equipment and semi conductors) where Chinese competition is slight (The Economist, 2003b).

I think China's commitment to financial reform and aim of opening up its capital account in a long term is commendable given the amount of bad debts in existence. I can see a possible movement of fixation to a basket of currencies that would include the Euro as the next step, but at least not before the Olympic Games in Bejing in 2008.

China has just escaped the painful process of deflation and to echo Mundell is emerging as one of the brightest lights in the world economy. The basis of its competitiveness is not an undervalued Yuan, but factors that lend to much longer-term economy that the world relies on in its manufacturing supply chain.

I conclude that the US is using China as a scapegoat for its self-imposed current account problems and subsequent weak currency. The Bush administration is acting in a blatantly neomercantilist fashion, over its balance of payments deficit and in turn using protectionist policies (I estimate the Dec tariffs are not to be the last, especially in an election year) to countervail its biggest threat that has ironically long supplied it with cheap imports and components.

Bibliography:

Chan, J. (2003), "US Pressure Continues Over Value of Chinese Currency" [Online]. *World Socialist Website*. October, 2003. Available from: <www.wsws.org>.

Honohan, P. and Lane, P. R. (1999), "Pegging to the Dollar and the Euro." *Trinity Economic Paper Series*, Paper No. 99/6

Mundell, R. (1961), "A theory of optimum currency areas." *American Economic Review*, Vol. 51 (4)

Obstfeld, M. (1994), "Risk Taking, Global Diversfication and Growth." *American Economic Review*, Vol. 84 (2).

People's Daily (2003), "Is China's Interest Rate Hike Looming" [Online]. 08/09/03. Available from: <english.peopledaily.com.cn/200309/08/eng20030908-1239321.>

Roache, S. (2003), *Getting China Right*. Morgan Stanley Research. Mimeo.

Sapir, A. et al. (1994), "The Impact of Exchange Rate Fluctuations on European Community Trade", *Discussion Paper Serie B*. Germany: University Bonn.

The Economist (2003a), "Flying on One Engine-A Survey on the World Economy." 20th Sept.-26th Sept.

The Economist (2003b), "Mr. Wen's Red Carpet." 13th Dec-19th Dec.

AN ECONOMIC ANALYSIS OF THE IRISH RETAIL PHARMACY MARKET WITH A FOCUS ON COMPETITION POLICY ISSUES

BY DAVID POWER

Senior Sophister

This essay by David Power provides a comprehensive analysis of the Irish retail pharmacy market. The author, starting from a brief market overview, proceeds to scrutinize the market structure, nature of the competition and existing barriers to entry. The paper is concluded by the critique of existing regulation in the market and suggestion of several reforms that would benefit Irish consumers.

Introduction

This study covers the provision of retail pharmaceutical services in Ireland. In particular, it examines whether the interests of consumers are best served by the current control of entry regulations. These regulations place restrictions on how and where contracts to dispense 'pharmacy-only' medicines in Ireland are awarded. The sector is clearly over-regulated, and entry is restricted to protect the profits of the incumbents. In this essay, I propose to demonstrate that deregulation will serve consumers interests, thus refuting the claim that restrictions are necessary to protect the markets and incomes of pharmacy owners.

The structure of this essay is to first, describe the retail pharmacy sector in Ireland and then to undertake an economic analysis of this market. The latter forms the core of this study and will involve a look at the nature of competition within in the sector, as well as the entry barriers that currently exist. The arguments for entry barriers will be discussed and rejected, and finally I will offer some suggested reforms to deregulate and then conclude.

The Competition Authority (2002) has recently investigated the regulatory situation in Ireland with respect of retail pharmacies and found that margins on private prescriptions in Ireland are the highest in the EU. The margins for a typical pharmacy outlet are 45% on private prescriptions, over double the margins of 19% on prescription medicines dispensed under General Medical Services (GMS). Bacon (1999) has shown that in a European context, these figures are even more staggering. Irish pharmacy margins on medicines, at 33%, were just under 1.5 times the EU

average (23.2%), and nearly 5 times the corresponding figure for the UK (7.5%). The reasons for such wide differences should be investigated in terms of regulation, competition policy and entry barriers.

Market Description

The continuing development of Irish retail pharmacy chains, the arrival of UK multiples, economies of scale, professional marketing techniques, and cost effective purchasing power has combined to put pressure on the traditional pharmacist to become increasingly commercially focused in their approach to community pharmacy.

Product Market

A retail pharmacy provides services in respect of four main categories of products:

1) Prescription medicines, i.e. a doctor's prescription is required for purchase;
2) Pharmacy-only Over-The-Counter (OTC) medicines;
3) Unrestricted OTC medicines;
4) A range of non-medicinal products; often referred to as cosmetic, toiletries and sundries (CTS).

Table 1 shows that estimates of nationwide turnover figures for each of the four categories indicate that the dispensing of prescription medicines forms the core business of pharmacies (61% of sales).

Table 1: Retail Pharmacy Sales in Ireland by Product Type

Category	Product Type	Sales €m 2001	Sales %
Pharmacy-only medicines	Dispensed medicines	702	61 %
	Pharmacy-only OTC medicines	179	16 %
Retail products	Unrestricted OTC medicines	18	2 %
	Cosmetics, toiletries and sundries	253	22 %
	Total	1,152	100%

Source: The Competition Authority (2002)

Pharmacy-only medicines distinguish a pharmacy from other retail outlets. The Pharmacy Act (Ireland) 1875 prohibits non-pharmacy retail outlets from dispensing prescription medicines. Unrestricted OTC medicines may be sold in any

type of retail outlet. However, only 10% of OTC medicines fall into this category. The total turnover of retail pharmacies in 2001 is estimated by at €1,152 million; for pharmacy-only medicines the corresponding estimate is €881 million.

Geographic Market

Demand from pharmacy customers tends to be local in nature. A number of factors influence the consumer's choice of retail pharmacy, including its location, price competitiveness, product range, and quality of service. This market is therefore hard to generalise, as it depends on location and varies in size and scope for each pharmacy. However, work by Hoban (1999), presented in Table 2, identifies three groups of pharmacies separated according to profitability.

Table 2: Types of Pharmacies in Ireland 1996

Class	Location	Net Profit Margin	Turnover	Number of outlets	% of Total Profits
A	Prime retail locations	15% +	£700,000 +	250	50%
B	Cities/ Large towns	7-15%	£400-700,000	550	40%
C	Rural	0-7%	£0-400,000	400	10%

Source: Hoban (1999)

The first, Group A, includes pharmacies in prime retail locations where economies of scale have enabled the net profit margin to grow to over 15% of turnover. They account for a disproportionate amount of total industry profits (50%), and are thus very attractive as commercial entities. Group B is the most numerous category and are generally located in large towns or cities or are the sole trader in smaller towns. Class C earn low or zero profits, and in effect provides a social service and a cost efficient distribution method for the Department of Health and Children.

Market Structure

As at March 2001, there were 241 hospital pharmacies and 1,202 retail pharmacies in Ireland (TCA, 2002). Retail pharmacies have traditionally been small, independent, single-location operations, but chains of pharmacies are becoming a more prominent feature of the Irish market. In 2001, 32% of retail pharmacies in Ireland belonged to a chain of two or more outlets. Most outlets (90.4%) are owned by pharmacists (Bacon & Associates, 1999).

Table 3: Ratios of Community Pharmacists in Europe (1993)

Country	Inhabitants per Community Pharmacy	Inhabitants per Community Pharmacist	Pharmacists per Pharmacy
Austria	8,000	2,224	3.6
Denmark	17,000	4,525	3.8
Ireland	3,070	2,800	1.1
Sweden	11,600	11,600	1.0
United Kingdom	4,810	2,741	1.8

Source: Based on MacArthur, D. (1995) *The Growing Influence of the Pharmacist in Europe: opportunities in a changing market*. Table 1.3. London: Financial Times Publishing

Concentration

Table 3 shows that concentration of pharmacies varies throughout Europe. The result is that Ireland ranks 5th in Europe in terms of the number of pharmacies per head, meaning that access is relatively good for consumers. Concentration levels are low, with the largest chain holding 4.3% of the total market. However, this figure is not particularly important, as it is the local market and not the national market that is the relevant focus.

Economic Analysis

It is clear that regulations restrict who may operate a retail pharmacy and where the pharmacy may locate. Adopting the approach of the Office of Fair Trading (1997) in the UK, this investigation is guided by the principal that competitive markets to which there are no barriers to entry generally serve best the interests of consumers. Therefore, I will consider whether consumers will be better served if the highly regulated and closed shop pharmacy sector was to be deregulated. However, I will also remain mindful of the public policy objectives of the government.

The Nature of Competition

Price competition and location competition between retail pharmacies has been regulated to a very large extent, therefore non-price competition is strong. Four dimensions of competition are usually identified (Ibid):

1) Pricing
2) Geographical location
3) Product selection
4) Level and quality of retailer service

Price Competition

Normal forces of supply and demand do not apply. 61% of sales come from dispensed medicine, which requires a doctor's prescription. Also, the GMS scheme set prices at wholesale levels. However, private customers are charged the wholesale price with a 50% mark-up rule plus the dispensing fee. This horizontal agreement is not explicit and so may vary, however pharmacists implicitly collude through the Irish Pharmaceutical Union (IPU) to agree to this general mark-up rule.

Location Competition

As outlined earlier, demand from pharmacy customers tends to be local in nature, as customers are generally not willing to travel long distances to obtain medicines. While there is no location restriction preventing a registered pharmacist from opening a pharmacy without a GMS contract, most pharmacies would consider a GMS contract essential. The 1996, 'location restrictions' effectively capped the number of pharmacists in each locality. In 2002, their removal was welcomed.

Product Selection

In recent years, the range of products available has increased dramatically, as cheaper generic substitutes for prescription medicines have entered the market. Secondly, the increasing entry of European wholesalers such as Gehe has provided access to a larger distribution system of more products. Thirdly, there is also a trend away from semi-ethical prescription medicines towards OTC self-medication medicines. This widens consumer choice, and also favours the broadening of distribution outlets at retail level to include supermarkets, etc. thus forcing manufacturers to become more customer-focused.

Quality Competition

Quality of care is an important aspect of the service that pharmacists provide as community health care professionals. Customers' value has come from a strong pharmacist-patient relationship that has been effectively provided by independent Class C pharmacies up to now. However, the entry of new chain retailers such as Boots has not resulted in reduced level of service, as they recognise the importance of quality competition in the market. Indeed, it could be argued that they have increased dynamic efficiency by introducing new innovations such as late evening and weekend opening hours.

Entry Barriers

Typically, entry restrictions to any market result in prices being higher, innovation lower and quality of service poorer. A considerable body of regulation surrounds the education, registration, employment and professional practice of pharmacy in Ireland. This regulatory framework began with the Pharmacy Act (Ireland) 1875, which set out that only Pharmaceutical Society of Ireland (PSI) registered pharmacists would dispense prescriptions. The Health (Community Pharmacy Contractor Agreement) Regulations (1996), '1996 Regulations', further restricted competition in the sector, by effectively limiting the number of GMS dispensing pharmacies through 'location restrictions'.

There are two main arguments the IPU and others advance in support of regulation. However from a competition policy perspective these restrictions are detrimental to consumer welfare.

(a) The 'Quality of Care' Argument

The main rationale put forward for the restrictions on competition introduced in the 1996 reforms was that it is necessary to protect the markets and incomes of pharmacy owners. This is expressly stated in Regulation 2(1): a new pharmacy should "not have an adverse impact on the viability of existing community pharmacies in the area." The aim of this legislation was to allow pharmacists to develop the quality of the service they offer as community health care professionals, in accordance with the long-term strategy of the Department of Health and Children. However, it does not follow that quality will be improved by ensuring that there is no local competition. Bacon (1999) proposed that the most appropriate way in this case to improve quality is not by restricting competition but to regulate for quality.

From a competition policy perspective, regulating for quality is appropriate as it can deter 'free-riding' by pharmacies that fail to meet the Department of Health and Children's requirements for pharmacies to deliver a quality service as community health care professionals. Pharmacists that fail to provide these pre-sale services could increase profits by free-riding on the quality certification that the other pharmacists would confer on OTC medicines and CTS products. A community pharmacist is providing a service, not just a product. Thus, pharmacies must provide a whole range of advisory services on a daily basis, as well as delivering the various GMS schemes, and not just dispense medicines. It is difficult to regulate for quality, but one possible way is to grant or renew contracts subject to an inspection or audit of services provided. This would help achieve the National Health Strategy goal that a close on-going relationship between providers and users of primary health care services is important.

Another reason why services may not be reduced is that competition among pharmacies is not price driven, but focuses on non-price competition. Of vital importance to this is the level and quality of services available. Therefore, as Gehe have outlined (TCA, 2002), service levels at multiples may in fact be higher. Rigorous staff training in customer care was provided, and considering that over 90% of managers were pharmacists in Gehe pharmacies, the level of care is not compromised. Secondly, the entry of multiples and resultant increase in competition improves dynamic efficiency. It has served as a catalyst to improving service and introducing innovations. Longer opening hours, weekend opening and home deliveries are some examples of innovative measures that have improved customer service.

(b) The 'Protecting Strategic Interests' Argument

A second argument identifies an interest of society in general arising from the protection of the ownership structure currently existing (Hoban, 1999). Basically, this relies on an argument that locally owned pharmacies are in the interests of society. However, this argument does not make economic sense and has not been proven. It is not clear, what, or how great, the benefits from having profits accruing to shareholders of smaller businesses (i.e. independent pharmacists or small chains) rather than shareholders in larger companies (i.e. larger multiples), will actually be. The fear is that larger pharmacies will fall into the hands of highly efficient operators and will crowd out smaller operations.

However, the 1996 regulations have only served to make the already viable pharmacies even more valuable. The IPU's argument that smaller 'Class C' pharmacies in more rural areas would be under threat and possibly disappear is unfounded, since no one would wish to start up a pharmacy to compete in such a locality anyway. They seem to present 'Class C' pharmacies as providing a social service to customers and operating at a loss, whereas it is clear that no business would continue to operate at a loss in any area. In fact, it has been the larger 'Class A' pharmacies that have become even more profitable, because competitive entry was restricted. So, despite the restrictive entry barriers, multiples will be attracted, rather than deterred, to the market, due to these high profits.

Principal-agent concerns are also relevant in this case. Opponents of deregulation would also argue that restrictions should be retained, as the growth of multiples will result in services being reduced. Small, independent, pharmacist-owned and run pharmacies have a long-term commitment to the locality, whereas financially motivated multiples may not demonstrate this concern. I believe there is merit in this argument as employees in multiples may be faced with principal-agent issues. Since they do not own the pharmacy, there is less of an incentive to develop strong client-customer relationships and there is likely to be a higher turnover of

professional staff. However, I believe that a partial solution to this problem is to place limits on ownership to maintain local ownership by pharmacists.

The Consumers Association of Ireland (CAI, 2000) challenged the 1996 regulations, by showing that while they may have been well intentioned, they have had the opposite effect. The regulations introduced unnecessary barriers to entry and resulted in a closed shop. They created monopoly power for incumbents, which I believe is anti-competitive and protectionist at its worst, as it is not in the best interests of consumers. On foot of the CAI's challenge, the regulations were finally dropped in January 2002.

Suggested Reforms

To improve the competitive situation of the market, I believe the government could adopt a number of measures. Firstly, the market should be opened up to competition. As has been shown with the failed 1996 regulations, placing restrictions on entry and thus limiting the number of pharmacies is detrimental, as it increases monopoly power of the incumbents and excludes competition. The granting of public service state contracts to new pharmacies should be open to all qualified pharmacists. If there is a sufficient population to grant a new contract then surely this is sufficient evidence of viability and the adverse effects on existing pharmacies should not be considered. Otherwise, inefficient pharmacies will be protected from direct competition, and the deadweight loss to consumers will be greater.

Secondly, limits on ownership should be introduced to ensure pharmacy services are kept localised. This measure may not necessarily be economically efficient, however wider social goals regarding access to pharmacy services take priority. The CAI suggests limiting ownership to two or three pharmacies. At present a review group is drafting a new regulatory system that is considering capping ownership at 10% of pharmacy contracts in each health board area (The Pharmaceutical Journal, 2002).

Thirdly, subsidisation could be introduced in order to meet the government's wider social goals for pharmacy services. The long-term strategy of the Department of Health and Children wants pharmacists to deliver a quality service as community health care professionals. This service is, in essence, a public good. Central to this idea is equal service to all; therefore, a subsidised system of financial support for pharmacists operating in remote and poorer areas would ensure the preservation of pharmacy services in these areas.

Fourthly, the supply of pharmacists should be increased. The Bacon report (1999) has analysed the labour market for pharmacists and concluded that the profession has been experiencing excess demand against a background of restricted supply. Recent moves by UCC and RCSI to introduce pharmacy degree programmes,

along with TCD's existing programmes, have more than doubled supply to 170 graduates per year (The Irish Times, 2001). Restricted supply also results in a misallocation of talent as entry requirements are high, which is costly from a societal point of view. Fingleton (1994) see this as a resource allocation problem, as pharmacists are involved in rent seeking because their talents are being under-utilised.

None of this is an argument for a free market in the pharmacy sector. This is not possible for a number of reasons and indeed many of the regulations are beneficial. For example, the system whereby the Department of Health and Children negotiates wholesale prices with the manufacturers appears to be a very effective means of counteracting the extreme market power that the drug companies would otherwise enjoy in negotiations with individual pharmacists.

Effects of Deregulation

The Office of Fair Trading (1997) in the UK admits that it is difficult to estimate precisely the potential benefits to consumers that would derive from deregulation, in the form of increased price and quality competition. However, they do point out that prices may drop, and cite the example of some national supermarket pharmacy chains that offer substantial price savings on OTC medicines of up to 30%.

While direct price benefits would not accrue to customers under the GMS scheme, private customers of pharmacy-only medicine would surely see price competition reduce prices, which would reduce excessive profitability among many pharmacies, and would result in longer-term improvements in the efficiency of pharmacies. This could further fuel cost savings, which would benefit pharmacies and could also be shared with consumers.

Conclusion

Returning to the introduction, The Competition Authority (2002) has shown that margins for a typical pharmacy outlet in Ireland are 45% on private prescriptions and 33% on all medicines. This is just under 1.5 times the European average. In this essay, I have attempted to analyse the reasons for these large profits in terms of regulation, competition policy, and entry barriers.

Typically, entry restrictions to any market result in prices being higher, innovation lower and quality of service poorer. The Irish retail pharmacy market is heavily regulated. As mentioned, I advocate the principle that competitive markets to which there are no barriers to entry generally serve best the interests of consumers. Two main arguments are offered in support of regulation, the 'quality of care' and

the 'protecting strategic interests' arguments. However, I have shown that both these arguments fail to benefit consumers and are only advanced to protect the interests of incumbents by restricting competition. Reforms could be introduced to help speed up deregulation: opening up the market to competition, limiting ownership, subsidising less profitable pharmacies to meet wider social goals and increasing the supply of pharmacists.

Regulation is not always bad, but the onus lies with the government to present conclusive arguments in favour of restrictive legislation, such as the 1996 regulations. Otherwise, legislation risks favouring particular groups, to the overall detriment of society.

Bibliography

Bacon, P. & Associates (1999), *Assessing Supply in Relation to Prospective Demand for Pharmacists in Ireland,* Available from: <http://www.hea.ie/publications/ Pharm%20Report.pdf>

Carlton, D.W. & J.M. Perloff (1994), *Modern Industrial Organization*. 2nd ed. Glenview: Scott, Foresman & Co.

Fingleton, J. (1994), "Competition Policy and Employment", *Economic and Social Review*, Vol. 25(1).

The Office of the Attorney General Health. (1996) *Community Pharmacy Contractor Agreement Regulations*. Government of Ireland. S.I. No. 152/1996.

Hoban, F. (1999), "Economics of Pharmacy: Analysis of Industry Structure", *IPU Review*, Vol. 24(1).

MacArthur, D. (1995), *The Growing Influence of the Pharmacist in Europe: opportunities in a changing market*. London: Financial Times Publishing.

Office of Fair Trading (1997), *Competition in Retailing*, Research Paper 13, September.

House of Lords. Select Committee, *Pharmacy Act (Ireland)* London 1875, HMSO.

The Competition Authority (2002), *Report of Investigation into Gehe Ireland Ltd.*, Available from: <http://www.tca.ie>

The Irish Times (2001), "UCC cleared to open new pharmacy school", *The Irish Times* 23/10/01. Available from: <http://www.ireland.com/newspaper/ireland/2001/1023/archive.01102300015.html>

The Pharmaceutical Journal (2002), "Irish post-deregulation review to curb monopolies" *The Pharmaceutical Journal*, Vol. 269(7225).

The Pharmaceutical Journal (2000), "Irish Pharmacy Restrictions Challenged" *The Pharmaceutical Journal*, Vol. 265(7120).

Tirole, J. (1988), *Theory of Industrial Organisation*. Cambridge: MIT Press

A CASE FOR RE-INTRODUCING THIRD-LEVEL FEES? AN ECONOMETRIC ANALYSIS.

BY TARA MCINDOE

Junior Sophister

While possible re-introduction of third-level fees in Ireland has been a central topic for policy debate and a great worry for the Irish student body for two years now, very little research has been done to date concerning the effect of the free fees initiative. Did free fees really have a positive impact on participation levels of the lowest socio-economic groups? Tara McIndoe examines the equity-enhancing effect of government expenditure on third level education and of the free fees initiative in particular at a most pertinent time. Using an econometric model of her own design to conduct the analysis, she concludes that although in general government expenditure on tertiary institutions contributes towards equitable participation, the free fees initiative has not been a success we might have thought it was!

Introduction

Education is an arguably important part of the reason that Ireland was able to experience the Celtic Tiger years that occurred between the late 1980s and 2001.

The Irish government has, for decades, prioritised education as a corner stone of the state's contribution to national development and as a means of achieving equity among its people. As such the 1960s saw the introduction of free secondary education; in the 1970s and 1980s education policy was focussed on raising participation levels of all socio-economic groups at primary, secondary and tertiary education levels; and the 1990s saw a strong movement to counter educational disadvantage in Irish places of learning (McCoy and Smyth 2004).

As Ireland enters the 21st century, however, education remains a topic of hot debate among policy makers, students and those who provide this integral service. Although there are many contentious issues surrounding education in Ireland, I would like to focus on that of the free fees initiative and government expenditure at tertiary level institutions.

Empirical Questions

I propose to carry out an econometric regression analysis on government expenditure at third level and the free fees initiative introduced in 1996 examining exactly how these have affected the participation levels of the lowest socio-economic groups of Irish society in third-level publicly funded institutions.

I anticipate this to be a linear regression relationship over the given time period. I expect to find that although government expenditure is positively related to these participation levels, this expenditure, which represents a significant proportion of public spending, needs to be more wisely allocated in order to have the maximum equity enhancing effect. Secondly, I anticipate that the effect of the free fees initiative on increasing the level of the lowest socio-economic groups' participation at third-level has been negligible or negative. I expect this due to current discussion in the public arena indicating that policy makers view the initiative as having failed to achieve this crucial objective.

Literature Review

Ireland is a small, egalitarian society that has commissioned a vast amount of literature over the years documenting the effects of government-funded schemes aimed at enhancing the education system within its borders. The most recent and pertinent examples of these come from the Economic and Social Research Institute and the Department for Education and Science.

The ESRI paper by Selina McCoy and Emer Smyth specifically investigates the implications for equality from educational expenditure from primary to tertiary level. Their detailed analysis concludes that:

> "In the Irish context, initial educational qualifications are highly predictive of employment chances, quality of employment and pay levels … The benefits of education accrue not only to individuals but to the broader society with increased educational investment associated with a reduction in welfare costs and crime levels." (McCoy and Smyth 2004).

Although they did not carry out a specific econometric analysis they found that despite increased government focus on eliminating education disadvantage via increased government expenditure and the introduction of the free fees initiative in 1996; this has had no noticeable effect on the equality of participation at third-level education.

The second report, for the Minister of Education and Science, was based heavily on Patrick Clancy's work in the area of education equality. This mirrored the ESRI paper, maintaining that:

> "For tertiary education to be equitable, early childhood, primary and secondary education, too, must be equitable – otherwise the pool of students is not available." (Department of Education and Science, 2003).

Significantly this work was commissioned under a rationale centred on: "the lack of success of the free fees initiative" (Ibid). It took into consideration action group feedback stating that: "it is not equitable or efficient for students from more advantaged social backgrounds to be in receipt of financial aid from the State to attend higher education" (Ibid).

