Financing Ireland's Transport Infrastructure During a Fiscal Crisis: the Viability of the Public-Private Partnership (PPP) Mechanism

Síofra Moriarty
Senior Sophister

In the face of fiscal tightening by the Irish government, Síofra Moriarty undertakes a thorough, topical analysis of the Public-Private Partnership mechanism, as it applies to the Irish transport sector. The paper concludes that the PPP option may not be so attractive when one assesses it through transaction cost economics and game theory.

Introduction

It has been over a decade since the first systematic programme of Public-Private Partnerships (PPPs) was introduced in Ireland, and until the financial crisis penetrated the Irish economy, Ireland pursued an extensive PPP strategy, ranking third in the world in terms of the maturity of its PPP markets in 2007 (Deloitte, 2007). The PPP mechanism has been particularly effective in the Irish transport sector, with all of the road schemes within the National Roads Authority's first PPP roads programme being delivered ahead of schedule and on budget (KMPG, 2011). Today, it is argued that because of Ireland's current fiscal pressures and the urgent need to regain competitiveness, there has never been a better time to utilise PPPs for the development of Irish infrastructure (KPMG, 2011).

However, since the advent of the financial crisis three years ago, no major PPP project has secured funding in Ireland (Irish Times, 2011). This trend seems likely to continue, reflected in Minister for Transport, Leo Varadkar's recent (2011) announcement of the "deferral" of both the Metro North and DART Underground projects. The fact that the official database used by the government to monitor PPP projects, ppp.gov.ie, has not been

updated since March 2010 further illustrates the current inactive state of the PPP mechanism in Ireland.

Up until now, little or no objective economic research has been in conducted in an Irish context in terms of gauging the viability of pursuing PPPs in the transport sector (Connolly, 2009). Using a combination of economic theory and global empirical evidence, this paper seeks to redress this void. Section I begins with a brief description of the PPP concept. Section II applies economic theory to examine the proposed advantages of PPPs. In Section III, consideration is given to existing empirical evidence. The issue of the transaction costs that arise under PPP projects is considered in Section IV and a model for mitigating the risks that result from the strategic behaviour of the PPP parties is developed in Section V.

PPP Concept- The Basics

According to the Department of Finance (2003: 4):

> "A PPP is a contractual arrangement between the public and private sectors whereby the delivery of public infrastructure is carried out by the private sector as opposed to being provided through traditional public-sector procurement."

There may exist different categories of PPP projects, depending on the degree of private and public sector ownership and commitments related to the projects. The model that has been adopted in most Irish cases, however, is the Built-Operate-Transfer (BOT) agreement, whereby the private sector builds and operates an infrastructural transport project and following completion, transfers ownership to the State (Reeves, 2011).

Proposed Objectives of PPP Projects

The Department of Finance (2001) asserts that the PPP mechanism delivers value for money by establishing: (i) contestability in the bidding market, (ii) bundling of project elements (iii) relief of budgetary pressures and (iv) risk transfer. I will now apply economic theory to explain such reasoning in a wider context.

Contestability and Efficiency

Much of the salutary effect of PPP schemes derives from the belief that the State's pursuit of multiple objectives (e.g. economic, financial, social and environmental) when delivering transport projects is likely to impede the levels of

efficiency achieved (Engel, 2009). On the other hand, the tendering process inherent to the PPP mechanism introduces an element of competition, ex-ante, in the form of a contestable bidding market for PPP projects (Demsetz, 1968). This, coupled with the contractor's one objective of making profits, ensures that the selected bidder proposal will be the one that is (i) the most productively efficient, ensuring outputs will be produced at the lowest possible cost (Demsetz, 1968) and (ii) the most dynamically efficient, ensuring that optimal investments in process improvements are made so that firms can more efficiently address consumers' needs in the future (Hodge et al, 2011).

Bundling
PPP projects typically encompass a wide range of activities - design, construction, operation and maintenance. The PPP mechanism bundles the activities of the construction and operational phases, so that the same agent undertakes them, thereby engaging in life-cycle costing (Engel, 2009).