None of the authors carried out an econometric analysis of their data, a task I will undertake in an attempt to verify, or otherwise, their conclusions.

Econometric Analysis

Data

Dependent Variable (SES[1])

The dependent variable in this analysis results from the simple sum[2] of the participation of Semi-skilled and Unskilled SES groups in third-level education. SES is expressed as the proportion of school-leavers leaving school at the Leaving Certificate level and entering third-level institutions immediately.

Data was sourced from the ESRI School-Leavers Surveys from various years between 1979 and 1998, complied and collated by ESRI statisticians. Unfortunately the latest School-Leavers Survey has not as yet been published. The latest data available for this variable then refers to school-leavers entering the tertiary-level in the year 1997/1998. Key information on whether the free fees initiative has had a visible, permanent effect on SES participation will only become available when this latest data is published.

The data quality, sourced from the ESRI through an effective *Annual School Leavers Survey* is assumed to be adequate. Correction however for both lagged population growth[3] and inclusion of a broader range of third-level entrants

---

[1] SES means 'Socio-economic Status' but is also used as an abbreviation of socio-economic.
[2] It must be noted that data for 1995 were not available and were constructed as a simple average of immediately preceding and following figures. This was unfortunate as 1995 represents the first year of the free fees initiative when fee levels were decreased by 50%. (Department for Education and Science, 2003)
[3] Impossible to obtain exact data as population censuses are only carried out once every 5 years.

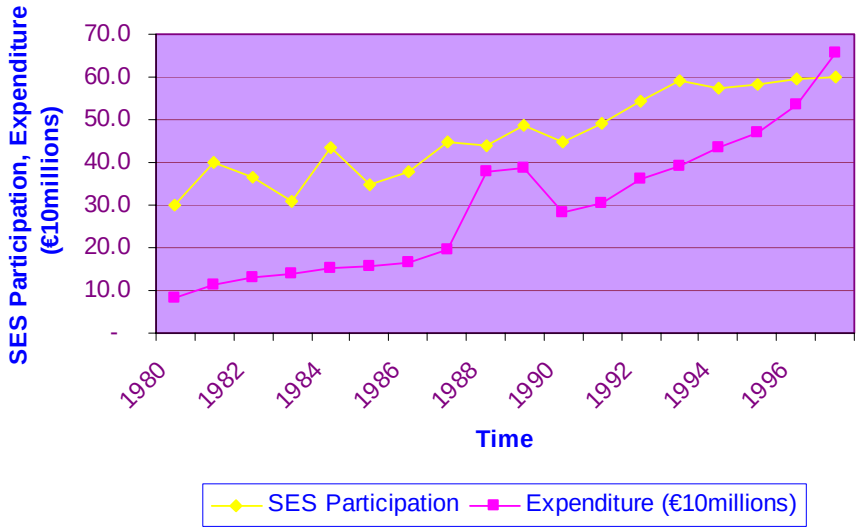
(for example second-chance[4] and mature students) would have been desirable, had the data been available.

First Independent Variable (E)

The first X variable measures government expenditure on third-level education. These observations are made up of the simple sum of total current and total capital expenditure at the third-level, deflated for alterations in the Consumer Price Index.

The quality of the data, measured at source (from the department of Education and Science) is again assumed to be adequate for the purposes of this analysis, that is, as a representative sample. Although a heavier weighting for government expenditure specifically aimed at increasing SES participation at tertiary institutions could conceivably be more instructive.

Figure 1: SES, E Against Time



This simple plot of SES and E against time shows that there appears to be a positive relationship between the two variables over time.

[4] i.e. students who have attended but not completed approved courses and who return after a break of at least 5 years to pursue approved courses (Ibid).

Second Independent Variable (F)

The final X variable is a dummy variable indicating the presence or otherwise of compulsory fees for all third-level students. As previously indicated, the Irish government introduced this initiative in 1995. 0 indicates the years that fees had to be paid in (including 1995 when the fee level was decreased by 50%), 1 indicates the years in which no fees were required by the state.

It is unfortunate that the dependent variable observations curtail the use of this second independent variable. Analysts eagerly await the release of the latest School-Leavers Survey in order to understand more fully the impact of the free fees initiative on SES participation at the third-level.

Excluded Variables

Variables such as indicators on government programs designed specifically to decrease educational disadvantage at third-level institutions (including the presence or otherwise of the Trinity Access Program) have been omitted from this analysis, it is recognised that their inclusion in future investigations may be highly instructive.

Regression

SES on E and F

Table 1: Summary Regression Result. SES on E, F
Dependent Variable is SES
18 Observations used for estimation from 1980 to 1997

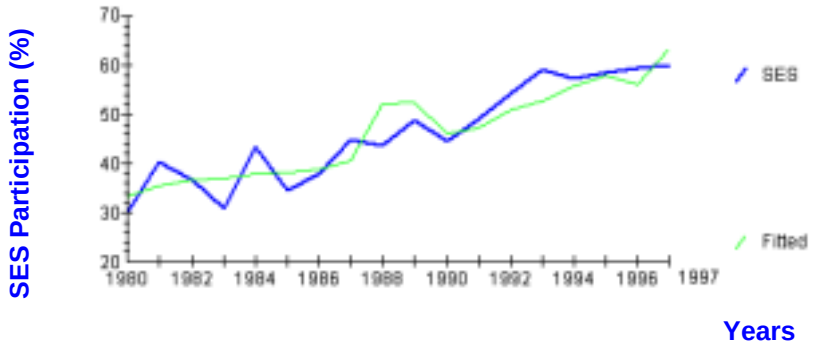
Regressor	Coefficient	Standard Error	t-Ratio	Probability
C	28.4469	2.5351	11.2212	.000
E	0.062608	0.0088009	7.1139	.000
F	-6.0181	4.4956	-1.3387	.201
R-Squared	0.82511			
F-Statistic	(2, 15)		35.3835	.000

The Coefficient for determination is an indicator of the goodness of fit of the model. R-Squared for this multiple regression is 0.825 indicating that the fitted regression line explains more than 82% of the variation of E and F in terms of SES.

The t-statistics of the multiple regression analysis individually test the hypotheses that each variable's coefficient is equal to zero. For the SES and E variables, these probabilities are of negligible value indicating that the null hypotheses in both cases can be rejected. For the F variable the null hypothesis can only be rejected at a significance level of more than 20%. Thus the regression analysis suggests the coefficient of the dummy variable is statistically not different from zero. Explicitly the high probability value of the dummy variable t-statistic indicates that the presence or absence of fees at tertiary-level institutions in Ireland has little or no effect on the third-level SES participation of the lowest income groups in Irish society. It is important to note also that the coefficient displayed for the F is negative. This implies that if the absence of fees had an effect on lower SES participation at tertiary-level institutions that this effect would be negative. This displays intuitive sense as equal government subsidy of both high and low SES groups at third-level institutions is an inefficient use of funds and will subtract from the resources available to fight disadvantage on other important fronts.

The F-Statistic displays the overall significance of an observed multiple regression that is, it simultaneously tests the hypothesis that the coefficient of all the variables are zero. In this case the probability of the obtaining an F-statistic of more than 35.38 is negligible. This is a very positive result for the model and will allow a significant degree of confidence in forecasting and inference, despite the poor t-statistic relating to the dummy variable as discussed. It is interesting to note also that in the case of simple regression of SES on E barring the test for functional form (which is significantly different from the corresponding multiple regression result) all other diagnostic and inference statistics confirm the validity of this model.

Figure 2: SES Against Fitted SES



Indeed the figure above does indicate to the naked eye a good relationship between the fitted and actual SES values. Further inference will only be available on construction of relevant confidence intervals.

Table 2: Diagnostic Regression Tests

Test Statistics (LM Version)		Probability	
Durbin Watson		2.2199	
Serial Correlation	(1)	0.44714	0.504
Functional form	(1)	0.83155	0.362
Normality	(2)	0.83171	0.66
Heteroscedasticity	(1)	0.045448	0.831

The diagnostic results above are generally good, this is crucial in order to preserve the critical assumptions underlying the Classical Linear Regression Model (CLRM) on which the Ordinary Least Squares estimation method is based. It must be noted, however, except for the Durbin-Watson test statistic, that these tests rely on large sample sizes, a feature that the current regression lacks. This will invalidate their high probabilities to some extent a situation that may have to be addressed.

Firstly, the Durbin-Watson statistic, a test for the presence of autocorrelation, is close to 2 (2.22 above) thus indicating that there is little or no evidence of first order auto-correlation. This is a better test for correlation than the serial correlation test which relies on large sample sizes.

Secondly, the result on the Ramsey test[5] is not significant at over 36%, however, this test relies on large sample sizes. In order to attempt a suitable analysis of the sufficiency of the linear functional form in this small sample regression four other good indicators of functional form will be examined. Namely, the preceding R-squared result, t-ratios, Durbin-Watson result and the signs on the estimated coefficients however all indicate that functional form is adequate despite the small sample size analysed.

Thirdly we examine the Chi-squared test for normality. Although the high probability value of 66% indicates that the null hypothesis[6] cannot be rejected; it is noted that again a larger sample size is required to accept this result with any confidence. Rather then a histogram plot of residuals should be examined:

[5] Null Hypothesis states that a linear functional form is correct.
[6] Null Hypothesis states that the OLS residuals display no skewness and normal kurtosis.

Figure 3: Histogram of Residuals and the Normal Density

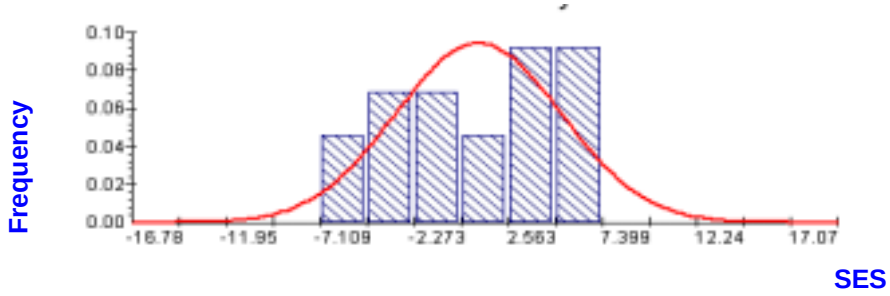


Figure 3 indicates that the residuals of the model do not conform to a normal distribution. This is a significant weakness in the data. Various corrective measures including increased sample size may alleviate this problem. In view of this both the t and F statistics should be treated with caution.

Multicollinearity is not a significant issue in this analysis given that the second explanatory variable itself a dummy variable with no obvious linear dependence on the E variable.

Homoscedasticity is the last Classical Linear Regression Model (CLRM) assumption to be explicitly tested. The OLS regression has indicated an 83.1% probability value for the heteroscedasticity test, this is not significant and implies that the regression model displays significant homoscedasticity. It must be noted that the small sample size employed negates to an extent the validity of this diagnostic test.

Confidence Interval

Policy makers typically use econometric results to make specific policy proposals. In order to offer the policy maker the maximum amount of information, this econometrics based report includes the following: a 95% Confidence interval on E, the coefficient for expenditure is found to be bound by the limits (0.08136, 0.04385). The two-tailed 95% interval is fairly narrow: the coefficient of interest has a value of 0.06261, the CI binds this value at 29.96% of the coefficient's value above and below the actual estimated value

Forecast

Figure 4: Model's Forecast Power for the 1990s

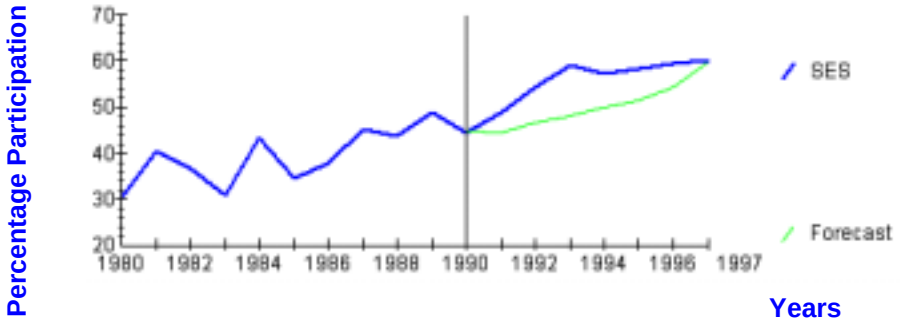


Table 3: Single Equation Static Forecasts

****

Based on OLS regression of SES on:
C E
11 observations used for estimation from 1980 to 1990

****

Observation	Actual	Prediction	Error	S.D. of Error
1991	49.0000	44.2195	4.7805	4.7380
1992	54.4000	46.6602	7.7398	5.0027
1993	59.0000	47.9766	11.0234	5.1832
1994	57.3000	49.9508	7.3492	5.4964
1995	58.4000	51.5029	6.8971	5.7738
1996	59.4000	54.4274	4.9726	6.3571
1997	60.1000	59.6296	0.47043	7.5385

****

The figure and results above show that although the prediction power of the test seems quite poor in the mid-1990s, at the beginning and end of the 1990s, its prediction power improves dramatically. This is mirrored by the low structural stability of the test for example the F-statistic probability value is low at 3%. The predictive failure test gives a high F-statistic probability value of more than 35% however, inferring that the model has weak predictive power throughout the 1990s.

The loss of predictive power during the mid-1990s may be explained by various shocks to the model. For example, various government schemes aimed at

decreasing educational disadvantage not accounted for in this model may have had significant effects on lower SES participation at third-level institutions. It is possible that these effects had stabilised towards the end of the 1990s although a detailed investigation of these would be highly instructive.

Conclusion

It has been shown then that the conclusions reached by both the Economic and Social Research Institute and the Department for Education and Science have been validated by the above econometric analysis.

Specifically, government expenditure on third-level education has a positive relationship with increased participation in third-level institutions by low socio-economic groups in Ireland. This is a long-term stable relationship and it would be advisable for the government to continue and increase its current and capital expenditure towards increasing third-level education as a proven way to decrease inequality of opportunity and outcome as regards the Irish education system.

Secondly, the econometric findings in this report have indicated that the introduction of the free-fees initiative has had a negative effect (although the coefficient is not significantly different from zero at the 5% significance level) on the SES participation of lower income groups at tertiary-level institutions. As previously stated this makes intuitive sense. Literature on educational inequality has shown that the earnings and welfare possibilities for non-graduates are depressed in relation to those opportunities for graduates, this is detrimental to Irish society as a whole.

Although this report bases its analysis on limited data the main conclusions are inescapable. In order for Ireland to maintain its competitive skills level and continue to strive for equality of educational opportunity and outcome it would be advisable not only to increase the scope and volume of government expenditure at and towards tertiary-level education but to re-evaluate the third-level free fees initiative which has not achieved one of its primary objectives, that is to increase the equity of SES participation at third-level institutions throughout Ireland. Discussions towards these ends are urgently required.

Bibliography

Central Statistics Office (2004), Available from: <http://www.cso.ie> Dublin.

Department for Education and Science (2003), *Supporting Equity in Higher Education.* Dublin: Department for Education and Science, Statistics Section.

Economic and Social Research Institute, (1980-1997), *Annual School Leavers Survey.* Dublin: Department for Education and Science- Statistics Section.

McCoy, S & Smyth, E. (2004), *Educational Expenditure: Implications for Equality.* Dublin: Economic and Social Research Institute.

THE INTERACTION BETWEEN THE PRINCIPLES OF EFFICIENCY AND EQUITY IN IRISH CONSUMPTION AND WEALTH TAXES

BY CONOR DOYLE

Junior Sophister

Theory holds that a good tax system satisfies three conditions: it is efficient, equitable and simple. However, as is the case of many economic theories, in reality there are many trade-offs involved between the three. Conor Doyle examines the interaction effects between the two principles of a good tax system in Ireland – equity and efficiency – with relation to consumption and wealth taxes.

Introduction

This paper will show that these principles interact through systems of trade-offs between them in the Irish tax system. I will begin this paper by outlining the concepts of equity and efficiency, providing a framework for their analysis. I will proceed to describe each Irish tax on consumption and wealth, and examine the interaction between the principles of equity and efficiency in the operation of each tax. It will be shown that Irish consumption and wealth taxes are designed with the conflicts between these principles in mind, and that they are doing well at establishing at achieving a workable trade off between them.

Equity and Efficiency

These are two desirable components of a tax system, which often conflict in the recommendations they put forward in designing the tax system. We will look at the theory of each.

Efficiency

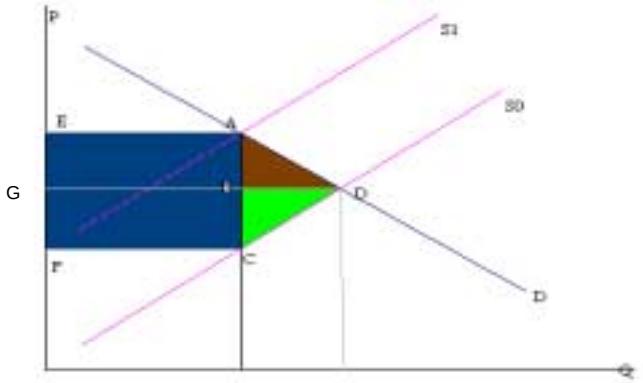
This principle is based on the theory that we would make our optimum utility maximising choices in a perfectly competitive market. An efficient tax is thus one which is "not distortionary" (Stiglitz, 2000) – i.e. it does not affect the choices, which would have been made in the absence of the tax. To design an efficient tax is

to attempt to minimise the costs choices foregone in imposing a tax. The principle of efficiency breaks down into three areas:

Economic Efficiency

A tax has been introduced into a market which charges the supplier per unit produced. This tax causes the supplier to charge more for each unit. Supply shifts from S0 to S1, causing a movement along the demand curve to D1. Graphically:

Figure 1: Deadweight Loss of a Tax



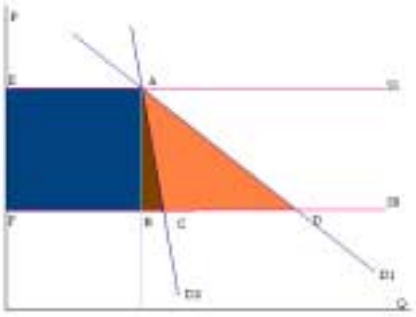
Revenue raised by the tax is measured by area EFCA, while there is a loss of utility, termed the deadweight loss (DWL) which corresponds to DCA. The higher the DWL, the greater the incentive for black market activity. The smaller the DWL, the less effect on choices made, for a given level of revenue, the more economically efficient the tax. Thus, the less the price elasticity of demand (PED)[1] of a good. Imagine demand for a second good, D2, with the same supply curves,[2] but a greater PED.

[1] The smaller the change in consumption for a change in price, the smaller the deadweight loss a tax on that good will impose.

[2] We use a horizontal supply curves to exaggerate this effect. In the same situation, non-horizontal supply curves will lead to a greater reduction in excess burden than a reduction in tax revenue, remaining relatively more efficient. See Stiglitz (2000) for mathematical derivation of this relationship.

Both collect the same revenue, but a tax on good 2 imposes a smaller deadweight loss. The difference between the deadweight losses is called the excess burden of the higher tax[3] (Stiglitz, 2000).

Figure 2: Excess burden of a tax



Administration Costs

A tax also imposes a cost of collection. The higher the ratio of revenue collected: the more administratively efficient the tax (Cullis and Jones, 1998).

Compliance Costs

Taxes further impose costs on those who have to pay it, in the form of time and resources needed to comply (Ibid). This may include lawyers, accountants costs, physical stress of compliance, time away from productive work, etc. If compliance costs are higher than those of avoiding the tax legally, resources may be moved from productive activities to tax avoidance, lowering overall welfare. If compliance costs are higher than the costs of illegal evasion, there is an incentive for black market activity.

[3] The size of the overall tax burden is determined by the government through a range of considerations, not all of which are economic. The overall burden of tax is thus irrelevant to this paper, only the relative excess burden.

Equity

The definition of an equitable tax is the subject of many arguments (Kelly, 2000; Cullins and Jones, 1998). It seems reasonable that individuals in the same circumstances should pay the same amount (Horizontal Equity) and thus that individuals in different circumstances should pay different amounts (Vertical Equity). But even these seemingly reasonable assertions beg the question; which differences are relevant? How should people in different situations be treated? What should be taxed? (Stiglitz, 2000)

This paper is not intended to answer these questions and will not calculate 'optimum taxes,' which are themselves subject to value judgements. I intend only to look at the value judgements on equity evident in the Irish consumption and wealth taxes. The most important elements of a tax system from an equity viewpoint are:

Progressiveness

A tax system is said to be progressive, if it takes proportionally more tax from those with a higher ability to pay (i.e vertically equitable). We can take a simple measure of the progressiveness of a tax using:

$$(T_1/I_1) - (T_0/I_0) / (I_1 - I_2), \quad T_1 > T_0$$

Where T_1 and T_0 are the tax liabilities at income levels I_1 and I_0 (Rosen, 1992).

Incidence

The incidence of a tax refers to who actually pays it. Referring to Figure 1, the incidence is split between the amount the supplier pays (ABGE) and the amount the consumer pays (CBGF). This occurs despite the fact that the tax was levied on the supplier. Who pays how much of the tax depends on the relative PEDs of the demand and supply curves. According to vertical equity, the incidence of a tax should fall on those more able to pay i.e. luxury goods/goods with high PEDs should be more heavily taxed.

Equity vs. Efficiency – Trade Offs

It should be clear that the design of any tax system faces conflict between these objectives. As regards consumption tax, efficiency perspectives recommends we tax necessities, equity says we should tax luxury goods. As regards wealth tax, efficiency says we should tax all wealth to minimise incentives for tax evasion, while equity says we should tax the wealthier more heavily.

The Irish Tax System

We now examine the specific operation of each tax, and the interaction between the principles of equity and efficiency in their operation.

We should first note the relative importance of these taxes to the Irish Exchequer. Irish tax receipts are more dependent than most of our EU partners on returns from consumption. Nonetheless, when we come to our conclusions, it should be borne in mind that income tax is still the most important source of tax revenue.

I would also note that administration costs are negligible. The Revenue Commissioners received a budget of €262 million in 2002, or 0.01% of government expenditure from tax receipts (Department of Finance, 2003)

Table 1: Tax Revenue Structures in the EU and USA

	Property Taxes	Consumption Taxes	Social Security Contributions	Income and Profits	Other
USA	10.1	13.8	23.3	49.1	3.7
Austria	1.3	26.8	34.2	28.7	9
Belgium	3.3	23.3	30.9	38.6	3.9
Denmark	3.3	30.9	4.6	58.9	2.3
Germany	2.3	27.2	39	29.8	1.7
Ireland	5.6	35.6	13.6	42.2	3
Italy	4.3	25.8	28.5	34	7.4
Holland	5.4	26.2	38.9	25.3	4.2
Sweden	3.4	20.1	28.1	41.6	6.8
UK	11.9	30.8	16.4	39.2	1.7

Source: OECD (2002)

Consumption Taxation

There is one major and two minor consumption taxes in Ireland:

1. Value-Added Tax (VAT)

"A general sales tax applied at all stages of production and distribution to the supply of taxable goods or services" (Revenue Commission, 2004).

While it may be economically efficient to tax consumption of commodities with lower PEDs, by their nature, such goods are often necessities. Taxing them will lead to proportionately higher taxes on poorer people, who spend proportionately more of their income on necessities. VAT follows a strongly progressive system,

which insures tax incidence is on those buying luxury goods and minimises regressive distribution impacts. Necessities are mainly zero-rated for VAT, including food, educational services, children's clothing, and medical services & supplies. Eating out and most services are charged at 13.5%. Luxury goods are penalised at 21%, including all consumer durables, business supplies, and most manufactured goods (Ibid.) Incidence falls on those buying higher more "luxury" goods.

Clearly this system is economically inefficient, as it imposes an excess burden on society as a whole. The Government recognises the inefficiency of VAT and explains the logic of the interaction between the two principles under VAT thus:

> "Valid arguments have been made by the NESC in the past that maintaining zero rates of VAT is economically inefficient… Notwithstanding… it appears that it would be difficult to gain public acceptance for any removal of the existing zero rates… It would not be possible to compensate all those potentially affected in an exact manner. There has been pressure from time to time to introduce a second lower rate possibly as low as 5 per cent. This option has been rejected, in the past, mainly due to the cost to the Exchequer… the existing single reduced rate works well in terms of maintaining a balance between collecting revenue and reducing the development of black economy activity." (Tax Strategy Group, 2001b)

The government recognise that the tax is considered roughly equitable, and do not wish to upset this perception. Therefore, efficiency is maximised within the current equity framework. Compliance costs for consumption taxes are clearly negligible:

Table 2: Payment compliance

	1999			2000		
Payment Compliance All Cases	Due Month	Due Month + 1	Due Month + 2	Due Month	Due Month + 1	Due Month + 2
VAT	73%	91%	96%	77%	93%	97%

Source: Tax Strategy Group, 2001a.

The use of a higher tax rate across many sectors leads to administrative inefficiency. However, this is apparently negligible, and the impact of many small market distortions tends to be less than the impact of one large distortion (Stiglitz, 1999)

2. Specific Consumption Taxes

A number of specific consumption taxes have been levied, the most important being on alcohol, tobacco and petrol. These taxes manage to combine equity and efficiency. The goods taxed have high PEDs - but are also seen as social vices rather than necessities. If people use less of them, the inequitable impact of these taxes can be relieved with a socially acceptable outcome (Revenue Commission, 2004).

3. Withholding Tax on Professional Fees (WT)

"A withholding tax, at the rate of 20 per cent, is deductible at source from payments for 'professional services' made to individuals and companies by 'accountable persons' (Government Departments, local authorities, health boards, State bodies, etc.) (Ibid).

WT, a small part of 'other taxes' has little impact on the economy. It imposes little economic inefficiency and has few other costs as it is 'deductible at source'. It remains generally progressive and equitable, in that it major services (medical, education, social) are exempt, being covered by PAYE and PRSI.

Wealth Taxes

"Strictly speaking, the term wealth tax can be used to cover a tax on the transfer of wealth (such as a gift tax, estate duty, or capital acquisition tax), a tax on appreciation of wealth (a capital gains tax), as well as a tax on the stock of wealth" (Sandford & Morrissey, 1985).

Ireland has two Wealth Taxes[4] :

1. Capital Gains Tax (CGT)

"A capital gains tax is chargeable on the gains arising on the disposals of assets other than that part of a gain which arose in the period prior to the 6th of April 1974. Any form of property (other than Irish currency) including an interest in property (as, for example, a lease) is an asset for Capital Gains Tax" (Revenue Commission, 2004).

Thus, property taxes correspond to CGT. However, CGT encompasses a raft of exemptions clearly designed to remove the incidence from those with a low stock of wealth. The first €1,270 of personal capital gains in a year are excluded, as are sale of one's home, if it is your only house and sale of a life assurance policy not

[4] Deposit Interest Retention Tax is a tax on income accruing from a stock of wealth and is therefore not included

bought from another individual. There are further exemptions for small firms (one owner), depreciation, sporting bodies, charities and trade unions (Ibid).

Gains accruing to other/larger businesses, [5] and wealth accruing to individuals beyond the levels above, are taxed at 20%. Thus the tax is proportional and avoids the economic inefficiencies

2. Capital Acquisitions Tax (CAT)

"Capital Acquisitions Tax comprises Gift Tax, Inheritance Tax, Discretionary Trust Tax and Probate Tax".

Gift Tax (GT) and Inheritance Tax (IT)

GT and IT operate such that inheritances within the direct family are not taxed heavily, while other gifts and inheritances are heavily taxed. Accordingly, there is a tax free threshold of €441,198 for gifts or bequests to a son/daughter and a special exemption for nephews/nieces inheriting from childless aunts/uncles. All other inheritances and gifts are taxed at levels above of €44,120. From an equity standpoint, incidence is removed from individuals and families.

Given recent rises in property prices, this tax does impose a considerably larger deadweight loss on such 'other' groups of inheritors. However, given the small amount collected through this tax, this deadweight loss is relatively small from the standpoint of society as a whole. Irregularity of collection minimises compliance and administration costs.

However, while the tax is proportional the fact that the tax is levied on those with higher stocks of wealth makes the cost of tax evasion much lower in relation to costs of paying. There may be black market activity generated. As Sandford and Morrissey (1985) show, the direct wealth tax, which acted as an annual gift and inheritance tax, applied to so few that its returns hardly merited its costs.

Probate Tax (PT) and Discretionary Trust Tax (DTT)

> "Probate Tax is charged at the rate of 2% on the estates of persons dying…A once-off Inheritance Tax applies to property subject to a discretionary trust…The current rate of tax is 6%. An annual Inheritance Tax at the rate of 1% applies to property subject to a discretionary trust" (Revenue Comission, 2004).

[5] In the interests of administrative and compliance efficiency, many capital gains accruing to business other than sale of land, are taxed under Corporation tax, at the same rate as CGT (Ibid)

PT exemptions include estates below €50,790 in 2000, property passing absolutely to a surviving spouse, the principal private residence, property passing to a charity, heritage property.

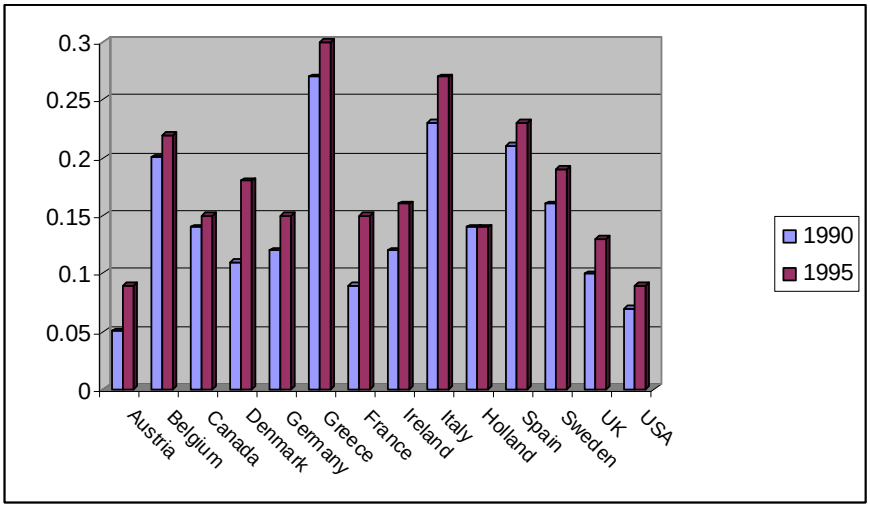
The low level of tax, irregular payment and large exemptions allow equity and efficiency to interact in the same way as for IT and GT. There are no exemptions on DTT.

Outcomes of the Interaction between Efficiency and Equity in Irish Consumption and Wealth Taxes

The Irish tax system seems to have gotten the balance between equity and efficiency about right.

As regards equity, Ireland has managed to keep an average score on the GINI index of income inequality relative to other industrial nations.

Figure 3: Gini coefficients in the EU and USA[6]



Source: World Bank, 2004.

[6] No full set of GINI coefficients is available for more recent years. In fact, the World Bank has collected only two relevant figures – Belgium, 0.29-1996, and USA, 0.41-1997 .

Table 3: GINI coefficients in the EU and USA

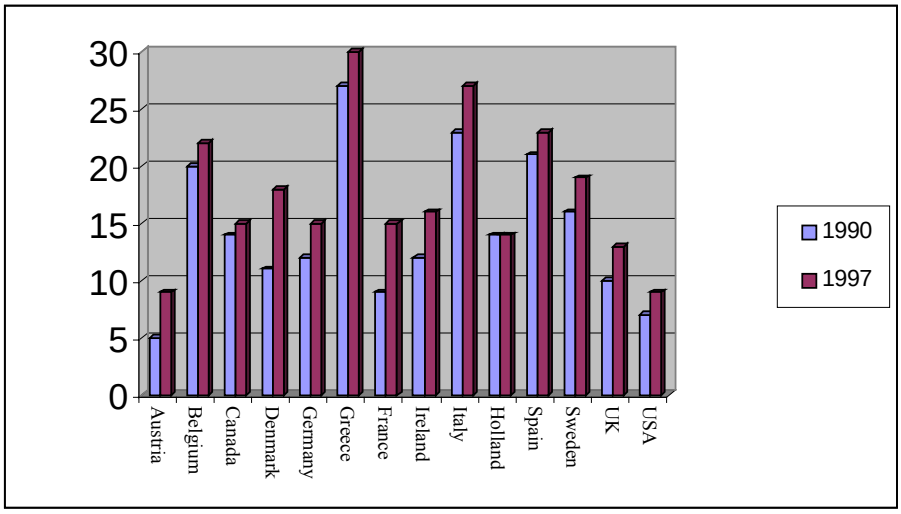
	Austria	Belgium	Canada	Denmark	Germany	Greece	France
1990	0.05	0.2	0.14	0.11	0.12	0.27	0.09
1995	0.09	0.22	0.15	0.18	0.15	0.3	0.15

	Ireland	Italy	Holland	Spain	Sweden	UK	USA
1990	0.12	0.23	0.14	0.21	0.16	0.1	0.07
1995	0.16	0.27	0.14	0.23	0.19	0.13	0.09

Source: O'Hagan (2000).

However, efficiency also remains relatively good. Despite market distortions, the level of black market activity is average.

Figure 4: Level of black market activity in the economy



Source: McAleese, (2001)

Clearly, this is not solely due to wealth and consumption taxes. As noted above, income tax is the largest tax revenue source in Ireland. Other factors such as government corruption, tradition, social values, etc. also have a (non-quantifiable)

part to play. However, consumption and wealth taxes constitute close to half of all taxes in Ireland, and are thus significant factors in determining the overall interplay between the principles of equity and efficiency in Ireland.

Conclusions

It is clear that the taxes discussed have been set up with the trade offs between efficiency and equity in mind. They are consciously designed to maximise both efficiency and equity objectives. While this paper can only provide a discussion of half of Irish tax revenues, as the introduction stated, the Irish tax system is doing rather well at creating a system, in which equity and efficiency interact to achieve both a relatively equitable distribution and a relatively efficient tax regime.

Bibliography

Cullis J. and Jones J. (1998), *Public Finance and Public choice.* 2nd. ed. Bath: Oxford University Press

Department of Finance (2003), "Budget 2003" [Online]. *Department of Finance*. Available from:< www.finance.gov.ie/budget/budget2003/wptabs.htm.>

Kelly, J. (2000), "Theories of Distribution: Rawls vs. Nozick." *Student Economic Review* 2000. Dublin: Trinity College.

McAleese, D. (2001), *Economics for Business.* 2nd. ed. Dorset: Prentice Hall.

O'Hagan, J.W. (2000), *The Economy of Ireland*. 8th. ed. Dublin: Gill & Macmillan

OECD (2002), *Revenue Statistics 1965-2000*. Paris: OECD.

Revenue Commission (2004), "Tax Duty and Type" [Online]. *Revenue Commission*. Available from: <www.revenue.ie>

Rosen, H. (1992), *Public Finance 3rd. Ed.* USA: Irwin.

Sandford C. and Morrissey P. (1985), "The Irish Wealth Tax – A Case Study in Economics and Politics." *ESRI*, no. 123

Stiglitz J. (2000), *Economics of the Public Sector.* 3rd.ed. New York: Norton.

Tax Strategy Group (2001a), "Tax Compliance (01/10)" [Online]. *Department of Finance.* Available from: <www.finance.gov.ie/publications/tsg01/0110.doc>

Tax Strategy Group (2001b), "Vat Issues (01/08)" [Online]. *Department of Finance*. Available from: <www.finance.gov.ie/publications/tsg01/0108.doc>

World Bank (2004). "Development Data". [Online]. *World Bank*. Available from: <www.devdata.worldbank.org>

AN ECONOMIC PERSPECTIVE ON DRUG PROHIBITION

BY IVAN HARKINS

Senior Sophister

This original essay by Ivan Harkins addresses an issue that many would find controversial – that of legalising certain illicit substances. The author suggests approaching the problem from the economic perspective. He examines the history of past experiences of drug decriminalisation and rigorously sets out the background for a proposed future cost-benefit analysis and further policy development.

"Prohibition goes beyond the bounds of reason in that it attempts to control a man's appetite by legislation and makes crimes out of things that are not crimes. A prohibition law strikes a blow at the very principles upon which our government was founded"

Abraham Lincoln. (The Quote Cache, 2003).

Introduction

The debate on the legal status of drugs has a quite extensive and often inconsistent history. While all sides acknowledge the destructive societal effects of drug use, the sources of these effects are disputed. The purpose of this essay is to discuss the pros and cons of legalising certain illicit substances.

Within the literature, there is a lack of any well-structured argument defending prohibition. One such paper stemmed from a meeting of the AER chaired by Irving Fisher (1927). Basing their argument around the increased savings and productivity of workers perceived during prohibition of alcohol, they declared the evils of alcohol to be detrimental to society. Their discussion was neither empirically comprehensive nor theoretically solid. As one speaker mentioned the proposed increased in savings was more attributable to changes in the banking system than to any change in the drinking patterns of consumers.

This essay will begin by outlining a brief history of prohibition in the US where there is a more widespread literature available. From there it will examine the theory surrounding prohibition and drugs. Then the discussion proceeds to talk

about the two 'natural experiments' of note in the prohibition debate: the prohibition of alcohol in the 1920s and decriminalisation of marijuana during the 1970s.

History

In the late 19th and early 20th century drug use was widespread. Opiates, cocaine and marijuana were part of a recreational life as well as part of many patent medicines and herbal remedies. Indeed in 1619, Virginia assembly passed the first law concerning cannabis, requiring every household to grow it; the fibres were legally used for sails and rigging. Marijuana use is believed to date back many centuries. The Chinese dynasties reportedly used what they called "ma", as part of a mixture for anaesthetic (Schlosser, 2003).

Public opinion on drugs, in the US, changed early in the twentieth century. With a rising tide of immigrants from China and Mexico, a strong racial sentiment altered attitudes toward drug use: marijuana became associated with the Mexican, as did opiate use with the Chinese. As movements began to restrict the influx of migrants, so also began a new stance against drugs.

The American Pharmaceutical Association declared: "if the 'Chinaman' cannot get along without his 'dope,' we can get along without him." In 1909, the United States' international war on drugs began when California prohibited the importation of smokeable opium (Drug Policy Alliance, 2003).

Similarly, bigotry and suspicions of Mexicans and later black jazz musicians were exploited to ban the use of marijuana. It was "depicted as an alien intrusion into American life, capable of transforming healthy teenagers into sex-crazed maniacs" (Schlosser, 2003). In 1914, a law was passed in the State of Texas banning the possession of marijuana and by 1931 twenty-nine states had followed suit. In 1937, the US Congress endorsed the Marijuana Tax Act, which criminalised the possession of cannabis throughout the United States.[1]

By the 1960s, the political culture was changing. There was a growing movement of marijuana users. Among college students and 'hippies' the 'weed' was gaining popularity. As a result of this shifting culture:

[1] "The first victim of this war was neither a foreign drug lord nor an organised crime boss, but rather a 58 year old farmer named Samuel R. Caldwell. [He] was apprehended three days after the law was passed selling a couple of joints to his buddy Moses Baca. These guys lived in Denver, Colorado and probably did not have C-Span [an American television station] to let them know of the newest laws on Capitol Hill. Mr. Baca received 18 months in Leavenworth Penitentiary, a maximum-security unit infamous for its history of violent offenders. Mr. Caldwell was fined $1000 and also sent to Leavenworth where he served out a four-year sentence. He died less than a year after his release" (Samuel Caldwell's Revenge, 2003).

> "in 1972 the National Commission on Marijuana and Drug Abuse recommended that the country 'decriminalise' the use and possession of a small amount of marijuana, following that recommendation, 11 states repealed the criminal prosecution of marijuana users, replacing it with citations and small fines" (Lee, 1993).

Cannabis was re-criminalised again in 1984 under Ronald Reagan after political pressure from parents associations and other political influences. It has remained criminalised, with increasing penalties.

In the modern day, drugs are playing a growing role within the media. In film and music industries drug use is being romanticised. There is an increasing need in today's context to 'de-mythologise' and separate the real economic and societal effects of drugs from those caused by prohibition.

Theory

The need for regulation in a market, according to economic theory, is driven by imperfections. There are two primary imperfections that exist in the drugs market. The first of these is informational asymmetries, for example, general lack of knowledge about the longer-term effects of drugs (perhaps explaining observed 'myopia') and the unknown quality of an individual suppliers quality of product. A second imperfection arises from externalities. This includes negative externalities such as, increased health costs or escalated violence. Is 'blanket-prohibition' of drugs, the best policy-solution to these market imperfections?

The discussion will look at each of the factors within the debate on 'prohibition of drugs'. The choice of whether drugs should be legal or illegal is ideal for study in an economic context. There will be a review of the literature concerning the 'natural experiments' mentioned at the outset, and these arguments will be developed below.

Level of Consumption

The first factor is the possible change in consumption by new and existing users. Some argue, intuitively, that a legal market will result in a higher demand because the ascribed price[2] of drugs will have gone down. Others refer to such 'natural experiments' as the alcohol prohibition where demand did not respond dramatically in the long run, to either prohibition or the eventual abolition of the law. Both viewpoints are valid and cannot be completely refuted. However it is unlikely given the evidence of alcohol prohibition, that there will be the 'flood' of drug users

[2] The price to the consumer includes the cost of legal punishment, which will be removed under legalisation, as such the price perceived by consumers will be lower.

that some commentators suggest, especially in the long run when "tasting trials"[3] will smooth out.

Poverty

There are those who argue of the possible effects of producing and exacerbating poverty levels. This stems from the belief that alcohol and cigarette consumption are often strongly correlated with poverty. However, causality is not well defined here. We do not know if poverty induces drug use or whether drug use causes poverty.

Kaestner (1998) looks at the potential causal effects that illegal drug use has on poverty. He finds that after controlling for other characteristic variables such as age, family background, etc., there is some causal effect: "a large preponderance of the estimates indicated that marijuana and cocaine use significantly increase the probability of being poor". However, Kaestner qualifies his findings with warnings about "several empirical limitations that make this a less than definite analysis"(Ibid.).

On the other hand, it is feasible that the casual effect, should it exist, will not be as strong as one might think *ex ante*, as will be discussed later. Illegal drugs such as cannabis may be strong substitutes for alcohol and cigarettes, and any effect that they (illicit drugs) cause will be, as a result, diluted.

Purity of Produce

A strong argument for legalising drugs is the improved purity content of the drug sold as a result of legalisation, and eventual regulation. It is widely contended that the majority of deaths from drug use are due to impurities in the product sold. Presumably, a regulated legal market would result in fewer deaths among users. This is the heart of the argument put forward by a Mr. Fulton Gillespie (2001) to the House of Commons in England. After the tragic death of his son, a drug addict, Mr. Gillespie discovered that his son died as a result of the toxic content of the heroin he consumed. He expresses in his memorandum to the commons, a view that is reflected by number of people: "drugs are a public health, not a criminal matter"(Ibid.) He cites three reasons for this:

a) "They should be taken out of the monopoly clutches of criminals;
b) The billions saved in law enforcement, street and property crime etc. should be channelled into control, licensing, education, prevention and treatment

[3]"Taste trials" refers to users who may sample drugs out of curiosity, but who may not have done so under prohibition. This effect will smooth out over the long run, to a stable level, i.e. one where as each generation comes of age they will 'taste trial'.

c) That thousands more like my son will die from criminally-supplied impure drugs unless western governments recognise that the present war against drugs is un-winnable and counter-productive." (Ibid.)

Crime Associated with Users

Another argument for legalising drugs is that the crime caused by users in order to finance their habit would diminish. Crime of this nature 'feeding a drug habit' currently accounts for a third of all crimes in the UK: "Feeding a drug habit can cost £20,000 a year – which many addicts fund largely through theft." (10 Downing Street, 1999).

Crime is an obvious externality of drug use, which is not only motivated by the addictiveness of the drug, the intensity of consumption, but also the price. It can be argued that a legal market could facilitate the optimal price level in order to minimise these negative externalities.

Clague (1973) sketches rough outlines of some possible legal strategies for dealing with heroin addictions.[4] Clague, interestingly, concludes that a heroin maintenance scheme where heroin is distributed freely and safely at depots scores 'best' on a number of points including crime reduction and minimising disease (a factor, which will be discussed later).

Crime Associated with Dealers

One of the central arguments legalisation advocates is the proposed reduction in crime associated with gaining market control. This crime, widely reported in the press, is depicted as mostly extremely violent and is considered by many analysts to be escalating problem. As Miron and Zwiebel explain, "prohibition is likely to lower marginal costs and raise marginal benefits to violence in an industry". They provide several reasons for this fact, including factors such as: "participants cannot use legal means to solve problems, participants are already breaking the law and concealing another crime is marginally cheaper." (Miron and Zwiebel, 1995).

This point is also illustrated in a United Nations Educational, Scientific, and Cultural Organization (UNESCO) report:

> "All over the world, the existence of laws that prohibit…compel practices and organisational forms that are both underground and violent in their means of negotiation (threats, intimidation, blackmail, extortion) or in conflict resolution (aggressions, murders, terrorism) to react accordingly to either their commercial or private disputes" (Schiray et al, 2002).

[4] While Clague avoids a discussion of legalising heroin, his analysis insightful for a discussion of legalising the drug.

Disease

Another implication is the reduction in the spread of disease, especially that of AIDS (associated with sharing of needles in the use of heroin, or through association with other industries such as prostitution). This is a widely recognised problem and has motivated many policy changes across Europe. According to a UNESCO report: "many drug addicts become ill from drug taking and may therefore become the 'high-risk group' for transmitting HIV and AIDS." (Ibid).

Terrorism

A very topical point within this debate is that allegedly, the proceeds from illegal drug trade are often used for the funding of terrorist organisations. It is alleged that in countries like Afghanistan, the drugs trade funds terrorist organisations, namely in the case of Afghanistan, the Al-Qaeda. If the market for drugs was brought under government control then selectivity in suppliers could be emphasised.

Opportunity Costs of Enforcement

A classic economic perspective involves the opportunity costs of law enforcement and the opportunity cost of maintaining prisoners. The money spent on the 'war on drugs' and maintaining the growing prison population could be spent on educating potential users to the dangers of drugs. Some figures may help elucidate the problem. In the UK during the year 2000, c. 111,000 people were dealt with for principal drug offences, c. 47,000 were cautioned or fined and c. 64,000 went through the courts of which c. 55,000 were found guilty. Some might argue these people are inherently dangerous. However when one looks at the figures more closely 74,000 ca. of those people were dealt with for possession of cannabis, a relatively harmless drug (Corkery, 2002). Whether one believes that this is a waste of resources or not is subjective, but the foundations for the argument are there. One should also keep in mind the theory that prisons introduce people to criminal networks and as such create criminals rather than rehabilitate them. In the UK: "60% of all offenders go on to commit more crimes after their sentence has ended" (10 Downing Street, 1999). Apart from the reduced crime, there would also be the benefit of additional tax revenue from the legalised drugs market. For instance, the drugs market in the UK for the year 1998 was estimated to be c. £6,613.5 million, a mere 1% of that brings in £66million (Bramley-Harker, 2001).

Miron (2003) from Boston University has carried out research assessing the budgetary implications of legalising marijuana. He estimates that it could save the government of "Massachusetts $120.6 million per year in government expenditure on criminal justice enforcement of prohibition." His report also indicates that legalisation would yield additional tax revenue of $16.9 million annually.

Complementary Illicit Markets

There are also possible knock-on effects on other illicit markets such as prostitution. It is commonly publicised that a significant portion of women involved in prostitution are drug users. A recent report by the BBC showed that the majority of prostitutes in Bristol are 'feeding' their drug habit. While this is not a conclusive survey it is a clear indicator of a definite causation. BBC Points West reporter Scott Ellis said: "What is different in the city now is that women are turning to prostitution because they are addicted to drugs" (BBC News, 2003). Perhaps by decriminalising drugs women with addictions will not be forced into these situations, either through 'pimp-pushers' imposing extortionate prices or through being stigmatised by society.

Utility of Consumption; Rational Addictions

There is a further argument, a corner stone to much economic theory; that of the utility of the consumption of drugs. Utility derived from drug consumption should be accounted for in assessing a cost-benefit analysis of legalising the market. The choice to consume drugs is, under the assumptions of rationality, like any other good. It is therefore said that the consumer will maximise their utility such that the marginal utility divided by price (i.e. MU/P) is constant across all goods. By making a product illegal, the government is minimising the consumption possibilities of consumers, and distorting the shape of their utility maps.