Such bundling has both quality enhancing and cost minimisation implications for the infrastructural project. For instance, high quality infrastructure generally gives rise to reduced operation and maintenance costs[1]. This positive externality induces the contractor to internalise costs, to prioritise quality concerns and to choose the efficient level of quality that minimises all operational, maintenance and construction costs. The private party may also be enticed to introduce innovation in service delivery, further enhancing quality (OECD, 2010). Bundling also makes the firm maintaining a transport project more accountable to users than it would be with the traditional approach, where separate activities are carried out by various different agents (Engel, et al. 2011). In addition, if activities are held together, the prevailing internalisation of life-cycle costs enhances the opportunity to exploit economies of scale. The lower unit costs that then ensue shift the private firm's average cost curve downwards, resulting in lower overall project costs (OECD, 2010).

Relief of Government Budgets
The Department of Finance (2001) also postulates that the PPP mechanism relieves strained budgets by allowing the Exchequer to spread the cost of transport infrastructure over a long time-frame, as opposed to being subject to large upfront payments on project delivery. This frees up government resources, which can then be spent on other projects with high social returns (Engel, et al. 2011). Also, when both construction risk and demand/avail-

1 Winston (1991) reports that small increases in road surface thickness can dramatically lengthen the life of a road and reduce maintenance costs.

ability risk are transferred to the private partner, the European Commission permits that related expenditures do not count as part of government borrowing[2]. In this way, the respective year's budget deficit is reduced (Eurostat, 2004). Such arguments are gaining particular momentum in the current economic climate, due to the Irish government's ever intensifying borrowing requirements, coupled with the Exchequer's commitment to keeping within the fiscal constraints imposed by the European Union (KPMG, 2011).

Risk Transfer

It is also argued that PPPs achieve optimal risk management in that they provide an opportunity to exploit the parties' relevant competencies and optimally apportion the project risk accordingly. They therefore supposedly achieve increased value for public services and result in more informed, efficient investment decisions (OECD, 2010). In the literature[3], there is a general consensus that because of superior project-management expertise, private firms are better equipped than the public sector to manage construction, timing, market and demand risk. On the other hand, the public sector is deemed to be more proficient at managing systematic risks, which are non-specific and result in broad economic conditions (Alexandersson & Hultén, 2007).

Empirical Evidence

The consideration of the available empirical evidence, however, raises questions as to whether the proposed advantages provide a valid basis for pursuing a PPP strategy in the Irish transport sector.

Contestability and Efficiency

For example, a prerequisite for reaping the potential benefits from Demsetz's (1968) auctions of PPPs is that there is real competition for the contract. Barriers to entry and outright collusion often indicate the contrary (Engel, 2009). For instance, in 2005, the three main public transport operators in France were condemned by the French Competition Commission to pay 12 million euro for collusive strategies during the PPP bidding process (Amaral, 2008). In Ireland, given the large size and nature of PPP road projects, coupled with the fact that competition for PPP contracts is limited to a small number of bidders, the risk of collusion and barriers to entry dilutes the competition argument put forward by the Department.

---

2 The NRA's PPP Roads Programme did not count as government borrowing, resulting in a €2.1billion reduction in the government deficit (KPMG, 2011).

3 See (Engel, et al. 2011), (Engel, 2009), (Alexandersson & Hultén, 2007), (OECD, 2010).

Bundling

The bundling argument also seems to be lacking. The European Investment Bank, for instance, in their (2005:8) study on large-scale European PPP projects, found that an emphasis on cost-minimisation resulted in, a "lower quality of public service" and "reduced safety of service". A report, made by the French Court of Auditors following the 2005 Roissy Airport Terminal crash, found that Aeroport de Paris was "wearing too many hats", acting as promoter, builder and owner of airport assets. Such bundling was deemed to be the primary cause for the suboptimal quality of airport infrastructure (Amaral, M, 2008). In addition, the London train collision of October 1999 was cited as being a direct result of the maintenance and expenditure of Railtrack (the private party) being much lower than agreed (EIB, 2005).

Relief of Government Budgets

The budgetary argument is also diluted when one considers the inter-temporal nature of the government budget, where initial savings of government under a PPP are found to be equal, in present value, to the amount it surrenders in tolls that could have been collected under the traditional approach (Engel, 2009). Also, spreading the costs of the transport infrastructural investment over time merely converts a present budget deficit into future budget deficits.