The Demand for Substitute Goods

There is also the sometimes forgotten effect, on the consumption of substitute goods such as alcohol and cigarettes. Both the aforementioned goods are similar in nature to 'harder' drugs, with characteristic traits such as addictiveness. Both cigarettes and alcohol carry negative externalities, the most pronounced being, especially for cigarettes, increased health care risks. A well-founded theory in economics is that the consumption of a good is negatively correlated with its substitutes. It is possible therefore that the perceived negative externalities associated with increased consumption of drugs will be counteracted by the reduction in consumption of alcohol and cigarettes.

Forgotten Medicinal Benefits

Many commentators advocate the medicinal benefits of drugs. Marijuana, it is asserted, has useful medicinal benefits that are not utilised due to its illegality and then the stigma attached to that.

Schlosser (2003), taking reference from a book entitled [*Marijuana], the Forbidden Medicine* by Dr. Lester Grinspoon, an emeritus professor of psychiatry at Harvard Medical School, outlines some of the proposed benefits. These include

relief from nausea associated with chemotherapy, prevention of blindness induced by glaucoma, an appetite stimulant for AIDS patients, an anti-epileptic, help warding off asthma attacks and migraine headaches, alleviation of chronic pain, and reduction of the muscle spasticity that accompanies multiple sclerosis, cerebral palsy and paraplegia.

With a legal market, it is possible that the proposed benefits of marijuana can be explored further, both by government and informed consumers. But as Dr. Grinspoon says: "You cannot patent this plant" (Ibid), so it is unlikely it will receive the research it requires without government assistances.

The Stepping Stone Theory

One of the few arguments against legalising drugs is that perhaps lighter drugs act as stepping-stones to harder drugs. However, there is evidence, from research carried out that this effect is overstated. Indeed, in the Netherlands where marijuana use has been decriminalised, there is evidence that the places of purchase for the various drugs differ substantially - cafes are the key outlets for marijuana purchases, yet cocaine and other drugs are rarely purchased in these outlets (Cohen et al., 1999). Similarly, a study in the UK finds that after controlling for the role of unobservable factors (such as a social or psychological predisposition toward antisocial behaviour) the spurious association between 'soft' and 'hard' drugs dissipates (Pudney, 2002). The researchers came to a conclusion that the proposed 'gateway' effect is minimal and as such, economically insignificant.

Topic for further study: Cost-Benefit Analysis?

The discussion above provides a possible starting point from which to initiate a more quantitative analysis of the effect of legalising drugs. Distinctions need to be made between physically addictive and non-addictive drugs, which were largely ignored in this essay.

Also I would note that as with any cost benefit analysis there will be different weights attached to each factor. From a generalised viewpoint I believe those factors (i.e. externalities) that affect non-users should be weighted heavier than those that affect users, the rationale being that the user has made a 'choice' where as the non-user has not, i.e. the user should have 'priced' the costs into his decision to consume.

Natural Experiments

In all branches of economics, there is a need for empirical backing of theory. Within the drugs debate instances of such possibilities for research are limited. The following are applied cases of legalisation of drugs in action, and provide us with real world settings from which to study the effects of prohibition.

Decriminalisation in 1970s USA.

Lee (1993) demonstrates with reference to data from the era of 'decriminalisation' of marijuana in 1970s USA that harassing drug users is an ineffective means of reducing consumption. Lee outlines an explanation for this assertion by investigating the dynamics of an illicit market, which are quite distinct to that of the traditional supply and demand framework (Ibid.).

Traditionally, it has been viewed that an illicit market works similarly to that of a legal market, where the threat of penalties raises the cost for the seller and the price for the buyer. As such a stricter enforcement policy is more effective in combating drug use. However, Lee contends that this is not necessarily the case. Lee's argument is based on the concepts of the factors that determine the choices made by sellers and buyers. Firstly, he posits that users of drugs hold a stock of drugs that they maintain. He then identifies two penalties that motivate the purchase decisions made: possession penalty and transaction penalty. These factors work in opposite directions: the possession penalty encourages the buyer to make many small purchases, while the transactions penalty pushes the buyer to make few large purchases. Similar motivations are determined for the supplier.

Lee explains the reduced consumption in the eleven states that 'decriminalised' marijuana as being driven by the changed dynamics of supply and demand (inexplicable in a traditional framework, where demand would increase due to decreased penalties and as such total consumption should rise).

Under the 'decriminalised' market, the use and possession of a small amount of marijuana no longer resulted in criminal prosecution, small fines. Citations still occurred, and states continued to criminally prosecute supply. This divergence in the treatment of suppliers and users of drugs meant that demanders of drugs were "less fearful of open market transactions involving a small amount…of marijuana". However, suppliers still faced large penalties as before. Therefore, while demand increased, supply of small quantities was less responsive, and possibly even decreased: "when the criminal justice system punishes dealers much more severely, the supply could decrease much more than the demand would increase, so that the net effect would be less consumption." (Ibid). Lee also suggests another possibility, that of an unusual demand behaviour that he dubs "forbidden fruit syndrome". Yet, he declares that no noteworthy adjustment in thinking or feelings were evident, on the part of demanders, from research at that time (Ibid.).

Alcohol Prohibition 1930's USA

Miron and Zweibel (1991) studied alcohol consumption during prohibition in 1920s USA. As they comment themselves: "although the parallels between the criminalisation of alcohol and the criminalisation of drugs are not exact, prohibition

provides a natural setting in which to examine the impact of legal restrictions on the use of substances such as alcohol or drugs."

After recognising the difficulties of gauging the size of an illegal market, they looked for proxies for the consumption of alcohol and then managed to create a statistically significant explanatory model.

Their findings are that in the initial years of prohibition there was a dramatic reduction in the amount of alcohol consumed, but over time consumption levels began to increase and by the abolition of the 'dry laws' consumption was already up to 60-70% of pre-prohibition levels, and rising.

The reasoning behind this is that while demand never changed, supply took time to accommodate itself to the new mechanisms of an illegal market. This makes intuitive sense, as networks of supply are difficult to establish, especially under the auspices of an illegal market. Also of note is that post-prohibition total consumption did not rise dramatically. Clearly any consumption control strategy had serious problems.

What about some of the other factors mentioned earlier? It is difficult given the era to find data relating to this information. However, Miron (2001) has written papers, such as *Violence, Guns and Drugs: A Cross-Country Analysis,* which outlines some of these effects.

He illustrates that guns[5] are not an indicator of violence.[6] He demonstrates, however, that expenditure on enforcement of drug prohibition is a significant explanatory variable for violence. He explains this phenomenon as arising from, as discussed earlier, the lower marginal costs of violence for drug dealers. This translates into a greater perceived demand for guns by the public that in turn contributes further to a violent society.

Conclusion

A number of aspects within the debate on the legalisation of drugs have been outlined. It is important to acknowledge some of the limitations associated with this essay. Of note are the important distinctions between addictive and non-addictive substances that were not drawn out in this analysis. There is a lack of extensive literature on 'natural experiments', due to the nature of the markets being studied. As Lee (1993) illustrates, inferences drawn from illegal markets about legal markets are not applicable. Nevertheless, there would appear to be sufficient economic grounds to infer that a legal, regulated market for drugs is superior to the

[5] When America is removed from samples.
[6] Measured as number of homicides

current illegal market. The discussion is a strong starting point for studies such as quantitative cost-benefit analyses that might verify the current information available, and allow a more precise examination of this misunderstood industry.

Acknowledgement:

The author would like to thank fellow students Brendan Graham and Eoin Denvir for their valuable commentary and feedback.

Bibliography:

10 Downing Street (1999), *Crime and Justice: Annual Report 98/99* [Online]. 10 Dowling Street. Available from: <http://www.number-10.gov.uk/output/page216.asp#>

BBC News. (2003), "Drugs Lead to Prostitution Boom." [Online]. *BBC News.* Available from: <http://news.bbc.co.uk/go/pr/fr/-/1/hi/england/bristol/2931983.stm>

Bramley-Harker, E. (2001), "Sizing the UK market for Illicit Drugs" [Online]. *Research and Development Statistics Directorate (RDS).* Occasional Paper No 74. Available from: <http://www.homeoffice.gov.uk/rds/pdfs/occ74-drugs.pdf>

Clague, C. (1973), "Legal Strategies for Dealing with Heroin Addiction" *American Economic Review*, Vol. 63 (2).

Cohen, Peter, D. A. et al. (1999), "Licit and illicit drug use in the Netherlands, 1997" [Online]. *Centrum voor Drugsonderzoek, Universiteit van Amsterdam*. Available from: <http://www.cedro-uva.org/lib/abraham.npo97.pdf>

Corkery, J., M. (2002), *Drug Seizure and Offender Statistics, United Kingdom, 2000,* Home office: London: 2002.

Drug Policy Alliance (2003), "The History of Prohibition" [Online]. *Drugs Policy Alliance*. Available from: <http://www.drugpolicy.org/race/historyofpro/>

Fisher, I. (1927), "The Economics of Prohibition." *American Economic Review,* Vol. 17 (1).

Gillespie, F. (2001), "Memorandum Submitted by Mr. Fulton Gillespie" [Online]. *Appendices to the Minutes of Evidence*. Available from:

<http://www.parliament.the-stationery-office.co.uk/pa/cm200102/cmselect/cmhaff/318/318ap25.htm>

Kaestner R. (1998), "Does Drug Use Cause Poverty?" [Online]. *NBER Working Paper* 6406 Available from: <http://www.nber.org/papers/w6406>

Lee, L.W. (1993), "Would Harassing Drug Users Work?" *The Journal of Political Economy*, Vol. 101 (5).

Miron, J. A. and Zwiebel, J. (1995), "The Economic Case Against Drug Prohibition," *The Journal of Economic Perspectives*, Vol. 9 (4).

Miron, J., A. (2001), "Violence, Guns and Drugs: A Cross-Country Analysis" [Online]. *Journal of Law and Economics*, XLIV (2) Available from: <http://econ.bu.edu/miron/images/gd_isp.pdf>

Miron, J., A. (2003), "The Budgetary Implications of Marijuana Legalization in Massachusetts" [Online]. *Boston University*. Available from*:* <http://econ.bu.edu/miron/images/legalize_mj_ma.pdf>

Miron, J., A. and Zwiebel, J. (1991), "Alcohol Consumption During Prohibition" *The American Economic Review*, Vol. 81 (2).

Pudney, S. (2002), "The road to ruin? Sequences of initiation into drug use and offending by young people in Britain" [Online]. *London Home Office Research: Development and Statistics Directorate*. Available from: <http://www.homeoffice.gov.uk/rds/pdfs2/hors253.pdf>

Samuel Caldwell Revenge (2003), *The First Causality of the Drug War* [Online]. Available from: <http://www.samuelcaldwellsrevenge.com/>

Schiray et al. (2002), *Drug Trafficking, Criminal Organisations and Money Laundering*, United Nations Educational Scientific and Cultural Organisation, and United Nations Office for Drug Control and Crime Prevention: New York.

Schlosser, E. (2003), *Reefer Madness and other tales from the American underground*. London: Allen Lane The Penguin Press.

The Quote Cache (2003). Available from: <http://quotes.prolix.nu/Drugs/Alchohol/>

DUBLIN'S REAL ESTATE BOOM – A RATIONAL BUBBLE?

BY DAVID COMERFORD

Senior Sophister

David Comerford examines the Dublin housing market that has been experiencing significant price increases over the last 10 years. The question he tries to answer is whether this price bubble is rational in nature. He reaches a positive conclusion by examining the market structure and trends in market fundamentals.

The New Palgrave (1987) defines a bubble "as a sharp rise in the price of an asset or a range of assets in a continuous process, with the initial rise generating expectations of further rises and attracting new buyers". This definition seems entirely consistent with Dublin's real increase of 195% in house prices between 1995 and 2002 (TSB/ESRI, 2003) and the continued strong speculative demand thereafter. The question I seek to answer in this project is whether the "expectations of further rises" that are currently fuelling price increases are rational.

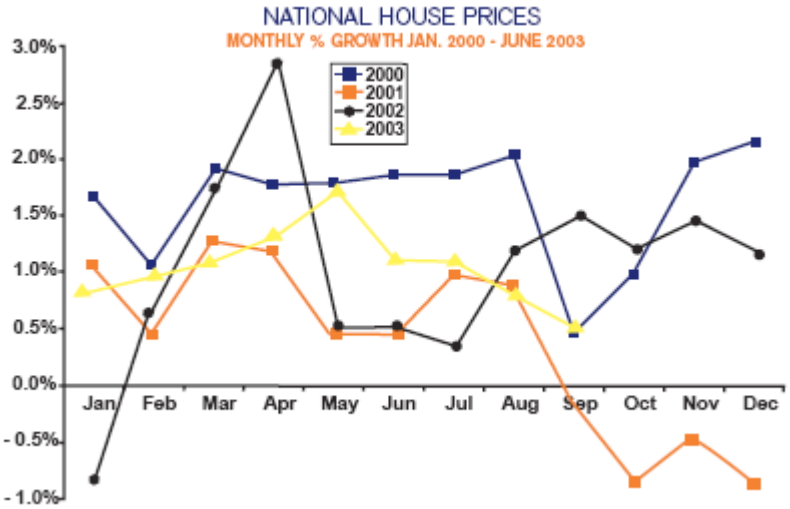
The ubiquitous example of speculative irrationality is Dutch Tulipmania of the mid-17th Century. The New Palgrave (1987) formalises the association by rendering irrationality and Tulipmania synonymous. It defines Tulipmania as "situations in which some prices behave in a way that appears not to be fully explainable by market fundamentals". Indeed its definition eschews any etymological specifics and jumps straight into the general case in which "the price of land rises essentially because it is expected to rise". Again this is redolent of Dublin's property boom.

The core issue then is whether the magnitude of the expected rise in land prices is fully explicable by a rational interpretation of market fundamentals. I shall answer this question by identifying and then analysing the changes in 'market fundamentals'. Garber (2000) defines these as the variables that "are believed to drive asset prices". The answer I arrive at and the thesis of this project is that the property market exhibits symptoms of bubble behaviour not because individuals are behaving irrationally but because it is impossible to quantify many market fundamentals.

Introduction - Dublin's House Price Experience:

There has been an explosion in the value of house prices in Ireland over the course of the past decade. A survey of property, published last June in the Economist, reveals that the nominal value of the average house has increased by 219% since 1995. Though this report was pessimistic about the sustainability of Irish house price inflation, the latest figures from the ESRI (The Economist, 2003) show that prices have continued to rise and have done so at an increasingly rapid rate. The latest TSB/ESRI (2003) national house price index for the third quarter of 2003 shows a rise of 9.4% over the course of the year thus far. This compares with a figure of 9% for 2002. Moreover, a month on month increase has been observed. In September house prices increased by 0.5%, whereas October witnessed a rise of 1.4%.

Figure 1: National house prices monthly growth



Source: TSB / ESRI House Price Index (2003)

As can been seen in the above diagram, house price changes in 2003 have been consistently positive and relatively stable. Using this as a guideline, there is little evidence of confidence erosion in the Irish housing market.

At first glance, such surety in the market belies the bubble thesis propounded in the Economist survey (2003). In *Another Bubble Fit To Burst*, the influential publication calculated that Irish property was overvalued to the tune of

42%. The Economist forecasts a fall in nominal prices of 20% over the course of the next four years.

Inflation in house prices comes despite a steady and unremitting increase in the supply of housing. A record number of houses have been completed in each of the past eight years (UNECE, 2003). Last year there were 57,625 completions and annual increases have been in the region of 9% since 1997. This means that over the past decade approximately 380,000 new housing units have been constructed. Although my crude statistical analysis takes no account of housing stock depletion (for which figures are unattainable), it seems inconceivable that the construction of a new dwelling for one out of every twelve members of the population should do so little to satiate demand. In order to see whether this is consistent with market fundamentals I shall examine the peculiarities of demand and supply in the housing market.

The Nature of Housing:

The first thing to note about the housing market is that it is not a conventional market. Houses tend to be indivisible, relatively illiquid and involve high transactions costs. In calculating these transaction costs, we must add the taxes, legal and administration costs (which the Economist (2003) calculates as being approximately 6% of purchase cost for Ireland), the costs of market research prior to purchase and the high psychic costs of upset caused by relocation. Though the costs of relocation only affect owner-occupiers and may actually confer utility in some cases, researching the market is a considerable cost to all actors. This is aggravated because, as markets go, housing is plagued by information failures.

Another remarkable feature of housing is its status as a necessity. It is not necessary to own a house, however historic and cultural evidence suggests that the Irish find it desirable to do so. Over 80% of Irish householders are homeowners, the highest level of owner occupation in the EU (Drudy, 1999). Public policy has been charged with promoting this goal. For example, mortgage interest relief, the first time buyers grant, the abolition of capital gains tax on the sale of the principle property and the remission of residential rates all serve to incentivise home ownership.

Crude theory would suggest that a house, as a necessity would be relatively income inelastic. According to this logic we would not place great weighting on income as a market fundamental driving house prices. However, this logic neglects the fact that a house has various functions. Along with its perfunctory role of providing shelter, a house also confers luxurious benefits. A home is a stake in a community. Rational people typically wish to live in as nice a community as possible, one with good facilities, low crime, aesthetic appeal etc.

Unlike many other goods a home is invariably perceived as a reflection of its owner. As such it is a source of pride. Conspicuous consumers, lets call them the Jones family, who derive utility from having a nicer house than the Smiths, will rationally invest in home improvements. If the Smiths are also conspicuous consumers they will want to keep up with the Joneses. This pattern means that houses are heterogeneous and that many of them are constantly improving – hence house price indices fail to compare like with like. There is a broad and convincing literature to support the claim that consumers derive utility from relative status (Veblen (1899), Frank (1985), Scitovsky (1976)). Though it is irrational for an atomistic individual to consume a good in greater quantities in the wake of price increase, a socially interdependent individual will rationally exploit culturally projected utility. Though social status is empirically a factor in determining house prices (note the frequency with which terms such as "prestigious" are employed in property advertising copy), it is frequently ignored when examining market fundamentals.

Because of social culture and the policy climate, renting is empirically a stopgap measure. Only 5% of the Irish population lived in privately rented accommodation in 2001, according to a survey conducted by European Community Household Panel (2004). Moreover, the vast majority of these people tended to stay in such accommodation for under three years. This culture inhibited the development of the privately rented sector in Ireland when compared with the rest of Europe. The rental market is a purgatory for those who cannot afford to "climb to the castles in the air". As we shall see, I will show that the number of householders falling into this category is set to increase considerably over the next decade. Policies that sought to promote the privately rented sector have been supply oriented (e.g. Section 23 and Section 27 income tax allowances for construction of rental accommodation). In a market with inelastic demand that demographic trends are constantly augmenting, the economic consequence of this policy is not to reduce the rents paid by tenants, but rather to increase the rents enjoyed by landlords. These factors conspire to make house purchase the most attractive option, not least for investors in the rental sector.

These investors form the third category of actors in the housing market. Whereas owner-occupiers invariably, and renters frequently, have an emotional stake in their actions in the property market, the stake held by investors is purely financial. Keogh and D'Árcy (1993) formalise the exceptional characteristics of housing from an institutional economics perspective. Those purchasing housing as a dwelling will find housing stock heterogeneous, rife with externalities and public goods. This fundamentally alters their market behaviour. Hence, while a rational retiree might prioritise the proximity to established social networks, a speculator has

no such concerns. These motivational differences have repercussions for the efficiency of the market. Investors have far more flexibility.

Within the category of investors a qualitative distinction is often made. Some investors purchase property on the firm foundation of price-earnings ratios, looking forward to a reliable income stream. Others purchase on the basis of the 'greater fool theory', buying in the expectation of further rises in property prices and hoping to sell before everyone else does. Market sentiment is a very real market fundamental. Behaving rationally, each investor has to predict the likely behaviour of all other investors. In order to predict the behaviour of other investors, each investor needs to predict their predictions.

A further peculiarity of property facilitates speculation. Because property transactions are so capital intensive, it is usually necessary to borrow in order to enter the market. This gives rise to several distortionary consequences.

The Consequences of Credit Dependence

Firstly, reliance on credit effectively makes the property market an options market and hence subject to intense speculation. For the would-be investor a prosperous property market is very appealing. By way of illustration, I cannot think of a better example than one cited in the Economist:

> "Suppose you had invested $20,000 in shares, which after five years are now worth $40,000, including reinvested dividends, implying an annual return of 15%. Then suppose you had used the $20,000 as a deposit on a $100,000 house that over five years had risen in value by a more modest 7% a year, to $140,000. Assume, for simplicity, that mortgage-interest payments and maintenance costs exactly offset the rental income. The average annual return on your deposit would have been almost 25%" (The Economist, 2003).

Secondly, it makes each transaction on the property market more complex and hence more rigid. In order to transact in the property market one must first transact in the credit market. This gives rise to time lags.

Thirdly, the assumption of rationality asserts very plausibly that lenders are also motivated by profit. They have incentives to lend as much as possible when property prices are rising as the wealth effects of property price increases the collateral that banks hold on existing loans. Moreover, given the fact that both borrower and lender agree that property is an attractive destination for funds, financial institutions give even greater concessions to property speculators. In a bid to overcome the rigidities of indivisibility and illiquidity that deter investment, they have established pooled property investment portfolios in which shares can be bought.

Lastly, and perhaps most importantly for those concerned with rational consumption patterns, it means that the relevant price to the consumer is not the

price of a house. Rather it is the price of paying for a house. As we shall see in Dublin's case, this makes mortgage repayments a fundamental source of change in house prices.

What Are The Market Fundamentals?

The very fact that the property market is predominantly made up by a core of inflexible actors yet encourages activity at the margins by flexible speculative investors makes it a prime candidate for bubble behaviour. Moreover, the nature of housing makes it hard for investors to find information, and when it is found it frequently comes from sources with a vested interest in overstatement, for example, Irish house price indices are undertaken by estate agents and mortgage providers. We have already seen that the motivation for positive bias is a rational economic characteristic of these actors. The scope for bias in index construction is also great. In a working paper, the ESRI methodologists Conniffe and Duffy (1999) point out that "the compilation of an adequate measure of house price changes is not at all a trivial task". It should be noted that the ESRI's own index is produced in conjunction with Irish Permanent, a mortgage provider who have an incentive to overstate the attractiveness of property investment.

Such indices, as are currently used, tend to use hedonic pricing techniques, whereby a similar property is taken as a proxy measure. As we have seen properties are heterogeneous and so it is impossible to compare like with like. Other approaches discussed include the use of resale values for the same house but the Irish case reveals a paucity of data that limits its applicability.

Abstract measures of property valuation are divided on the relevant market fundamentals. Garber (2000) tells us that market fundamentals are "a collection of variables that we believe should drive asset prices". He does not tell us who 'we' are, and that makes a large difference in what 'we' consider relevant to our willingness to pay.

For a hardheaded investor, price-earnings ratios based on discounted rents are the standard means of determining future income streams. For the home purchaser high notional returns based on imputed rents are of little practical use. For example, the Bacon Report (1998) shows that the value of single-bedroom apartments in Dublin's city centre rose by 30% in 1998, whereas those for a four-bedroom semi-detached house rose by a relatively slight 18% in the same period. Though the one bedroom apartment was the better investment, the rational purchaser may have trouble explaining that rationality to the two children who are forced to sleep on the floor of their parents' bedroom.

For such a purchaser, a more relevant market fundamental is income. Economists formalise this relationship and use it as a predictive basis for price behaviour through the house-price-to-income ratio. The logic underlying this ratio is

that people tacitly decide to spend a certain proportion of the entire income that they envisage earning over the course of their lifetime on housing. The corollary of this logic for bubble watchers is that if house prices go out of line with this proportion, people are behaving irrationally. Average household disposable income is the proxy most commonly used to determine average permanent income. However, as a mean measure of income it is subject to a skewed distribution. This has bearings on the calculation. The Economist (2003) tells us that Ian Morris of HSBC believes median income, which is not subject to the vicissitudes of outliers, is a preferable measure. The difference between the verdicts reached by these two measures is striking. The mean measure sees American property prices as overvalued to the tune of 5%. The median measure revises that figure to 14%.

The point here is that even those who subscribe to the hard rationality criterion of New Classical economics will have to concede that information failures abound. To Keynes, it seemed investors realised this and so most of them "are largely concerned not with making superior long-term forecasts of the probable yield of an investment over its lifetime, but with foreseeing changes in the conventional basis of valuation a short time before the general public" (Malkeil, 1999). In short, investors are concerned with market fundamentals only insofar as the market believes them to be fundamental. Even if all market actors are subject to the same misinformation, its insuperable lack of precision means that actors at some point have to rely on their gut feeling. An economic truth becomes an economic truth not because it is true, but because I believe that you believe that most people believe it is at least plausible enough to be worth considering in evaluation.

So pervasive and costly are information failures in the property market that it may well be rational for investors to herd into a market. According to the soft criterion of perfect information, investors will cease to assimilate information when the costs of research outweigh the likely benefits accruing from it. In such a case, great savings can be made by free riding on the research of ostensibly better-informed investors and following them into the market.

In my discussion thus far I have stressed the personality of the economic actor – whether people who demand housing effectively do so as consumer, hardheaded investor or animal-spirited speculator. It strikes me as anecdotally plausible to consider house buyers to be a composite. This has repercussions for the house-price-to-income ratio.

Firstly, it seems consistent in a nation in which home ownership is held in such esteem for consumers to buy the 'best' house they can afford. Moreover, behavioural theory suggests that as the prestige attributed to housing increases, so too does its importance in consumption priorities. These observations would suggest that people spend a residual income on housing rather than a proportion of income. If the relative prices in a consumption bundle remain constant there is no difference

between the two. As we shall see, that has not been the case in Ireland over the last decade.

Secondly, as Ireland's disposable income has risen by an unprecedented magnitude and for an unprecedented length of time, it seems plausible that more people are in a position to invest. Hence actors whose role in the property market was formerly entirely consumption-oriented have become somewhat investment-oriented.

Thirdly, it seems entirely rational to shift investment portfolios in favour of property given its favourable returns relative to those of other assets, especially given the availability of capital for such investment. Moreover, actors who would have saved in the past have no incentive to do since the advent of EMU and its accompanying negative real interest rate regime.

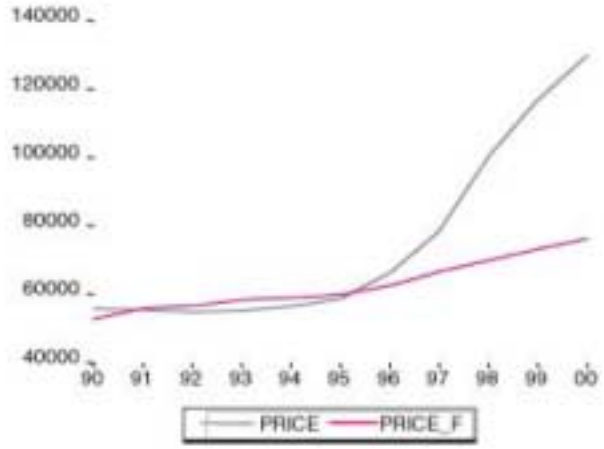
What this means is that more people will rationally spend more of their income on property, just as the house-price-to-income measure suggests. There is good reason to believe the unforeseen permanent income shift caused the ratio to shift and that a higher ratio, though perhaps not the current one, is sustainable.

The interest rate is a key variable in both abstract measures. It is therefore a market fundamental according to Garber's (2000) criterion of something believed to drive asset prices. In the next section I shall show that there has been a fundamental shift in mortgage rates in Ireland over the recent past.

Trends in Market Fundamentals

The Bacon report (2000) performed an econometric analysis of house prices. It concluded that "the structure of the housing market changed significantly after 1996".

Chart 4.2: House Prices: Forecasted (Price F) vs Actual (Price): 1996-2000



4.4.2 Forecasting House Prices

Source: Bacon, 2000

Now let us examine if this Ireland has built its house on a foundation of sand or if the fundamentals shifted with structure. The Irish economy, and with it the disposable income of Irish residents, underwent huge growth in the past decade. On average, disposable income has increased by 55% since 1992 (Lang, 2001). Increased money in an economy gives rise to increased effective demand. Ireland's openness meant that increased demand for traded goods could be catered for. The service sector, whose supply is constrained by the availability of a skilled workforce, suffered quite considerable inflation during this period but this was somewhat moderated by net immigration. Housing stock, which is fixed in supply in the short run, is consequently susceptible to inflation and indeed suffered it to greater extent than any other sector. Indeed the in-flux of service sector labour further increased demand for housing. An interdepartmental review group predicted gross immigration of 200,000 between 1998 and 2005.

We have already seen that there are high psychic costs involved in purchasing a house. Not only is it is a large financial commitment, but it is also a lifestyle commitment. Until the early nineties, Irish demography was marked by emigration. People were anxious about the long-term social desirability and economic viability of remaining in Ireland. With the turnaround in employment prospects arising out of the economic boom, Ireland, and especially the East region, became hubs of immigration. There was less uncertainty about future welfare and hence the risks inherent in purchasing a house deceased substantially. Anecdotal evidence suggests that young people who were living at home while considering emigration used their newfound affluence to buy themselves the independence and security of a home. Not only did homebuyers have more money, there were also more of them.

Given the swell of long-term positive sentiment that buoyed the economy in the mid-nineties, credit institutions were far more confident in their borrowers ability to pay. In recognition of this, the typical duration of a mortgage increased from around 25 years in 1993 to around 30 years today (Homeloans, 2004). This lengthier period means that the same weekly repayment buys a more expensive property. All buyers in the market are equally benefited by the windfall. However, it serves only to shift up the total purchase price.

Similarly, developments on a macro level have led to an increase in affordability. In the first quarter of 1993, the variable rate of mortgage repayment was 14%. By the end of that year it had fallen to 8% and it remained at this level until 1998. It has been around 4% for the past four years.

The most obvious reason for this is European Monetary Union. Since Irish interest rates are being determined by the wallowing economies of Europe's flagships, we have enjoyed a pro-cyclical monetary regime. In the short term this appears to be the best of both worlds. In the longer run it is sure to have the usual inflationary costs, such as loss of competitiveness arising out of inflation linked pay agreements. In the long run, however, it has the unimpeachable benefit of promoting a stable macroeconomic framework and moreover, one that is seen to be stable. Since macroeconomic stability reduces risk, this has lowered the costs of investment and hence has shifted market fundamentals. Davy Stockbrokers Equity Research bears this out:

> "For a variety of reasons, including the advent of low single digit inflation globally and EMU, expectations of the long run average interest rate were revised down substantially. Moreover, expectations of the amplitude of future interest rate movements are also likely to have been substantially revised downwards. As a result borrowers would have formed the belief that they could comfortably shoulder the costs of servicing much larger mortgages than in the past with much diminished risk to a sharp increase in interest rates" (Lang, 2000).

Davy's statistics (Lang, 2000) suggest that borrowers are right. In real terms, it costs the same to service a £160,000 mortgage today as it did to service a £100,000 mortgage in 1992.

To borrow a thousand pounds in 1992 cost £11. In 1999, it cost just £6. Despite the vast increase in house prices, and the far more attractive borrowing conditions, evidence suggests that mortgage borrowers are keeping their heads. In 1992 the loan-to-value ratio for all banks and building societies for the mean mortgage was 69%. In 1999 it was 67.7%.

Though the level of indebtedness relative to house values has remained stable, the fact remains that house values, and hence the absolute level of debt, may be at unsustainable levels. The loan-to-value ratio is a useful to lenders as an indicator of risk, but it is only relevant for borrowers in the event of repossession. By another level of affordability, the loan-to-income multiplier, borrowers are far more indebted. In 1994 the average borrower's mortgage was 1.27 times their income. Four years later the multiplier had risen to 1.94 (Bacon, 1998).

This rise can be attributed to the rationality of actors. In aggregate the pre-tax profit of the five major mortgage providers increased 15% in 1998. It is rational for an individual in the zero-sum game of fixed housing supply to take on as high a mortgage as possible since the house goes to the highest bidder. Access to credit confers bargaining power. Traditionally, household mortgages were limited to the sum of the second salary and two and a half times the first salary. Now borrowers are receiving up to triple the household's combined income.

It is unclear if the mortgages being shouldered by debtors are debilitating. As Lang (2000) states: "The usefulness of traditional multipliers as a guide to the ability of households to service debt is predicated on a more or less stable long run relationship between house prices and income. In circumstances where there is a change in the interest rate regime such a relationship (if it ever existed) is likely to break down". We have already seen a very fundamental change in the interest rate regime.

Moreover, increased competitiveness in the credit market has accentuated this change in the interest rate regime. Goodbody Stockbrokers' (2000) research tells us that formerly Irish mortgage lenders enjoyed the most lavish mark ups in the Eurozone, a spread of 2.5% over the wholesale rate. An upset occurred in August of 1999 when the Bank of Scotland entered the fray. The average variable rate at the time was 5.25%. By the tenth of September Bank of Ireland and AIB had cut it to match Bank of Scotland's 3.99% offer. An indication of the very real consequences of this change appeared in The Irish Examiner: "With competition at fever pitch in the mortgage market, some observers feel that many of the institutions, particularly the larger ones, will do their best to absorb the full rate rise" (Keane, 2000).

Conclusion - Are House Prices Sustainable?

The consequence of a bubble is a collapse in asset prices. We have already seen that house prices are rising month on month and show no signs of volatility. This however may not be as promising an omen as it seems. Bacon tells us "the speculative/transitory component of demand has risen in recent years". Indeed, last year 30% of house purchases were made by existing homeowners on second properties (TSB/ESRI, 2003). The conclusion that the Bacon Report (2000) drew is that this structural change in the housing market has given rise to a behavioural change whereby "previous year's house prices became more important after 1996." If this is the case, it could be said that the market is behaving irrationally since previous price changes are in no way indicative of future patterns, and hence they cannot be considered market fundamentals. However Garber's (2000) definition admits analytically irrelevant information as a market fundamental if we believe it to drive asset prices. The unhelpful conclusion that we can draw from this is that so long as actors continue to refer to previous years' prices the bubble will not burst.

I am not convinced by the inference of the Bacon report. There can be no doubt that the property market has been subject to a series of permanent shocks (EMU, demographic changes, income increases) that have shifted market fundamentals. These shifts only become relevant when they are recognised and internalised into decision-making. I would argue that this occurred in 1996. Quite apart from the quantifiable permanent shocks, I have suggested consistently that the property market has been influenced by shocks to unquantifiable market fundamentals such as consumer tastes and the social status attributed to housing.

Even if we admit Bacon's (2000) conclusion that the property market is subject to speculation that cannot be explained by a rational interpretation of market fundamentals, I do not consider it to be any less rational than any other. It merely suffers from worse information. Moreover, since individual actors are unable to change the market independently, the best they can do is to operate rationally within an irrational framework. It is my contention that this is exactly what most actors have been doing. With this in mind, I consider the Irish property boom to be a rational bubble.

Bibliography:

Bacon, P. (1998). *An Economic Assessment of Recent House Price Developments.* Dublin: The Department of the Environment, Heritage and Local Government (Housing Division).

Bacon, P. (2000), *The Housing Market in Ireland An Economic Evaluation of Trends and Prospects*. Dublin: The Department of the Environment, Heritage and Local Government (Housing Division).

Conniffe, D.and Duffy, D. (1999), Irish House Price Indices – Methodological Issues. *ESRI*. Working Paper (10)

Drudy, P.J. (1999) *Housing: A New Approach – Report of the housing commission* Dublin: Labour Party.

European Community Household Panel (ECHP) (2004). "ECHP Projects and Publications." [Online]. *Euro Panel Users Network*. Available from: <http://epunet.essex.ac.uk/echp.php>

Frank, R. (1985), *Choosing the Right Pond: Human Behaviour and the Quest for Status.* Oxford: Oxford University Press.

Garber P. (2000), *Famous First Bubbles* Cambridge: MIT.

Goodbody's Stockbrokers (2000), *Safe as Houses: revisited - they huffed and they puffe*. Dublin: Goodbody's Stockbrokers.

Homeloans, (2004). *Irish Homeloans Online*. Availible at: www.homeloans.ie

Keane, C. (2000), "Homeowners Get Temporary Reprieve from Rate Hikes." *Irish Examiner.* 10/12/2000.

Keogh, G. and D'Arcy, E. (1993), *Market Maturity and Property Market Behaviour : a European Comparison of Mature and Emergent Markets.* Reading: University of Reading, Department of Economics.

Lang, E. (2001), *Solid Foundations - Irish Mortgage Providers and the Housing Market.* Dublin: Davy Stockbrokers Equity Research.

Malkiel, B. (1999), *A Random-Walk Down Wall Street: Including a Life-Cycle Guide to Personal Investing.* New York: Norton.

Palgrave, R. (1987), *The New Palgrave: A Dictionary of Economics*. London: Macmillan.

Scitovsky, T. (1976): *The Joyless Economy*. Oxford: Oxford University Press.

The Economist (2003), "Another Bubble Fit to Burst: A survey of Property." *The Economist.* 31/05/2003

TSB/ESRI (2003), *Permanent TSB/ESRI House Price Index*. Dublin:ESRI.

United Nations Economic Commission for Europe (UNECE) (2003). "Ireland: National Market Report 2003." [Online]. *UNEC.org*. Available from: <www.unece.org/trade/timber/mis/market/market-61/ireland.pdf>

Veblen, T. (1899), *The Theory of the Leisure Class*. New York: McMillan.

AN ANALYSIS OF ROAD PRICING AND A STUDY OF ITS FEASIBILITY ON THE M50

BY COLM FAHEY

Senior Sophister

Traffic congestion has become a pressing issue in Dublin over the last few years. The ever-increasing negative externalities call for policy action. Colm Fahey evaluates the case of introducing road pricing on the M50. He uses economic theory of road pricing and illustrative cases of London and Singapore in building up his argument and, after considering congestion externalities, concludes that there is a case for the introduction of road pricing in Dublin.

Introduction

The theory of road pricing has existed for decades. This theory will be discussed and the link between economic theory and the reality of road pricing in Singapore and London will be investigated. Next the literature on road pricing in Dublin is reviewed. Returning to economic theory, the externalities of the M50 are discussed in order to establish if road pricing can internalise them. Next the feasibility of road pricing on the M50 is discussed by examining the capacity of the motorway and its connecting roads.

The Economic Theory of Road Pricing

Traffic congestion is an economic issue with an economic solution, road pricing. Traffic congestion occurs because a market for road space does not exist. Each vehicle on the road creates a negative externality. A negative externality occurs when those who are external to the market are adversely impacted upon. Each vehicle occupies road space and as a result increases the traffic congestion and delays all other motorists on the road. Road space is a scarce resource. At present it is allocated by queuing. The potential for an allocation improvement exists if the pricing mechanism is used to allocate road space. The economic characteristics of roads distinguish them from other public goods. "Road use is rival in consumption and also excludable with adequate costs of the pricing technique. For that reason

road infrastructure…[can be]…a private good with some degree of externalities" (Rolle, 1994).

The current system of road space allocation leads to a market failure as the supply and demand of road space does not equate. There is excess demand for road space. CBA does not recommend indefinitely increasing supply (Barrett, 2003). The ability to solve the problem by increasing supply is limited due to the high level of latent demand for road space. As the supply of road space increases, more of the latent demand for road space becomes actual demand (information provided by Eoghan Madden).

This reduces the effectiveness of increasing supply. The reason for the market failure is that a gap exists between the private cost to the motorist and the cost to society of the motorist's actions. Foster et al explain why this is occurs when road space is allocated by queuing.

> "Assume a road with fixed starting and end points. The impact of an additional driver can be measured by the average cost or the marginal cost. As the number of drivers increases, the amount of road space decreases and travel time will subsequently rise. In other words, the cost of *one additional driver* is an increase in travel time. This is the marginal or social cost. The marginal cost is the cost added to all drivers for one more driver on the road. Therefore, the marginal cost increases as the number of drivers increases. However, the average cost will decrease in the same period. The presence of a road price…on drivers is to increase the average cost and to deter those who value the use of the road as less, thus equating marginal and average cost" (Foster et al, 2003).

Road Pricing in Reality

Singapore

Road pricing has been a reality in Singapore for many years. Motorists are required by law to have an electronic device fitted into their cars. Each time they enter the Central Business District, their journey is recorded on the device through the use of laser. Cards can be slotted into the device, allowing drivers to deposit credit. The fee varies, from $3 to free, depending on congestion in the area 25 (BBC News, 2002). Electronic Road Pricing has worked well in Singapore. After its introduction, there was a reduction of 24,700 cars during peak times. Average traffic speed increased by 22% (Rolle, 1994). However, it has also been observed that as it approaches 7 pm cars queue, patiently waiting for the charge to disappear. This has caused gridlock of 15 minutes on approaches into the Central Business District (BBC News, 2002).

London

In 2003 a road-pricing scheme was introduced in London. A charge of £5 was placed on all cars that enter central London, with the aim of reducing congestion by 15%. The technology used was different to that in Singapore. 230 cameras are positioned at entry points to Central London. These scan the license plates. Any driver who has not paid by the end of the day then receives a fine of £80.

The level of congestion has decreased by 38%. It is expected to have raised £80 million by April 2004. The charge has been so effective, that it has reduced the number of cars entering the city by 50,000 per day (The Evening Standard, 2004). This is a remarkable achievement, considering that before the introduction of the charge, the same number of cars entered central London per hour at peak times (BBC News, 2002). Over half of the motorists who stopped driving into the city centre as a result of the charge, use public transport while approximately a quarter of them are now divert the zone (The Evening Standard, 2004). Simultaneously to the introduction of the charge, bus fares were reduced, bus capacity was increased, bus routes were improved and bus lanes were enforced (BBC News, 2002). 15,000 extra passengers are using buses in the morning rush hour. On average, 110,000 congestion-charge payments are made each day. Over 1.1 million fines have been issued. Worryingly, approximately half of motorists have won appeals against the fines (The Evening Standard, 2004).

The charge penalised cars because they are an inefficient use of road space. The charge benefited buses, which are a more efficient use of road space. Before road pricing, buses were being subsidised for sitting in traffic. The reduction in congestion has increased their productivity. Their load factor has also increased (Barrett, 2003). The charge prevented low utility journeys from being travelled. It appears that 38% of motorists gained less than £5 worth of utility from driving into London city center.[1] The charges have reduced the costs of business, as trucks are not stuck in traffic. It is estimated, that London's economy has benefited by £50 million per annum due to shorter journey times and fewer accidents (Ibid). The majority of motorists in London now agree that eliminating traffic congestion from the city centre is worth £5. Polls show that support for the charge is at 57% and that opposition is at 36% (Ibid).

Dublin?

The success of road pricing in Singapore and London makes us question its feasibility in Ireland. The main obstacle is the lack of enthusiasm amongst policy makers:

[1] In Singapore 24,700 motorists valued their peak time journeys at less than $3.

> "It is important that the Oireachtas Transport Committee relays a clear message…the London experience can never be repeated here…The cost of implementing congestion charges in Dublin would be prohibitively expensive compared to the revenue yielded…The cost would be crippling" (The Sunday Business Post, 2003).

The Oireachtas committee is wrong. It would appear that they only considered the technology used in London. The system used in Singapore would be far most cost effective for Dublin. The Oscar Faber report (1999) also disagrees with the Oireachtas committee: "The technology for such systems exists and implementation is feasible." The report lists a selection of low cost methods to implement a congestion charge. It confirms the potential of road pricing in Dublin. The report recommends that the canal ring around the city centre should be the cordon for a congestion charge. It concludes that a peak charge of €3.81 would cause a trip reduction of 8%. The revenue generated is expected to exceed €50.8 million. It is estimated that bus fares could be halved through the use of this money (Oscar Faber, 1999). This measure would increase the opportunity cost of travelling into the city by car and should result in a further reduction in congestion. Alternatively, the revenue generated could be invested in increasing the road capacity into and out of the city. Another study attempted to bridge the gap between the marginal cost to motorists and the marginal cost to society of a motorist. It recommended a congestion charge of €13.25 (Foster et al, 2003). It is my intention to analyse the feasibility of a price charge to be levied at the interchanges of the M50.

Externalities caused by the M50

Newbury identified accident costs, road damage costs, environmental costs and congestion costs as the four main externalities of motoring (Newbury, 1990). The externalities of motorways are not entirely consistent with the externalities of other road types.

Accident Costs

More traffic on the road increases the probability of traffic accidents (Foster et al, 2003). This is a serious externality in Ireland. A total of 376 people were killed in 346 fatal accidents on Irish roads in 2002. Based on fatalities and injuries sustained the cost of these accidents was €728 million. These figures give credibility to the argument that accident externalities are greater than the sum of all other motoring externalities (Newbury, 1988). However, this is not true for the M50. Motorways have a far higher safety rate. Statistics show that they have at least 65% fewer fatal accidents than ordinary roads (The Sunday Business Post, 2001). None

of the 376 fatalities in Ireland in 2002 took place on the M50. There was only one fatal accident on the M50 in 2001 and this was blamed on inadequate safety features on the motorway (The Irish Independent, 2001). The low level of accident externalities on the M50 must be accepted when it is considered that during the 4-year period previous to the 2001 fatality, over 90 million trips were made on the motorway, but there were only 2 fatal accidents. While the small number of fatal motor accidents on the M50 is tragic, they are insignificant on the overall scale. In years when there have been fatal accidents on the motorway, they account for approximately 0.2% of national fatal accidents. Road pricing would not be the most effective method of reducing accidents on the M50. The installation of median crash barriers for the entire stretch of the motorway would be a more effective method of reduced this externality. This is standard on most European motorways and it would have prevented the fatal collision in 2001 (RTÉ, 2003).

Road Damage Costs

Heavy vehicles cause almost all the damage that is done to the road surface (Newbury, 1988). It can be argued that heavy vehicles cover this externality as they pay a higher level of road tax. However, the local authorities collect road tax, whereas it is the Department of Transport that is responsible for the maintenance of the M50. This justifies the use of road pricing to internalise the road damage costs externality. The level of the charge could be set so that heavy vehicles pay the majority of the cost and cars only pay a fraction of the cost.

Environmental Costs

A third externality is the environmental costs associated with motoring. Motorists damage the environment on both a local and a global scale. As the government will be fined if Irish emission levels are not reduced in accordance with the Kyoto agreement, it is important that a method is found to internalise the externality. While road pricing is an option to internalise these externalities, it is not the most efficient option. A more cost effective method of internalising the externality caused by emissions is to increase the tax on fuel. Noise pollution is also a feature of motoring. Road pricing would be an effective tool in combating this.

Finally, the effect of land use is identified as a negative externality. This refers to the destruction of wildlife habitats and the visual impact on the landscape. In the absence of the M50, it is unlikely that the site would not be developed. Therefore I reject the argument that there is an opportunity cost regarding the destruction of wildlife habitats. Road pricing will have no effect on the landscape. Once the motorway is built the level of traffic on the motorway does not affect the landscape.

Congestion Costs

The final externality that was identified was congestion costs. The M50 has been called 'Europe's largest car park' due to its congestion problems. It is unclear how much traffic congestion costs the economy. In 1999, the Chambers of Commerce of Ireland claimed that traffic congestion costs the economy €12.7 million per working day or €2.54 billion per annum. In 2002, Seamus Brennan claimed that traffic congestion costs Dublin €635 million per annum. The following year, the Dublin Chamber of Commerce estimated that congestion costs the economy €1 billion per annum, and directly costs business €650 million. Despite the disparity of the figures there is consensus that traffic congestion costs the economy hundreds of millions per year and that it the cost is increasing each year. Dublin Bus provides reliable annual statistics to estimate the trend of how much traffic congestion is costing the economy. Their traffic congestion costs soared to €50 million in 2003. This is a 42% increase on 2001 (The Sunday Business Post, 2003). None of the above sources mentioned costs related solely to the M50. The M50 is a unique case for a congestion charge. It is both a cause of and a solution to traffic congestion.

The M50 has positive externalities as it diverts traffic from the congested narrow streets of the city to motorway that can support a higher capacity of traffic. This allows more road space to be used for shorter local journeys within the city. However, the volume of traffic that uses the entire stretch of the M50 is very small. It is estimated that, on average, 30% of the motorway traffic exits at each interchange. If the traffic flow is followed south from the M1 interchange, 30% of it will exit at the first interchange. 30% of the remaining traffic plus the traffic that entered on the first interchange will exit at the second interchange and so on (Information provided by Eoghan Madden). In Dublin, there is a trend that an increasing amount of journeys are only short distances (Keegan, 2004). If this trend is replicated on the M50 an even higher percentage will exit at each interchange. Therefore, the amount of traffic that the M50 prevents from passing through the city centre cannot be compared with the volume of traffic that uses the M50. Only a small fraction of M50 traffic would have passed through the city centre. This positive externality is smaller than it first appeared. The negative externalities are those mentioned in the section on economic theory. They exist due to the gap between the private cost to the motorist of motoring and the cost to society of each motorist. Road pricing is an effective solution to internalise this externality.