In addition, we have to be aware that realistically, capital is not free. From the perspective of financial markets, there is no safer borrower than the State - their monopolistic powers of taxation enable them to secure the best interest rate available, implying that overall borrowing costs are less when funds are acquired by a public agent (Clements and O'Mahony, 2005).

Finally, it has been found that PPPs have been initiated as a means of evading expenditure controls and hiding budget deficits. In Hungary, for example, PPPs for motorways were wrongly recorded off budget in 2005 and 2006; the reconciliation of these costs boosted the country's deficit by almost one percent of GDP that year (OECD, 2010).

We can thus conclude that, financially speaking, PPPs are simply a case of hire-purchase and in the long run: they do not provide the budgetary gains that the theoretical arguments propose. In the UK, Chantry Vellacott (2005), estimates that a typical PPP contract has an inherent cost of some 5% per year higher than if the Treasury borrowed the money directly.

Risk Transfer

Much of the empirical evidence relating to the risk transfer argument has also shown that instead of efficiently transferring risk from the private to public

sector, most PPP contracts include different forms of insurance against construction, maintenance and demand risk (Engel, 2009) For instance, in 2010, the National Roads Authority was obliged to pay the operators of the M3 motorway and the N18 Limerick tunnel €1.79 million as traffic volumes failed to meet "traffic-related guarantees" (Irish Times, 2011). The lack of data on the risk transfer argument raises further questions about its credibility. A study by Pollock (2005) found that of 622 PPP deals signed in the UK by October 2007, the National Audit Office had examined the relationship of risk transfer in only three; two of these studies were uninterpretable because of small sample size and selection bias.

Transaction Cost Economics

Following a consideration of the above complications illustrated by empirical evidence, one can view the problems posed by the PPP design as really being a special case of the famous principal-agent problem, whereby the principal wishes to procure services for the public, by using the services of private agents whose objective is to maximise profit. PPPs present an asymmetry of information in that both private and public partners are more informed about their respective areas of specialisation (e.g. government policy, construction sector practices). In addition, PPPs include transaction costs, which arise from organising a competitive tendering process as well as from writing, monitoring and enforcing contracts (EIB, 2011).

The shortcoming of the Department's case for the pursuit of PPP strategies can thus be seen to lie in the non-consideration of such agency problems, informational asymmetries and transaction costs. In theoretical terms, by merely evaluating the above four criteria, their assessment is broadly in accordance with neoclassical economics, which considers only those markets with perfect information and focuses only on production costs (Reeves, 2008). A wider perspective on economic efficiency, embracing transaction cost economics[4] (including the consideration of renegotiation, opportunism and adverse selection) is thus an important contribution to understanding the issues embodied in the PPP mechanism (Coase, 1937).

For instance, the recessionary environment has meant that project costs, market demand and other market conditions relating to PPPs have become significantly unfavourable. This volatility has triggered cases of financial renegotiation, whereby after PPP commencement, the private party is forced to renegotiate with the government for subsidies, in order to ensure project continuation and completion. The government is often tempted to accept the

4 Transaction cost economics are the economics of the costs incurred in making an economic exchange (Williamson, 1996).

renegotiation because of the gigantic transaction costs associated with re-tendering and because of its interest in guaranteeing service provision (Ping, 2005). In a situation exacerbated by the economic crisis, demand over-esti-mates of about 50,000 users a day for the Kilcullen-Waterford motorway have resulted in €1.5 billion of excess costs and over-investment by PPPs (Barrett, 2006). The Irish taxpayer is currently paying the price for such inaccuracy.

The very expectation that the Irish government will renegotiate if relevant market conditions deteriorate may cause opportunistic bidding on behalf of prospective private partners. Developers, anticipating government rescue if adverse risks prevail, may intentionally understate the possible risks involved in their PPP proposals in order to outperform other bidders and se-cure the PPP contract. This may result in cases of adverse selection, whereby "bad" parties are selected to engage in a PPP and the "good" ones are driven out of the contestable bidding market (Ping, 2005). Engel (2009) for example, shows that a 'renegotiation frontier' emerges, which trades off renegotiation ability with technical prowess, attracting firms that are more skilled at lobby-ing rather than those who are technically efficient.