The Capacity Dilemma of Motorways

Theory

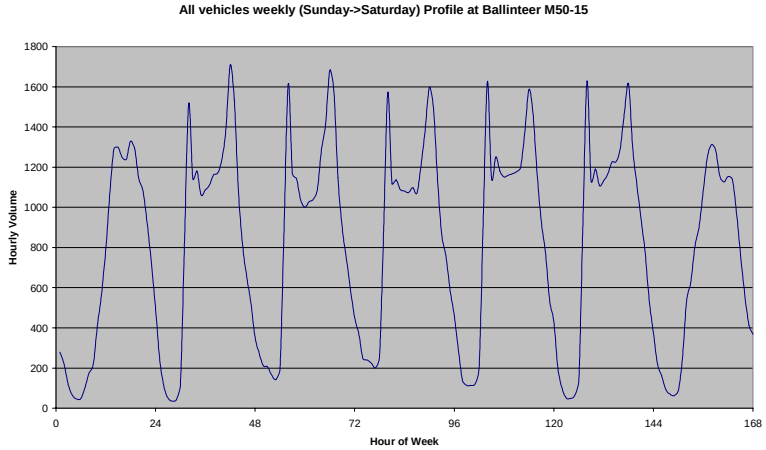
In this discussion of road pricing, I am not attempting to eliminate the gap between the social and private cost of motoring. I am discussing introducing a congestion charge that would ensure that the road network is operating to the maximum capacity possible, without causing traffic congestion. All roads have an optimum capacity of 1800 cars per lane, per hour, regardless of the road type. The optimum capacity is the maximum amount of cars that can safely travel on a road without congestion, assuming free flow conditions. In other words, assuming that there are no impediments such as traffic lights or junctions (Ibid). This is a fair assumption for a motorway. However this assumption cannot be made for the roads that feed into the interchanges. These roads are plagued with junctions, traffic lights and roundabouts. These combine to reduce the capacity of the roads to far below 1800 cars per lane, per hour. Therefore the motorway has a higher capacity than the interchanges. If the motorway is operating to full capacity the result will be congestion at the interchanges because the vehicles will have to queue for space on the local roads. The supply of space on local roads falls as peak hours are approached because local roads have the same peak times as motorways (from local traffic that does not use the motorway). So, at peak times the demand for road space on local roads is at its highest, but the supply of road space on local roads is at its lowest. The result is traffic congestion at the interchanges. If the situation is drastic enough a tailback onto the M50 will occur. This causes an impediment to free flowing traffic on the M50, reducing the optimum capacity of the motorway.

Therefore, a congestion charge that encourages a flow of traffic that ensures that the M50 is operating at its optimum capacity will not solve the traffic congestion problem. A higher charge is needed. This charge must be related to the availability of road space on local roads. This scenario will ensure that the motorway is not running to its optimum capacity. There is a trade-off between maximising the efficient use of the M50 and preventing congestion in areas surrounding the interchanges. This is the dilemma that is faced when trying to introduce road pricing for the M50.

Reality

The National Roads Authority's data confirms this theory.

Graph 1: Southbound to Ballinteer



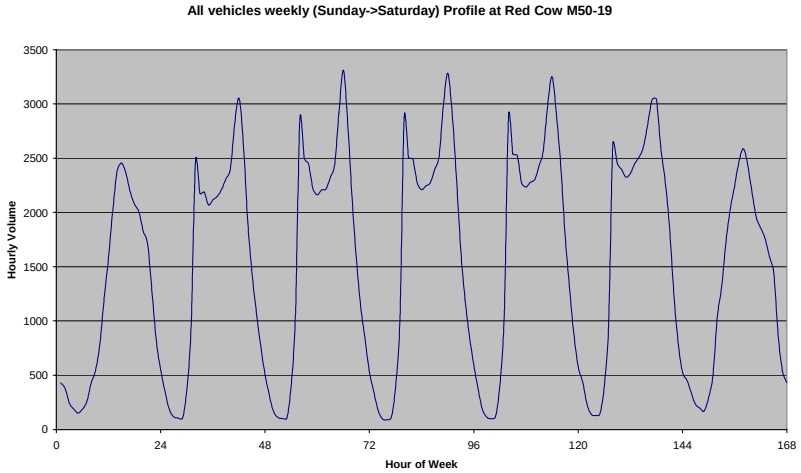
Source: National Roads Authority (NRA), Website 2004

This graph shows that the traffic volume on the final stretch of the M50, towards Ballinteer, never reaches even half of its optimum capacity. This explains why there is not traffic congestion on the motorway as Ballinteer is approached. All of this traffic must pass through the Ballinteer interchange, as this is currently the end of the motorway. Based on my observations, the overwhelming majority of this traffic then takes a one lane road to Dundrum. The capacity of this road is far less than 1800 cars per hour because of the presence of four roundabouts, several junctions, several pedestrian crossings, and traffic lights at the crossroads for Dundrum village and Sandyford. This is why congestion on the local road from Ballinteer to Dundrum is intense during peak times even though the motorway is operating at less than 1800 cars per hour. Unfortunately, Dun Laoghaire-Rathdown county council has not collected data on traffic volumes on this road, but from my observations the demand for local road space by vehicles that do not use the M50 would correlate very closely with the NRA's data.

At the Red Cow segment of the motorway (between the N7 and N4 interchanges) the M50 is not operating at its optimum capacity of 3600 cars per hour in either direction. The traffic congestion surrounding the Red Cow (also known as the 'Mad Cow') interchange has become legendary due to the gap between the

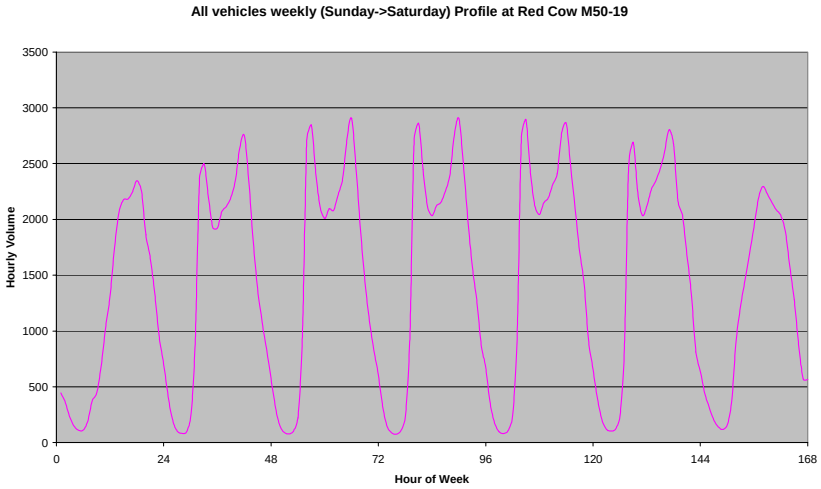
optimum capacity of the motorway and the optimum capacity of the local road network. This can be seen from graphs 2 and 3.

Graph 2: Red Cow Segment: Northbound towards M1



Source: National Roads Authority Website 2004

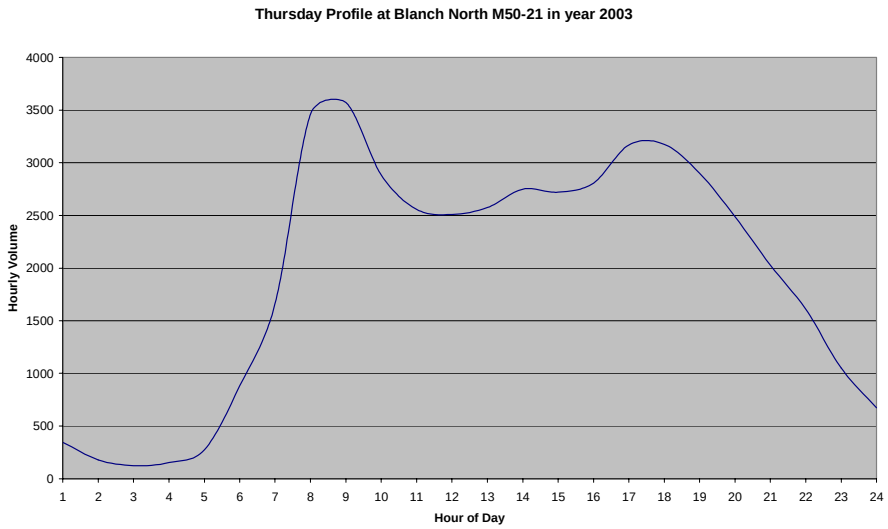
Graph 3: Red Cow Segment: Southbound towards Ballinteer



Source: National Roads Authority Website 2004

According to the NRA's data the only time when any segment of the M50 is operating at its optimum capacity is on Thursday mornings between 7am and 8am on the northbound segment of motorway between the N2 and N3 interchanges, as can be seen from graph 4. I assume that these interchanges are not famous for traffic congestion, because the local road network has a capacity that is high enough to prevent 'mad cow roundabout' style tailbacks.

Graph 4: Segment between N2 and N3 interchanges: Northbound towards M1



Source: The National Roads Authority Website, 2004

Varying Prices Depending on the Availability of Road Space in the Locality of the Interchanges

If the aim of the congestion charge was to prevent the M50 from exceeding its optimum capacity, it is clear that such a congestion charge would be of little benefit to Dublin. It would only be effective on one segment of the motorway for one hour of one day, per week. As the M50 is operating below its optimum capacity, its users are not causing congestion externalities while they are on the motorway. The externalities arise as vehicles queue for road space at the interchange. If road pricing is to become a reality at M50 interchanges, the charge must vary depending on the availability of road space on the local road network surrounding the interchanges. The local authorities must record data regarding the volume of traffic

that use these roads, if such a scheme is to be undertaken. Only those motorists that gain the highest utility from their journeys would travel on the motorway. Road pricing on motorways may cause other problems though.

Substitution Effect of Road Pricing

Motorists do not travel on the M50 for the pleasure of driving on a motorway. They travel on it to get to their destination in the quickest time possible. If the congestion charge is higher than the utility they derive from the time savings of travelling on the motorway, but the utility that they derive from the journey is greater than the extra time costs associated with using local roads to travel to their destination, they will travel to their destination using local roads. Local roads will then have a higher level of congestion. When they approach their destination, they are using up road space in the area surrounding the interchange. This increases the traffic congestion in the area surrounding the motorway interchange, and, as a result, increases tailbacks onto the motorway. This would reduce the effectiveness of a congestion charge. There is evidence from London of the substitution effect of road pricing. Approximately a quarter of motorists who stopped travelling into the centre of London due to the congestion charge are now diverting around the zone (The Evening Standard, 2004). Furthermore, approximately half of the aforementioned motorists are now using public transport (Ibid). While this is to be welcomed, it would not happen if road pricing becomes a reality on the M50 interchanges. Buses are not permitted to travel on the motorway. At present there are no public transport routes connecting the M50 interchange areas. This must be established if M50 interchange road pricing is to be successful. The absence such a public transport system would increase the number of motorists who occupy road space, causing congestion to motorway traffic, but dodge the congestion charge. This questions the viability of road pricing on the M50.

Conclusion

The idea of road pricing is based on sound economic theory. There is a strong match between economic theory and reality in Singapore and London. There would be benefits to Dublin of introducing a cost effective road pricing system in the city centre. Road pricing is not the most effective option of reducing accident externalities on the M50. The only environmental externality where road pricing is the most effective mechanism of internalising the externality on the M50 is noise pollution. Road pricing would also be effective at internalising the externality of road damage costs. Road pricing has potential to improve the congestion externalities caused by the M50. The M50 is not being used over its optimal

capacity. The lack of road space on the local roads that feed into the interchanges is the cause of traffic congestion. Road pricing at the interchanges is a possible solution, but the substitution effect of introducing a congestion charge at the interchanges is likely to be higher than has been observed in Singapore and London. Congestion on the M50 is caused by the shortage of supply of road space on local roads. Therefore, it would be more sensible to charge traffic at their destination. Electronic tolling must be introduced on the M50 toll bridge as a matter of urgency. The cost of such a system is small. A 'smart sticker' can be digitally checked as vehicles speed past the bridge. It costs 64 cent per sticker (Oscar Faber Consultants, 1999). The congestion charge recommended by the Oscar Faber Report would the most beneficial use of road pricing in Dublin. The level of technology used should be no less than an "automated data collection read only tag" and no more than "Electronic road user pricing", the system used in Singapore. When this has achieved an acceptable reduction in congestion, the issue of road pricing in the suburbs should be addressed. This would be effective in reducing demand for local roads, and therefore alleviating the difficulties associated with the M50.

Bibliography

Barrett, S. (2003), *Transport Economics Lectures* Dublin: Trinity College Dublin.

BBC News (2002), "Pay-per-mile Singapore Style." [Online]. *BBC News*. Available from: <http://news.bbc.co.uk/1/hi/world/asia-pacific/1839910.stm>

Chambers of Ireland (2004), *Chambers of Ireland.* Available from: http://chambersireland.ie/index.asp?iocid=82&docid=419

Foster, K., Lindberg, L. and McCarthy, C. (2003). "Road Congestion Pricing: Theory, Analysis and Critque" *Student Economic Review.* Vol.17.

Keegan, O. (2004), *Transport Economics Lecture*. Dublin: Trinity College Dublin.

Newbury, D. (1988), "Road Damage Externalities and Road User Charges." *Econometrica,* Vol. 56 (2).

National Roads Authority Website (2002). *Road Safety Records.* Available from: <http://www.nra.ie/roadsafety/roadsafetyrecords>

Newbury, D. (1990), "Pricing and Congestion: Economic Principles relevant to Pricing Roads" *The Oxford Review of Economic Policy.* Vol. 6 (2).

Oscar Faber Consultants. (1999), *A Study of Road Pricing in Dublin.* Unpublished, but available from ENFO, 17 St. Andrew Street, Dublin 2.

Rolle, C. (1994), "Road Pricing: The Case For and Against" *Student Economic Review.* Vol. 17.

The Evening Standard (2004), *The Evening Standard.* 02/2/04.

The Irish Independent (2004), "New Safety Barrier to Run Length of M50 Motorway" *The Irish Independent.* 15/5/01.

RTE (2004), *Prime Time.* Dublin: RTE. Aired: 11/11/03.

The Sunday Business Post (2003), *The Sunday Business Post.* 23/11/03

TAXI DEREGULATION: THREE YEARS ON

BY JENNIFER DALY

Senior Sophister

Approximately three years have passed since deregulation in the Dublin taxi market. In this paper, Jennifer Daly demonstrates how the taxi market has changed since deregulation to benefit the consumers. Although increased benefits to consumers are confirmed by the research, the author warns about the possibility of regulatory recapture and calls for further liberal reforms in the market.

According to the theory of W.J. Baumol, a contestable market is one in which there is complete freedom of entry and exit for firms:

> "It is a requirement of contestability that there be no cost discrimination against entrants. Absolute freedom of exit…is one way to guarantee freedom of entry…any firm can leave without impediment, and in the process of departure can recoup any costs incurred in the entry process" (Baumol, 1983)

This characterisation of a contestable market represents the very antithesis of the regulated Dublin taxi market up until November 2000. The idea of a contestable market approximates the 'ideal' perfectly competitive market outcomes by ensuring that the possibility of new entrants to the market disciplines the incumbents in the market to behave optimally.

Just over three years have passed since the High Court decision by Mr. Justice Murphy led to the removal of entry restrictions in the taxi market. Sufficient time has passed to determine whether the deregulated market represents a closer approximation to Baumol's contestable market and also the positive and negative aspects of the change. Ultimately, I would like to establish definitively, whether deregulation has led to a significant improvement in services, and also, the implications of taxi deregulation for other transport sectors in terms of privatisation policy, e.g. the proposed privatisation of some of the Dublin bus routes and the proposed shake up of the Aer Rianta company.

Up until 1978 there had been no restrictions in the number of taxis allowed to operate in Dublin. In the period prior to regulation, increased demand for taxis led

to a significant increase in the number of participants. Incumbents, fearing oversupply lobbied the government and forced it into a position of 'regulatory capture' whereby entry was restricted, safeguarding incumbents earnings potential. As a result of this action, in the period up until 2001, the number of taxis operating in Dublin increased only marginally. The intervention in the market disrupted the 'free hand' of the market to allow supply of taxis to be determined by demand. As a consequence, a number of negative externalities arose in the intervening years.

According to the 1998 Goodbody report (Faber,1998) on the taxi and hackney services in the Dublin area, taxi licences acquired a scarcity value and a secondary market was established whereby taxi licences traded for c. IR£90,000 and above. This was as a direct result of entry restriction, taxi licences themselves were merely a permit to operate and should not have had any significant value in their own right.

From the consumer's perspective, the results of market regulation were mostly negative. The limitation of entry placed a restriction on taxi supply at a time when demand was growing considerably. The growth in demand was due to a number of factors:

- Population growth: Between 1991 and 1996 for example, the number of adults residing in Dublin City and county rose by 10,000.
- Tourism growth: Tourist numbers also grew considerably, in the period from 1978-2000, the numbers of overseas visitors to Dublin grew from 2 million to 6 million each year (Barrett, 2003)
- Economic growth: Between 1985 and 1990, GNP grew by an average of 3.6% per annum. The arrival of the Celtic tiger in the mid-1990s brought a period of prosperity to Ireland. In the period from 1990-1997, GNP grew by an average of 5.6% per annum. Further economic developments meant that between 1991 and 1996, employment in the Dublin area grew by c. 65,000 (Faber, 1998).

These effects, together with a tightening of the drink-driving laws (Fingleton, et al., 1997) at the time translated into a significant increase in demand for taxis in terms of new users and also an increase in the frequency of travel by existing users. These effects led to a notable shortfall in taxi supply, which was estimated in the 1998 Goodbody report to be in the order of 2000 units, however, this may have been a conservative estimate.

This shortfall reflected unfavourably in the service available to customers. Waiting times at taxi ranks in excess of 90 minutes were not uncommon at peak

times. Also, 75% of people surveyed on street at the time felt that taxis were not easily available at peak times. In a similar vein, 46% of calls to cab companies led to wait times in excess of 20 minutes, while some calls were unable to secure a pick-up within a reasonable period. The overall customer perception from the study was that there was a lack of available cabs (Faber, 1998).

There appears to have been little reasoning for the 1978 regulation policy. General economic rationale for regulating entry may be based on a number of possibilities:

- Imperfect competition may result in a level of entry above that which would be socially optimal.
- A tendency towards inefficient entry may occur.
- Restrictions on entry may be used to correct for market failures other than excess/inefficient entry (Competition Authority, 2002).

In the first two cases here, it is probable that the regulator would not possess all the information necessary to determine which effect dominates and to be able to observe changes. We must not neglect the fact the regulation causes its own inefficiencies particularly in a situation in which incumbents are allowed to influence the regulator. The Competition Authority's report (2002) suggests that "regulatory capture is the most likely explanation for the quantitative restriction placed on taxis between 1978 and 2000". This indicates the possibility that the decision to regulate was not for consumer benefit, but rather reflected the preferences of vested interests, namely, the taxi drivers themselves. Furthermore, as regards the third rationale for entry regulation above, evidence from the 1998 Goodbody study indicates that restriction of entry has not altered firm behaviour in a socially optimal way (Faber,1998). The Competition Authority goes further by stating that, as a mechanism for correcting market failures; entry regulation is a "blunt instrument" going on to say that there are "no convincing economic rationales for placing quantitative restrictions on the number of operators in taxi markets".

The 21st of November 2000 signalled a change in the awarding of licences. A high court decision by Mr. Justice Murphy on the 13th of October taken by hackney drivers Humphrey and Others against the Minister for the Environment and Others led to a removal of entry restrictions in the taxi industry. The action by Humphrey et al. questioned the decision taken by the Minister for the Environment and Local Government to increase the supply of taxis in Dublin by awarding an additional taxi licence to each existing licence holder. The judgement by Mr. Justice Murphy indicated that such a policy would be discriminatory and contrary to EC Treaty rules, in that it would restrict new licence holders to those mainly Irish nationals who were already operating. The decision by Justice Murphy stated: "any suitably qualified individual may provide taxi services". It was also stated in the

report that "regulations which restrict the number of public hire vehicles contradicted the very concept of public service" (Murphy, 2000).

The subsequent removal of entry restrictions led to an immediate increase in the number of taxis operating around the country. Taxi representative bodies responded by organising strikes and demonstrations opposing the decision to deregulate. In this instance, the government did not bow to industry pressure. The Taoiseach asserted at the time that "there can be no change in the government's position on deregulation".

The persistence of the policy of regulation of the taxi industry for a period of 22 years indicated the strength of the tendency of the government to protect incumbents. In its 2001 report, the OECD stated: "policy must move away from emphasis on protection of incumbents against innovation and competition".

The report was also critical of government efforts in this area stating: "the coming cycle in Irish Economic Development justifies a more coherent and determined approach to regulatory reform than seen to date" (OECD, 2001). The 2000 deregulation policy was one step towards a "more determined approach" on the part of the government. The removal of entry restrictions effectively removed the secondary market for taxi licences and with it the exorbitant cost of entry for new participants. The change signalled a removal of "discrimination against entrants" and so is now more akin to a contestable market.

In order to determine whether the taxi market in the post deregulation period represents an improvement in services to passengers, I undertook a study of waiting times for both taxis and passengers at the College Green rank in the city centre on Thursday 29th of January. The study was conducted over three periods during the day i.e. 9am-10am, 4pm-5pm and 11pm-1am. I will compare these results to those obtained from the 1998 Goodbody study (outlined above), to determine the level of improvement in the service since the removal of entry restrictions.

- *9am-10am:*

  I considered this period to be a peak time given that it was a busy weekday morning and the level of congestion observable. During the period there were 57 pick-ups. I found that of these pick-ups, over half of passengers had no wait at all at the rank, while for those who did queue, the average wait for a taxi was only 3 minutes 20 seconds. The longest passenger wait was 9 minutes 20 seconds, but this was unrepresentative of the period as a whole.

  For taxis, 41 had to wait at the rank for hire, the average wait being 2 minutes 9 seconds. In general, over the period I found that taxis were waiting on passengers rather than the other way around and that service was brisk and queues rare.

- *4pm-5pm*

  Again, I expected this to be a busy period, though it was less busy than the morning period with 49 pick-ups. In this instance, no passengers had to wait for taxis. All taxis had to wait for hire at the rank, though these periods were generally short, the average wait being 3 minutes 21 seconds and the longest, 8 minutes 26 seconds. The average number of taxis at the rank was 3.75, the maximum being 6 at any one time.

- *11pm-1am*

  This period was the busiest with 95 pick-ups in the first hour and 87 in the second hour. Again I found that customer demand was met effectively, as only 12% of passengers had to wait for a taxi. However, the average wait in this instance was very small, the average being 21 seconds and the longest, 35 seconds. The average wait for taxis was again small, as taxis waited for an average of 1 minute 46 seconds in the first hour and 1 minute 52 seconds in the second hour. The longest wait for a taxi was 12 minutes; however, only 3 taxis wait approached this length. Only 6% of taxis had a wait of longer than 5 minutes. I observed that at this time there was very little congestion and an abundance of taxis visible for hire, both at the ranks and on street.

From my own study and in comparison with the 1998 Goodbody report, I can conclude that things are indeed better for taxi users in 2004, c. 3 years after deregulation than when the taxi market was regulated. In that, deregulation has had a positive effect in terms of service to customers.

The 1998 Goodbody report indicated that there was a shortage of taxis in operation at the time, that lengthy wait times (in excess of 90 minutes) were common and that there was overall dissatisfaction among consumers re: taxi availability. The report concluded that:

> "Restrictions on market entry should be avoided as regulatory authorities are unlikely to have sufficient foreknowledge as to provide the optimum supply of taxis. In the longer term, therefore, economic theory and experiences elsewhere support a policy of open entry in the taxi market". (Faber, 1998)

On the basis of my own findings above, I suggest that deregulation has had a positive outcome. Many of the negative externalities associated with supply problems have been eliminated, such that taxis waiting at ranks is the standard, rather than the lengthy queues mentioned in the Goodbody report. In contrast to the 90 minutes wait times mentioned in the report, the longest wait time observed in the

study on the 29th of January was only 9 minutes 20 seconds, which was by no means representative of overall wait times as the majority of passengers had no wait at all. Thus, even from elementary comparisons it is possible to conclude that there has been a dramatic increase in the availability of taxis for consumers. There are no supply problems even at peak times and taxi travel has become a feasible transport option for a greater number of people.

In spite of the success of deregulation in this regard however, there remains a threat to the current system in terms of possible 'regulatory recapture'. One could suggest that with current supply in excess of 10,000 units, if entry is once again restricted, the externalities will not be as severe as those previously. However, given current growth trends, the problems posed by regulatory recapture are not insignificant. Such a move would also prove detrimental to efforts to reduce entry barriers in other transport sectors. Efforts must be made to ensure that lobby groups do not have the power to sway government policy. Recent examples of attempts at regulatory recapture may be observed in recent press releases. For Example, a recent article in the *Irish Independent* entitled *How Safe are Dublin Taximen*, in which, the vice-president of the National Taxi Drivers Union (NTDU) tries to draw attention to 'problems' caused by deregulation. Safety fears are also highlighted. TD Roisin Shorthall is quoted as saying that:

> "Before deregulation, there was a fair degree of assurance that if taxi drivers had forked out up to €100,000 for a licence, they were highly unlikely to jeopardise their investment by carrying out any kind of assault. The very fact that the industry was so difficult to get into was a kind of guarantee of safety to passengers" (Irish Independent, 2004)

This kind of sentiment is misleading to taxi users, as this is not an adequate argument for regulation of entry, but rather for quality and safety standards for drivers and vehicles. A similar article from the *Sunday Independent* recently was headed *Rise In Attacks By Taxi-Drivers Leads To Calls For Checks*. In this article, Vinnie Jones of the NTDU claims: "even where applications for taxi licences have been refused, some applicants with dubious backgrounds have been successful in making challenges in the courts" (Sunday Independent, 2004). These types of scare tactics are, in my view another attempt at regulatory recapture and echo the sentiment expressed in the Taxi Forum release in 1998:

> "Taxis with regulated fares and high cost of entry need to be protected from the hackney business where fares are not regulated and significantly lower costs of entry apply" (Dublin Taxi forum, 1998).

This release also suggested a points system for the awarding of licences and a phased increase in taxi licences up to 3,200 in 2002 (Ibid). It is obvious that this

system would have been wholly inadequate in dealing with supply problems of the time and is merely an attempt to maintain some kind of restraint on the number of operators with a view to maintaining the monopoly rents earned by the incumbents.

What is required is the type of policies advocated by the 1997 report: *The Dublin Taxi Market: Re-Regulate Or Stay Queuing?* This report highlights the need for free entry to the market to be accompanied with a strengthening of fare controls and safety standards. They indicate that safety fears and quality standards do not provide a case for regulation of entry:

> "The only systematic problem that has accompanied entry deregulation is that quality standards have fallen as the market expands. However, this is a reason for improved quality standards rather than an argument against removing entry controls" (Fingleton et al, 1997).

It is imperative that this is the view taken and upheld by the government also. A possible source of concern in this regard is the appointment of an interim regulator in 2003, whose responsibilities include bringing stability to the industry and establishing lasting career opportunities for all those involved (Barrett, 2003). Although regulation is welcomed in the areas of fare and quality control etc., there are dangers in giving regulatory powers to one office in that this office provides an easy target for lobbyists leading to pressures, which could threaten the improvements brought about by liberalisation.

The breakdown of the current system could be harmful to policies aimed at increasing competition in other transport sectors (e.g. bus privatisation) for which the taxi model may provide a blueprint. For example, falling safety standards in the taxi market may be used as an argument against deregulation in other sectors.

Successes in this area invariably spread enthusiasm for the benefits of competition elsewhere, the deregulation of the city centre - Dublin Airport bus route being a case in point. TD Des O'Malley (2000) commented on the success: "deregulation should spread the benefits of competition through the country"

Conclusion:

There is ample evidence that the benefits of taxi market liberalisation have been substantial in terms of improving the availability and accessibility of taxis to consumers. The improvement of contestability has allowed the market to achieve more optimal outcomes. However, we must note that taxi liberalisation does not end with freedom of entry. To ensure the smooth running of the market, to eliminate the threat of regulatory recapture and to provide an exemplary model to liberalisation policies elsewhere, other reforms are also necessary. The Competition Authority presents this view in its 2002 report:

> "More systematic regulatory reform is required if the taxi market is to function efficiently and deliver maximum benefit to consumers, the taxi industry and the overall economy…otherwise there is a risk that the deregulation of entry to the taxi market in isolation, could undermine support for regulatory reform more generally" (Competition Authority, 2002).

It is evident that, in respect to deregulation and privatisation that the Irish Government lags behind some of its international counterparts (OECD, 2001). So it is imperative that the taxi liberalisation policy does not provide arguments against liberalisation in other sectors of the economy. In this way there will be no impediment to the benefits of competition in the economy and the government can switch its attention away from market intervention towards the provision of pure public goods, which is invariably the socially optimal outcome (Barrett, 2003).

Bibliography

Barrett, S.D. (1982), *Transport Policy in Ireland.* Dublin: Irish Management Institute.

Barrett, S.D. (1991), *Transport Policy in Ireland in the 1990's.* Dublin: Gill and Macmillan.

Barrett, S.D. (2003), "Regulatory Capture, Property Rights and Taxi Deregulation: A Case Study" *Economic Affairs*, Vol. 23(9)

Baumol, W.J. (1983), "Contestable Markets: An Uprising in the Theory of Industry Structure." *American Economic Review,* Vol. 72

Competition Authority (2002), "Submission on Qualitative Improvements in Taxi Services and Future Regulation of these Services to the Department of the Environment and Local Government" [Online]. *Competition Authority.* Available from: <www.tca.ie/submissions.html>

Faber, O. (1998), *Review of the Taxi and Hackney Carriage service in the Dublin Area: Final Report.* Goodbody Economic Consultants: Dublin.

Fingleton, J., Evans, J., and Hogan, O. (1997), *The Dublin Taxi Market: Re-Regulate or Stay Queuing?* Dublin: TCD Department of Economics.

Irish Independent (2004) *Irish Independent.* 20/01/04

Murphy, J. (2000), *Humphrey and Others v. The Minister for the Environment, Local Government, Ireland, the Attorney General and Others.* Mimeo.

OECD (2001), *Regulatory Reform in Ireland.* Paris: OECD.

O'Malley, D. (2000), "Sharing our Success." *Irish Independent*. 20/01/04.

Sunday Independent (2004), "Rise in attacks by taxi-drivers leads to calls for checks." *Sunday Independent.* 18/01/04.

THE CONTESTABILITY OF MARKETS

BY JONATHAN MILLER

Senior Sophister

In this essay Jonathan Miller draws our attention to one of the dominant issues in European Transport Policy – that of the contestability of markets. The author applies Baumol's theory to the airline industry, highlighting the barriers to contestability and providing an illustrative case study of the US airlines.

Introduction

European Transport Policy in the 21st century is dominated by four main issues: social cost, value for money, safety and contestability. This essay addresses contestability by introducing Baumol's Theory of Contestability (1982) and also by discussing deregulation. The theory will be applied to the transport market. The topical air transport industry[1] will be the primary focus within the market since Baumol's theory postulated that airline markets could provide a particularly close approximation to contestability (Baumol and Bailey, 1984). Furthermore, the airline industry gives the added bonus of providing many barriers to contestability. This will be commented on at a later stage, followed by a case study on USA airline industry.

In a contestable market, potential entrants can easily challenge the position of incumbents. A perfectly contestable market occurs when entry is free. This implies that there is an absence of sunk costs, costless reversible entry exits, and incumbents compete on symmetric terms with entrants.[2] Hence, any cost differences arise from price or quality differences in output since there are no regulatory restrictions or threat of predatory pricing. Under these assumptions the knowledge that firms will enter if supernormal profits are being earned means that even a small number of incumbents will sell at a competitive price. Baumol found that some

[1] Examples include Aer Lingus marketing and change in service, state aid to EU national flag-carriers, Ryanair's cheap tickets, the collapse of Swiss Air, Air France amalgamations, to name but a few.

[2] Access to the same production techniques and input prices exist and there is no subsidy being awarded to solely the incumbent or the potential entrant

industries with small numbers of firms are in fact highly contestable and claimed that high concentration with low prices can be seen as a virtue rather than a vice.

Welfare Features of a Perfectly Contestable Market

- Zero profits constitute equilibrium because if the incumbent earns any supernormal profits, it would automatically constitute an incentive for a potential entrant to 'hit and run'.
- There is an absence of productive inefficiency in long run equilibrium, or efficient entrants would contest the market.
- No product can be sold at a price greater than marginal cost in long run equilibrium therefore cross-subsidisation is impossible.
- These characteristics keep the incumbent efficient, as potential entrants can force monopolists or oligopolists to act in a socially optimal manner.[3]

Benefits of Deregulation

Microeconomic forces have provided increasing evidence that the costs of regulatory failure are not actually worth the reduction in market failure that they were designed to achieve. Examples include the efficiency savings of private bus operators in Sydney, British Coachways in Britain, Aircoach and other small private bus operators in Ireland. Extreme cases have led to regulatory capture where competing (usually the incumbent) firms are able to influence the regulatory body, e.g. CIE,[4] EU state airlines in the 1990s,[5] taxis in 1990s.[6]

Deregulation is considered procontestable as it causes a higher industrial output, which leads to greater output and productivity and lower prices. This confirms with the Theory of Contestable Markets and its applications, which were conceived primarily with consumers' interests at heart.

---

[3] Ryanair made British Airways and Aer Lingus change their entire approach to business operations. British Airways pulled out of low cost city routes, which were in direct competition with Ryanair, Buzz and Go. Aer Lingus has now become a "no frills" airline with much administration imposed on the passenger rather than the airline or travel operators, thus reducing cost.

[4] Especially during the period of nationalisation and controlled competition 1930-1974. Further liberalisation for private bus operations in Ireland is required to increase contestability and in turn competition. At the moment, it can take up to a couple of months for a private bus operator to obtain a route licence while CIE can obtain one nearly overnight.

[5] Aer Lingus, Air France, Iberia and British Airways

[6] A Dublin taxi plate valued at €100,000 before deregulation to €5,000 after. Only 3,913 taxis were on the streets on the eve of deregulation. This number had reached 11,630 two years later. See Fingleton (1998) and Barrett (1982).

There has been much deregulation throughout numerous sectors of the economy since the early 1980s with varying success. Freight haulage is probably the most successful deregulated transport sector where contestability has increased substantially. Ireland completed deregulation of the haulage sector by September 1988, with quality controls replacing quantity controls on market entry. Quality controls can be used to combat externalities and along with private sector investment, there exists no need for governmental intervention, the West Link and East Link toll bridges provided by National Toll Roads being a prime example.[7] Originally restrictions imposed on the haulage sector in 1933 were designed to assist the railways, but their main effect was to increase transport costs and to generate a scarcity value for haulage licences similar to those of taxi plates in Dublin in the 1990s. With deregulation, road haulage increased from 17% to 72.4 % of the market share, to the detriment of the rail freight (Barrett, October 2003)

The Airline Industry

A healthy airline industry is, in a similar fashion to telecommunications and finance, essential for the prosperity of other business. Around 1.25 billion people a year travel on aeroplanes, spending $250 billion and providing jobs for about 1.5 million people. Perhaps more importantly, airlines are a key element of the larger business of travel and tourism, which has revenues of $3,400 billion a year and which makes up about 10% of world GDP, in employing 200 million people (Asia-Pacific Technology Network, 2004).

For the past half-century, growth rates in the airline business have consistently been above world GDP growth, with the exception of 1991, largely due to the Gulf War, and in 2001/02 in the aftermath of September 11th. In spite of this strong growth, profit margins have typically been low and some large carriers have either made large losses or have disappeared altogether.[8]

Airlines have many of the features frequently found in nationalised industries. They have strategic importance, potentially large externalities, high fixed costs and contain elements of national pride. Unsurprisingly, for most of the post-war period the industry has been dominated by state-owned 'flag carriers'. Governments are able to control the amount of competition their domestic carriers are exposed to, which is a key factor in any airline's profitability, whether state-owned or private. For example, while some would argue that British Airways' post-privatisation record is due to efficiencies imposed by free market pressures, others

[7] It will be interesting to see of the tolling of the Irish motorways will also provide empirical evidence to the argument that regulation is not required and that private public partnership can optimise consumer welfare benefits.

[8] TWA, Swissair, PanAm

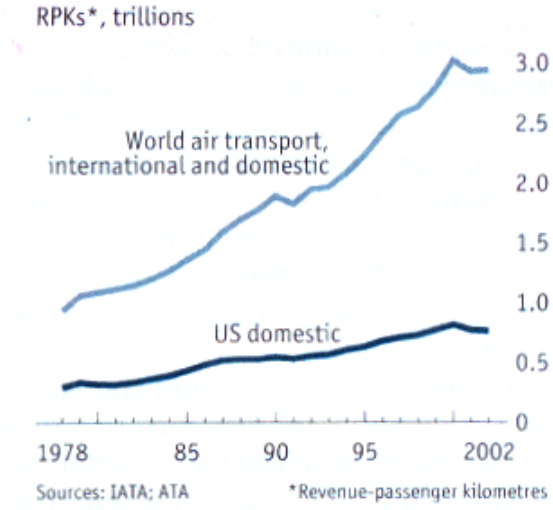
would claim that healthy profits are more attributable to the continued preferential treatment it receives at Heathrow airport through grandfather rights and slot control.

However, over the last 15 years or so the airline industry has been undergoing fundamental structural change. In the past, it was a global industry dominated by national firms. More recently airlines have revealed ambitions to become global firms, through a programme of mergers and alliances. The reasons are similar to those for telecommunications firms, rising fixed costs and increased importance of access to new markets in a global economy.

Liberalisation of the Air

The US Airline Deregulation Act, 1978, freed airlines of government control over airfares and routes, which allowed some airlines to develop large domestic control, like Delta Airlines. Bailey (1986, In: Williams G., 1994) states the benefits of procontestable policies as reducing average prices, time spent travelling, real cost of airline operation and cross-subsidisation, whilst increasing efficiency via computerised routing. These in turn benefit the consumer by providing a wider variety and better service at a lower price, leading to an upsurge in air travel, which rose by 150% over the last 20 years (See Figure 1). With the developed world no longer having a monopoly in technology, capital and human skills, this competitive spur is required to maintain its market shares (Scamorana, 1993).

Figure.1: Explosion in US Air Travel after Deregulation



Source: The Economist October 14th 2003

The gains from European airline deregulation have been more dramatic than in the USA[9] and of all European countries, Ireland was one of the first and most dramatic cases of fare reductions and traveller volume increases. On the first day of deregulation, 23rd May 1986, the Dublin-London airfare was reduced from £208 to £94.99 (€264.11 to €120.61), a fall of 54%. The current average fare per journey on Ryanair, the Irish start-up airline under deregulation, is €41 and the airline claims cuts of as much as 85% off the fares charged by traditional European national airlines. Passenger numbers between Dublin and London in August 1987, the first complete year of deregulation, were 92% above those in August 1985. Although the population of Ireland is just over 4 million, more people fly between Ireland and the UK than between the UK and France, Germany and Italy combined (Barrett, 2003). Ryanair was the first start-up airline in Europe to carry more passengers than its national airline, and with 24 million passengers in the current year, it is likely to become the largest international passenger airline in the EU in 2004. Ryanair's passengers may exceed 30 million passing British Airways and Lufthansa, each currently with c.29 million passengers in Europe (O'Leary, 2004).

Bailey's 1984 article led to the acceptance of the first of two hypotheses tested by Graham, Kaplon and Sibley that national airlines used to employ excess capacity relative to that in unregulated competition, resulting in inefficient use of resources. This fact can be easily seen by the fact that Ryanair's growth in the airline market has led to main airline carriers tightening staff quotas to cut unneeded costs. The second hypothesis that potential competition would keep fares at cost in highly concentrated markets was refuted. They found that potential competition alone is not enough to reduce fares since they were closely related to market traits such as time-distance sensitivity. Kenneth Button (1982) wrote that since buses are an experienced good, reputation was a significant factor. Potential entry was not enough to reduce fares. It would be interesting to see if this hypothesis still holds true in light of the events of September 11th 2001 and whether safety is now a factor that must be built into the business model and, as it is an externality, whether the model should be altered to reflect the recent change in people's concerns? Similarly, should externalities such as health concerns caused by deep vein thrombosis be incorporated into the 21st century model, or will large settlements arising from proceedings taken by claimants of the airlines be required before they are incorporated into economic models as a real and substantial cost?

Deregulation has allowed low cost entrant airlines into the industry who, unlike the incumbents, are not tied into high cost labour agreements with powerful

[9] European air fairs before deregulation were the highest in the world, according to an ICAO survey.

unions.[10] The incumbents' attempt to overcome the disadvantage of a different cost structure was originally a two-tier labour structure. This entailed existing staff being paid the original wage but new staff receiving a more competitive lower wage. This approach was thwarted by disruptions and strikes.[11] Baumol seems to have overlooked this dilemma when choosing air transport to support this theory.[12]

The existence of sunk costs provides an advantage to incumbents, which seems to prevent entry and cannot be simply legislated away. In this way, railways have not been considered a contestable market due to the existence of substantial sunk costs like the laying of tracks.[13] However, the airline industry has the ultimate in mobile assets and a highly developed resale market. Bearing this in mind, why in the past was so much of the market under state ownership and regulated?

National Monopoly or National Pride?

Airlines were traditionally seen as a public utility, which helps to explain the prevalence of state ownership and heavy regulation. However, they cannot be seen as a natural monopoly as the cost implications of more than one airline serving one route are not the same as those of more than one water company laying down pipes.

It is possible to argue that there exist elements of natural monopoly in the provision of airport facilities,[14] but the same is not necessarily true of the airlines, which serve these airports. Whilst airlines may have heavy fixed costs in terms of their aircraft and computer systems, the assets are both mobile and escapable. Statistical studies have shown that there is little relationship between size and average cost, once a minimum efficient scale of operation had been reached. Returns to scale are constant, so large airlines do not automatically have a cost advantage over smaller ones. It would therefore appear that governments choose to regulate or

[10] Ryanair's low cost base would cause Aer Lingus to make losses at Ryanair's prices

[11] Michael O'Leary often states that his workers receive a higher wage than high cost airlines due to the workers higher productivity. With 24 million passengers and 2,200 staff, Ryanair has 10,900 passengers per staff member per year in contrast with 800 in member airlines of the Association of European Airlines. Aer Lingus had much union trouble since the implementation of its restructuring plan in 2002 but the Aer Lingus Board headed by Tom Mulchay have recognised that this is the only plausible method to make the airline profitable again and eventually lead it to its initial public placing as was originally intended by the government in 2000.

[12] Baumol and Willig do state in a later article that TCM is an ideological economic market. Prefect contestability is not primarily a description of reality but rather a benchmark for desirable industrial organisation.

[13] Even the privatisation of locomotives failed to make the British railway profitable

[14] This argument is becoming obsolete as more people are willing to fly to smaller airports further outside the city in return for cheaper flights.

own airlines because of the externalities, the national image and the strategic factors involved. Much of the rationale for government regulation is based on non-economic factors. To understand this, consider the domestic reaction, politically and popularly, if a foreign carrier were to make a take-over bid for Aer Lingus and the union strife that would ensue.

Ports following Lines

Airport competition has been an unprecedented result of airline deregulation in Europe. The traditional European hub airport was a high-cost, non-competing facility, controlled by the host national airline through a scheduling committee, which allocated slots to airlines in order of seniority, through the grandfather rights system. Without airport competition, the Single European Market in aviation would have afforded few market opportunities for new market entrants who would have had freedom to enter the market but would not have had access to hub airports in a sufficient number of slots to make a significant market impact. Airports in Europe are thus becoming contestable markets with passengers quite willing to transfer to new airports and low-cost airlines providing services not supplied by the incumbent airlines in the past.[15] Airport competition appears sustainable because of the large number of underused airports in Europe. For example, Charleroi Airport, Brussels, has a terminal capacity of 1.8 million passengers and a runway capacity of 10 to 15 million operating 24 hours a day. In 2002, the number of passengers at the airport was 1.5 million. Airport and airline competition has also brought downward pressure on airport charges at the hub airports. At Dublin, Aer Lingus has joined with Ryanair in its criticisms of airport charges. According to Barrett, the scope for savings from airport deregulation ranges from 50% to 92% (Barrett, 2003). However, these savings could be diminished as Ryanair was ordered on the 3rd February 2004 to pay back €4.5 million in illegal subsidies in a landmark European Commission decision. The Commission said cut-priced rates offered exclusively to Ryanair for using Charleroi Airport in southern Belgium were illegal. The airline will not have to repay all its subsidies from the Belgian regional government, but Michael O'Leary sees the decision as a "disaster for consumers". It could lead to similar action been taken in other regional airports, in some cases making it unviable for low-cost airlines to continue operating certain routes (The Irish Times, 2004).

[15] New airport market entrants under deregulation include Stansted, Prestwick, Charleroi, Beauvais, Hahn and Lubeck

Economies of Scope

In the airline business the efficiencies due to size are more correctly termed economies of scope, rather than economies of scale. The latter are typically found in manufacturing when the cost of producing an extra unit of output is lower than the average cost of the output, which has gone before it. Essentially, this is due to the high cost of the investment needed before production can begin - what economists term start-up fixed costs.

The airline industry reaps economies of scope when the cost of supplying two products jointly is cheaper than producing them separately. These economies usually relate to the size of an airline's flight network. For example, advertising costs are not aimed at a particular route, but at the airline's whole network, which can be termed an economy of scope. Large networks also give opportunities for economies of scope in areas such as frequent flyer schemes to generate customer loyalty and computer reservation systems.

Deregulation tends to result in lower prices because in addition to stronger competition on routes, it generates increasing economies of scope by allowing network size to increase, whether by corporate expansion, mergers, or alliances. As a result, airfares have risen much slower than the rate of inflation in recent years (Asia-Pacific Technology Network, 2004).

Barriers to Contestability

In US internal aviation the consumer welfare gains from deregulation have been estimated at $20 billion a year, but incumbents have many structural and strategic anti-competitive advantages over potential entrants including:

- Hub airport dominance/Grandfather rights are the control of major airports by incumbent airlines and the restriction of access to these airports for new airlines thus reducing the latter's ability to entice passengers. Airport slots should be allocated by lottery, franchise or auction instead of senior scheduling committees. It was considered the most important barrier to contestability. However, as mentioned above, many customers have made the choice of flying to more remote destinations at a reduced cost with new low fare airlines.
- Ground Handling Monopolies were restricted to airlines with grandfather rights thus enabling them to charge prices for this service to smaller airlines. This limited the price-competitiveness of potential entrants. EU deregulation of handling rights lead to the appointment of competing specialist baggage handling groups,

which resulted in some cases to costs to smaller airlines being halved.[16]

- Computer Reservation Systems (CRS) of incumbents must be used by potential entrants because of economies of scale. CRS are used to bias travel agents' displays against small airlines although, neutrality should exist. This problem has been overcome in recent years with the advancement of internet technology. Selling tickets over the Internet is a more cost efficient and productive means of sale. The reduced cost of ticket sales has led to low air fares and more airlines dealing on-line, like *aerlingus.com*
- Geographical Price Discrimination is the strongest obstacle as it allows incumbents to undercut new entrants on competitive routes by cross subsidising with the profits of uncontested routes (Kahn, 1988). This was the most destructive post-deregulation obstacle to efficiency in the U.S.,[17] so the EU Commission have wisely initiated and implemented a package, which can investigate fares which are either too high or too low. Similarly price collusion between incumbents must be broken and national airlines' anticompetitive Article 81.3 status, exempting them from competition policy, must be lifted.
- Frequent flyer programmes place new airlines at a disadvantage as with a bigger network the passenger can maximise benefits by flying with a larger airline. It is also difficult to regulate. However, it is becoming an issue only affecting the executive traveller flying on company business who collects the air miles for his own use, rather than the tourist traveller who has little or no carrier loyalty.
- Mergers and Concentrations as potential entrants are taken over by larger airlines, continuously reducing the contestability of the market. Mergers[18] have made it more likely that deregulation will not reach its full potential (Morrison and Winston, 1986).

---

[16] British Midland International halved their costs when they started handling their own baggage instead of having to pay British Airways.

[17] See Case Study on Predatory Practices in the US Airline Industry.

[18] As a prime example, Air France and KLM Royal Dutch Airlines are to come as close to a cross-border merger as the industry's rules currently allow. They hope to circumvent the rules that tie traffic rights between two countries to domestically owned carriers. The two airlines will be owned by a joint holding company, but will operate as separate brands from their bases in Paris and Amsterdam. By obtaining KLM, Air France will dominate two of the four long-haul hubs in Europe and increase its annual passenger load to 42 million people. The deal has to win clearance from both the American and EU antitrust authorities.

thus merger policy should be more rigorously enforced. Strategic alliances also increase airlines economies of scope but without the complication of government intervention. The three main forms of alliances are code sharing, block spacing and franchising. Chataway (1992) suggests an independent, non-political European Agency, which would make decisions purely on competition grounds. This could be further developed to look after safety and health aspects, which now concern a majority of passengers.[19] This power could be removed from the bodies such as the CAA and IAA, and given to a European Agency. Whilst the CAA and IAA are not politically inclined, there must be sizeable duplication of costs by having so many regional authorities.

- State Aid in the past to national airlines seriously distorted contestability, allowing below cost selling by incumbents, therefore placing entrants at an unfair disadvantage. Although not the recipients of state aid, Chapter 11 bankruptcy laws still permit bankrupt US airlines to operate without having to make interest or pension fund payments, thus reducing ease of exit. This in turn reduces the benefit of procontestable policies (Barrett, 1994). In an attempt to create a more level playing field, regulations limiting when state aid could be provided were issued under Articles 87 and 88 of the EU Treaty. Aid is permitted in areas where there has been economic disruption. This is an area, which state airlines play to, examples including September 11th and possible terrorist fears weakening the aviation market. Whilst the administration and legislation is in place the European Commission must make sure that any aid given follows certain criteria, and that it is a once off improvement package, which has to be implemented correctly by the receiving airlines. Even though the EU has strengthened competitiveness by limiting when state aid is granted, it unfortunately lacks real enforcement.

Case Study on Predatory Pricing in the US Airline Industry in the early 1990s

In the early 1990s, after an economic recession, a new wave of start-up airlines began operating and by the mid-1990s, one in five travellers were flying with these small, low cost companies. It was therefore a matter of concern to the Department of Transportation (DOT) when the new entrants complained that they

[19] These could be incorporated into a minimum safety and health standard without which airlines cannot operate and the costs of which are borne by the airlines.

encountered sharp price-cutting by major incumbent carriers, particularly when entering concentrated hub markets. The DOT questioned whether incumbents were setting fares well below cost in an effort to divert customers away from the new challengers, seeking the demise of the new entrants in order to raise fares back to much higher, pre-entry levels. There were also reports of incumbents using their long-term leases and other airport contractual arrangements to exclude challengers; for instance, by refusing to sublet idle gates.

Concerned about possible predatory practices in the airline industry, and recognising the uncertainty and expense involved in trying to prove such conduct through the courts under traditional antitrust law, the DOT offered its own criteria for detecting predatory pricing. It proposed an administrative enforcement and investigative process to police unfair competition in the airline industry. Sharp price cutting and large increases in seating capacity in a city-pair market by a major airline in response to the entry of a lower-priced competitor would trigger an investigation and possible enforcement proceedings.

The DOT's proposal, in April 1998, prompted strong reactions. It was lauded by some, including many start-up airlines, as a necessary supplement to traditional antitrust enforcement, giving new entrants the opportunity to compete on the merits of their product. Others, including most major airlines, have criticised it as a perilous first step towards re-regulation of passenger fares and services and as incompatible with traditional antitrust enforcement. In addition, in May 1999, the Department of Justice filed a civil antitrust action against American Airlines, claiming that it engaged in predatory tactics

Conclusion

Privatisation assists in contestability being obtained in that any high airfares charged would be reflected in supernormal profits. These attract more attention than economic rents dissipated in inefficiency or labour costs. Privatisation also ends the monopoly of regulatory capture by Europe's airlines over their 'parent' government departments.

Baumol's theory was justified by the massive success of deregulation in all areas of economic life over the past 20 years including telecommunications and utilities. To prevent contestability becoming a monopolistic damp squib (Pryke, 1987) the government and EU must enforce and further develop the recently introduced procontestable policies. More than ever, regulation has an important role to play in tackling barriers to contestability. Contestability has been assisted greatly by advances in communication technology enabling consumers to obtain a high degree of information and price comparability at low cost e.g. the through internet

and its search engines. Lobbying against anti-competitive and concerted practices as illustrated in cases in the United States, Ireland and the EU[20] suggest an ever-increasing trend towards a level playing field. For the sake of efficiency and moreover for the sake of the consumer, let us hope that Baumol's ideological market structure can be illustrated by the airline industry.

Bibliography

Bailey, E.E, and Baumol, W.J. (1984). "Deregulation and the Theory of Contestable Markets". *Yale Journal on Regulation*, Vol. 1(111).

Barrett, S. (2003). *50 Years of Transport Research: Experience Gained And Major Challenges Ahead*, presented in Budapest, Hungry, October 2003.

Barrett, S. (2003). "Regulatory Capture, Property Rights and Taxi Deregulation: A Case Study". *Journal of Economic Affairs,* Vol. 23(4).

Barrett, S. (1994**).** "Aviation and the Pursuit of Contestability". *Journal of Economic Affairs,* Vol. 14(2).

Barrett, S. (1982). *Transport Policy in Ireland*. Irish Management Institute, Dublin.

Baumol, W. (1982). "Contestable Markets: An Uprising in the Theory of Industrial Structure." *American Economic Review*.

Baumol, W.J, Panzar J.C, and Willig, R.D, (1986). *Contestable Markets and the Theory of Industrial Structure*. Harcourt-Brace-Jovanoich, San Diego.

Button, K, (1982). *Transport Economics*. London: Heinemann.

Button, K. (1989). *EC Airlines and Their Problems*. Mimeo.

Chataway, C. (1992). *Competing for The Skies*. Mimeo.

[20] For example Ahmed Saeed Flugreisen (1989), Nouvelles Frontiers (1985), British Airways/ British Caledonian (1988), British Midland vs. Aer Lingus (1992)

Meyer, T. and Menzies, T. (1999). *"Airline Deregulation: Time to Complete the Job."* [Online.] *Issues in Science and Technology.* Available from <http:/nap.edu/issues/16.2/p_meyer.htm>, 14/01/2004.

Fingleton, J. (1998), *The Dublin Taxi Market: Re-regulate or Stay Queuing?*, The Policy Institute, Dublin.

Kahn, A. (1988). "Surprises of Airline Deregulation". *American Economic* Review. American Economic Association, Vol. 78.

Morrison and Winston, (1986). *The Economic Effects of Airline Deregulations*. Washington D.C.: The Brookings Insitute.

Pryke, R. (1987). *Competition Among International Airlines*, Ashgate Publishing, Limited.

Scamorana, (1993). *The Case for Competition*. Mimeo.

O'Leary, M. (2004). *Ryanair*. Trinity College Dublin, 21/01/04

Williams, G. (1994). *The Airline Industry and the Impact of Deregulation*. Avebury.

Sunday Business Post (2003). "Ryanair should not be victimised". *Sunday Business Post.* 28/09/03.

The Economist (2003). "Open skies and flights of fancy." *The Economist*. 4/10/2003

The Economist (2003). "A bet on the future, AirFrance-KLM." *The Economist*. 4/10/2003

The Irish Times (2003) "Commission orders recovery of €4.5m in Ryanair subsidies." *The Irish Times.*03/02/2004

The Irish Times (2003). "Ryanair condemns EU on impact of Charleroi ruling." *The Irish Times.* 04/02/2004

CRITICAL ASSESSMENT OF THE PROPOSED DUBLIN METRO

BY CLARE EGAN

Senior Sophister

The Transport Minister Seamus Brennan proposes the plan of a metro link between Dublin city centre and the airport. In light of the recent costs escalation and disappointing completion terms of the Luas project, investor confidence has been dampened. Clare Egan looks into the potential pitfalls of the CBA concerning the project, examines examples of similar rail systems in other countries and assesses risks and uncertainty associated with the project, concluding that Dublin Metro project should be considered with great caution and scepticism.

Introduction

In economic terms, rail transport should be able to justify itself; in reality this is rarely the case. With fixed track transport, analysis is centred on assessing the social need for loss making services. The Rail Procurement Agency has a budget of €9 million from the Department of Transport to cover metro-related research work. This paper attempts to examine the likely return on this investment to society. This is achieved through focusing on the following themes:

1) Background to Proposed Metro and Controversy Surrounding the Proposal.
2) Cost Benefit Analysis (CBA) and The Links Between Infrastructure Investment and Economic Growth
3) An Examination of Systems Implemented in Other Countries
4) Transport Mode Substitution: The multi-modal perspective and bus as an alternative
5) Risk and Uncertainty associated with the project
6) Background to Proposed Metro

Transport Minister Seamus Brennan is to bring to cabinet in the next two months his final proposals for a metro link between Dublin city centre and the airport. Dublin Airport is located in Fingal County approximately 10km north of Dublin city centre, between the M50 motorway and the town of swords, which lies

2km to the north of the airport. Over 56% of all passengers originate from the Dublin area, with 11% originating from surrounding counties and 33% from the rest of Ireland. According to an Aer Rianta passenger survey Dublin Airport handled 14.3 million passengers during 2001, 25% of these were flying for business purposes.

Mr Brennan is facing a tough political battle to convince his cabinet colleagues that the €2.4 billion plan for the first phase of a Dublin metro system is viable. The Department of Finance remain opposed to the project for a number of reasons; firstly the project is only feasible as the first stage in an overall scheme estimated at costing €20 billion. Secondly, the Department of Finance commissioned a report that suggests that a much cheaper alternative exists, the extension of the DART to the airport, via a new spur line.

The Rail Procurement Agency's (RPA) latest metro plan revised initial estimates of €4.8 billion down to €2.4 billion by reducing the number of stops and design specifications. Legislation to transfer underground property rights to the state is also being proposed in an effort to halve the time for the planning stage of the project from four years to two. The most contentious issue with the metro however, is the proposed route within the city centre. Most recently the RPA submitted a proposal whereby the metro stops in the city would be O'Connell St. and D'Olier St. Mr Rob Leech, Project Manager with the RPA, said this was due to "engineering restrictions on track angles and efforts to minimise costs" preventing integration with DART stations and the national rail network. Mr Brennan has rejected this plan insisting the metro provide a direct link to either Connolly or Tara Street Stations. The RPA's proposal for a 'travelator' – a moving walkway link running between D'Olier St and Tara St Rail Station has also been rejected. This amounts to a small victory for common sense as research has shown that each additional interchange imposed on customers carries with it a minimum 40% reduction in demand.

A business plan has been submitted by the minister to the cabinets transport subcommittee, based on the RPA's proposals for a public private partnership (PPP) metro contract. It involves repayments of €300-400 million over a 25 year leasing deal, to be paid when the metro goes into operation. The Department of Finance are concerned that ongoing commitments during a long period might compromise the ability of future governments to expand the network. The likely success of PPP in relation to the metro remains to be seen. In August 2002, Frank Allen, Head of the RPA himself admitted that doubts about the future of the proposed metro link to Dublin Airport was not inspiring confidence among prospective international investors. Finally in relation to the financing of the project fears over a likely breach of the EU Stability and Growth Pact must be taken into account. The RPA have commissioned a CBA of the metro and so the purpose of this paper is not to attempt

the same but to highlight some of the likely pitfalls associated with this particular CBA.

CBA and the Links Between Investment in Infrastructure and Economic Growth

> "Sustainable growth and development has the basic objective of maintaining growth in the national and international economies, but with the use of fewer resources, particularly non-renewable resources. This means that we would expect a continued growth in GDP but with fewer resources used in transport" (Banister and Berechman, 2000).

This does not automatically mean that there will be less transport, however it does mean that we have to become more efficient in our use of resources. In addition, the current technological revolution in communications and information may also weaken the links between transport expansion and economic growth. Substantial pressure is being placed on central government to finance major infrastructure projects on the understanding that this will lead to increased competitiveness and economic development. Banister and Berechman (2000) claim that based on historical evidence the case is far from clear, particularly in developed economies where additional investment has little impact on the overall accessibility within the transport system.

In the transportation field, the use of CBA has become part of the system of determining priorities for investment and the allocation of capital grants by Central Government. Stanley and Nash (1977) feel CBA as applied to the evaluation of transport improvements has been far from ideal. They claim the ready availability of market prices for many items has led to a failure to trace through the effects of schemes on the economy as a whole. They also note that items for which many values are not readily available have, with the exception of time been subjected to far too little scrutiny. "Evaluation has been used as a hurdle which a scheme prepared on other grounds must negotiate before implementation, rather than as an integral part of the planning process" (Stanley and Nash, 1977).

CBA is often preferred by those wishing to promote a specific option for a particular investment project, in which case the accuracy and impartiality of the analysis would require further checking. "Since it has become common practice to use CBA in the transport field, however, analyses are sometimes produced to support already committed political decisions without being used in the actual decision making process" (Hutton, 1978).

The Community Support Framework Evaluation (CSFE) Unit proposes a number of working rules for Cost Benefit Analysis. Their first recommendation is in relation to timing. CBA is a decision support tool and "It would be undertaken at an early stage in the project selection process before any commitment, provisional or

otherwise, has been entered into" (CSFE Unit, 2002). The CBA should also identify an alternative option for comparison with the project under appraisal, commonly used counterfactuals include 'do nothing' and 'do minimum' options. It is unlikely that the CBA commissioned by the RPA will seriously examine these counterfactual situations seriously. The possibility of the project displacing other economic activity should be specifically examined. The net benefit arising from the proposed project should be reduced in proportion to the estimated amount of the displaced activity.

Walsh and Williams (1969) recommend that the analyst should show the full effects of all aspects of the options on the group involved as well as presenting the Net Present Value (NPV). This could go someway towards combating the problems arising from the use of NPV "as a decision criterion in an economy where the tradition has been to use CBA for public expenditure advocacy rather than evaluation. NPV favours large projects" (Barrett, 2003). It is also noted here that the Strategic Rail Review's analysis of decision criterion recommends that caution is not necessary in interpreting NPV estimates. Is this appropriate in view of its bias towards larger projects with a weak tradition of public capital investment appraisal, especially in the transport sector?

Considering Ireland's proximity to full employment, we must bear in mind Keynes' words regarding diverting resources out of full employment. "Assuming a full employment economy, use of the resources for the project under consideration will prevent them being used on some alternative project and one must allow for the extent to which private investment will be displaced" (Hensher, 1978). Mishan (1988) points out that many externalities cannot be incorporated within the pricing mechanism, due to the nature of things. The existence of these externalities brings a divergence between social and private costs and benefits. Debate surrounding externalities is prevalent with regard to the environment. We must recognise that although electricity may appear to be a 'clean' fuel, electricity generating stations such as that at Moneypoint, Co. Clare, often burn fossil fuel turf and create pollution, so the use of electricity only puts pollution further back in the production chain.

The impact of transport investment on the value of land is a very complicated issue. Large infrastructure developments, such as a metro often have a substantial influence on the value of land in surrounding areas; this increase can in itself amount to a free gift to landowners. In the city of Montreal the council financed the construction of a transit line by borrowing the money against the expected increase in income from rates (land taxes), around the stations that would follow from its opening. Depending on the route eventually chosen it is questionable as to whether a similar policy would work in Dublin. With this in mind the next section of the paper attempts to examine whether we can learn from the experiences of those around us.

An Examination of Systems Implemented in Other Countries

Throughout my research, I have encountered examples from a large number of countries that indicate a certain level of caution is warranted, in relation to the proposed Dublin Metro. A close example is that of London Heathrow. The British Airport Authority undertook this project and it has proven to be far from successful. The high fare that was set succeeded in attracting only 8% of passengers and virtually no staff. The percentage of passengers travelling by car has actually risen since completion of this metro. Only three of the nineteen major European airports examined in a transport Economic study at the University of Rotterdam have a metro system in place: London Heathrow, Paris Charles de Gaulle and Madrid Barajas Airports.

Studies presented by Alan Abouchar of the University of Toronto show that it is unlikely that the subway there had a different impact on Toronto than on other major Canadian cities. "Urban projects and property value change study shows that subway had no effect on property values in subway areas, compared with non-subway areas" (Abouchar, 1977). In light of this one should question the rejuvenation of Ballymun being put forward as a likely benefit of the metro by those in favour of the project at the RPA. If Toronto is any gauge then we cannot in general expect that a metro causes significant benefits to accrue to persons other than the users (apart from pollution control to an extent and even that is not one hundred percent verifiable).

In Norway, the Alesund tunnels connect the town with surrounding islands including the airport. These tunnels are privately financed; the toll company is bankrupt and has been taken over by the banks with debts of NKr 900 million.

The main reason for this situation was the over estimation of traffic by 20-30% as the toll charges were set too high (Bannister and Berechman, 2000). The central problem here is one that could well be mirrored in Dublin; the market is not large enough to sustain the fare levels and to secure the necessary revenue to pay for costs of administration and collection, interest and loans.

Certain theorists would argue that the impact of investment is important in establishing the image of an area and hence its attractiveness to new development. This in turn can have a positive impact on the labour market. Transport investment may act as the trigger mechanism to this process. The alternative explanation seems to lead to the conclusion that only existing locations will be attractive as they have first mover advantages and will always be more accessible than new or peripheral locations. The Merseyside situation in the UK is informative here. It was expected that substantial programmes invested in during the 1960s and 1970s would lead to increases in population and employment, together with rising productivity and income. Bannister's study found that inadequate road networks were not a key

component of that restructuring process and investment was required in retraining, new industries and a regeneration of the local economy. Roads were originally justified on the basis of the expected growth in traffic and the necessity to accommodate and direct this growth. Subsequently, the same roads were being defended as a means to regenerate the local economy (Bannister, 1994).

A final example is that of the London-Oxford passenger route served by three main forms of transport: rail, long distance express coach and car using the M40. The introduction of lower fares on express coaches resulted in passengers transferring from rail travel to coach travel, particularly where journey times were similar (Cole, 1987). It is this topic: the ease of transport mode substitution that the next section examines.

Transport Mode Substitution: The multimodal perspective and bus as an alternative

Dublin Airport is accessible by over 700 buses daily. You can reach the airport by bus from many areas outside of Dublin. Eight separate bus operators service the airport, including Aircoach, a highly successful private venture, without subsidy or capital grant, carrying over one million passengers. The Government continues to finance the railway deficit without any measure of corresponding social benefits. No other operator is eligible for these subsidies. "There is strong support for railways to the detriment of an independent bus sector" (Barrett, 2003).

If the Government allowed contestable markets to exist in the transport sector then both operators and service users can benefit. This is supported by the deregulation of the Dublin-Galway bus route, the airport coach service and internal air services. The introduction of additional Quality Bus Corridors (QBC) can also have a positive impact. These cost €400,000 per km., compared with €28 million per km of Luas. Already numbers are up 54% across the QBC network in Dublin.

In a 2001 Aer Rianta passenger survey, 22.4% of passengers used bus with 43.5% using private car (See Table 2). The likely target market for the metro would exist predominately within the sector already using bus. Based on the relative costs of Irish railways, airlines and buses in a contestable market, railways lose market share due to the lower costs of independent bus companies. In the current economic climate surely the Irish Government should be supporting and encouraging competition, not frustrating it.

According to the conclusions of the round table conference of Ministers for Transport on airports as multimodal interchange nodes, the future of airports depends on technical opportunities to cope with congestion. "This concerns the airports themselves as well as the infrastructure of the interconnected modes, in

particular road infrastructure." (OECD, 2003) It is also noted here that larger airports may justify high investments in rail connections, but these links should be integrated with the regional and national railway system. Fixed track systems are limited to bringing you from A to B. If Dublin Airport is to succeed as a multimodal interchange node then market alternatives need to be considered and promoted. It is worthwhile noting here, that the Strategic Rail Review gave no serious consideration to public transport alternatives to the large railway investments proposed. In addition, the OECD suggested the following as essential for the development of an effective tourism policy: making better use of existing infrastructure rather than expanding it when demand rises and promotion of multimodal services.

Risk and Uncertainty associated with the proposal

According to Bannister and Berechman (2000) risk is used to indicate the likelihood of selecting the wrong project or a project that is economically non-viable. Sometimes the term is used to indicate the effect of a given project on the welfare of taxpayers, relative to the distribution of the project's indirect costs (negative externalities). The term uncertainty is sometimes used to indicate the degree of inaccuracy associated with the forecast of the project's future benefits and costs. Since there is always an intrinsic level of uncertainty regarding the future state of the economy, a project that is largely irreversible or that cannot be stopped without rendering its costs sunk, should be ranked inferior to one with the same NPV but which is flexible to a reasonable degree.

The Dublin Port Tunnel is a prime example of risk and uncertainty. Project costs soared from €165million to €670million, although the tunnel is still too small to take high vehicles (Tobin, 2004). The inefficiency of The Irish Construction Industry results in bills being doubled. Before additional track systems are built the success of the Luas should be assessed. At that stage contracts tendered are likely to be lower as a result of the addition of Eastern European countries to make up the EU 25. Skamris and Flyvbjerg (1997) have concluded that in the case of Danish bridge and tunnel projects, on average, construction costs were consistently 50% to 100% undervalued, whereas traffic forecasts were about 60% overestimated.

The conclusion seems to underlie an established maxim in transportation CBA, which says that in order to arrive at the correct benefit and cost values of a transport infrastructure project one should halve the project's predicted benefits and double its estimated costs (Bannister and Berechman, 2000).

Data on completion times of projects such as the Dublin Metro are very discouraging. "The fact that many tunnels being built 50-100 years ago had similar

total construction times as today's tunnels brings into focus the inefficiencies" (Einstein, 1987).

Conclusion

In summary, the concept of opportunity cost presents us with the question 'would we be better off without it?' One could argue that the proposal is justified on the basis of timesavings, crowding relief and total public transport user benefits. In spite of this, that is not the conclusion reached here. Transport investment does not cause growth but allows it to take place. We must consider the small element of production that railways contribute, at such high cost, in terms of economics rather than emotions. One has to consider with great scepticism projects prepared by bodies that are directly linked to the project and stand to gain from its implementation. Stricter control and regulation of Cost Benefit Analysis by the Department of Finance could go someway towards counteracting the problem. A secure Independent wider evaluation framework is also required. The proposal may seem affordable and therefore worthwhile now but the question remains: Will future generations funding the project agree?

Bibliography

Abouchar, A. (1977), *Transportation Economics and Public Policy with Urban Extensions.* Canada: University of Toronto.

Banister, D. & Berechman, J. (2000), *Transport Investment and Economic Devlopment.* London: UCL Press.

Banister, D. (1994), *Transport Planning.* London: Spon.

Barret, S. (2003), "The Irish Strategic Rail Review: A Critique" In: *Quarterly Economic Commentary,* Autumn.

Cole, S. (1987), *Applied Transport Economics*. London: Kogan Page.

CSFE Unit (2002), "Cost-Benefit Analysis in the Community Support Framework: A Critical Review" In: Mulreany, M. (ed.) *Cost Benefit Analysis Readings*. Dublin: Institute of Public Administration.

Einstein, H. (1987), "Influencing the Cost" In: Davidson, F. (ed.) *Tunneling and Underground transport developments in technology, Economics and Policy.* New York: Elsevier.

Hensher, D. (1978), Urban *Transport Economics.* Cambridge: Cambridge University Press.

Hutton, J. (1978), *Cost Benefit Analysis and Transport Planning*. Glasgow: Glasgow College of Technology.

Mishan, E.J. (1988), *Cost Benefit Analysis.* 4th ed. London: Allen & Unwin.

OECD (2003), "Airports as Multimodal Interchange Nodes" *Conclusions of Round Table* 126, European Conference of Ministers for Transport. OECD: Paris.

Skamris, M. & Flyvbjerg, B. (1997), "Accuracy of traffic forecast and cost estimates on large transportation projects" *Transportation Research Record*. Available from: <www.publicpurpose.com/pp-infra.htm>

Stanley, J.K. & Nash, C.A. (1977), "The evaluation of urban transport improvements" In: Hensher, D. (ed.) *Urban Transport Economics*. Cambridge: Cambridge University Press.

Tobin, L. (2004), *Integrated transport plan is way forward.* Dublin. Mimeo.

Walsh, H.G. & Williams, A. (1969), "Current Issues in Cost Benefit Analysis" *Civil Service College Occasional Papers*. Dublin: Civil Service College.