Game Theory and the Renegotiation Problem

I propose, however, that the use of game theory could offer researchers a framework and a methodology to understand and analyse the behavioural dynamics of the parties in a PPP. The study of such analysis could aid govern-ment in formulating effective management policy to solve the renegotiation problem.

Consider the basic dynamic game illustrated below. Following the prevalence of adverse market conditions, strategic behaviour carried out by both parties of the transport project under a PPP arrangement results in an identification of 3 Nash equiliria:

- The developer chooses "project bankruptcy", a no rescue equilibrium. The government will restructure the project and will face a political cost of restructuring, $-b(G + \pi)$, where G is the least required government funds for restructuring a project and π is the opportunity cost of replac-ing developers including the costs of retendering and interruption.
- The developer will "request a subsidy", such as a debt guarantee, and the government will "reject", a no rescue equilibrium, whereby the payoff will be the same as for bankruptcy, $(0, -b(G + \pi))$.
- The developer will "request a subsidy", and the government will "negoti-ate a subsidy", by offering gU, where g is a "rescuing subsidy"- a ratio between 0 and 1 – representing bargaining. This is a rescue equilibri-

um. The payoff to the developer here is gU. The payoff to government is [-b(gU) – r(gU)], which relates to the political cost of restructuring, where b is the function of political cost of budget overspending and r is the political cost of over-subsidisation (including a loss of public trust and resource misallocation) (Ping, 2006).

Negotiate Subsidy: $\in gU$ $[gU, -b(gU) - r(gU)]$

Irish Govt.

Request Subsidy: $\in U$

Private Partner

Reject $[0, -b(G + \pi)]$
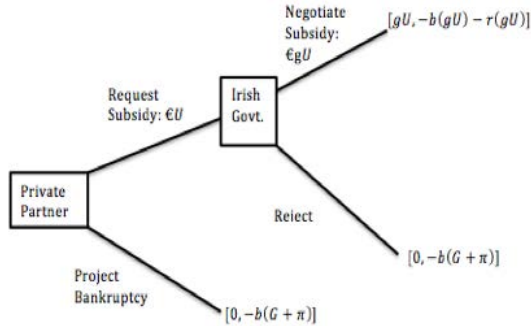
Project Bankruptcy $[0, -b(G + \pi)]$

Figure 1: Game Theory Application of the Renegotiation Game
Source: Adaptation of Ping's (2005) Renegotiation Game's Equilibrium Path

What policy implications can be obtained from the above game theory analysis? As illustrated in Figure 2, when G is less than or equal to S (the intersection of the curves b(gU)+r(gU) and b(G + π)), the "rescue equilibrium" will be obtained. Hence, the most important policy implication here is that the Department's policies relating to PPPs should try to reduce the magnitude of S so as to decrease the possibility of opportunism from developers.

    How can this be achieved? As shown in Figure 3, policy makers could firstly attempt to increase the political cost of over-subsidisation, r, as when this function becomes steeper, the magnitude of S will be reduced significantly. For instance, laws may regulate the renegotiation and negotiated subsidy, and such laws will increase the political cost when the subsidy offered is not considered to be justifiable. Similarly, the probability of reaching the "rescue equilibrium" could be reduced by introducing strategies that reduce π, the cost of replacing the developer. This could be achieved for instance, through the introduction of a good monitoring or 'early warning' system, administered by third party experts, that would give government more lead time to replace the developer with minimal impact.
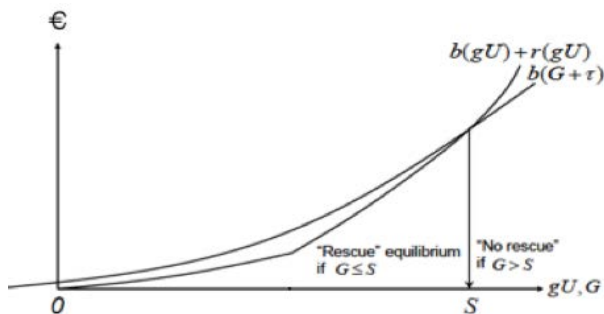
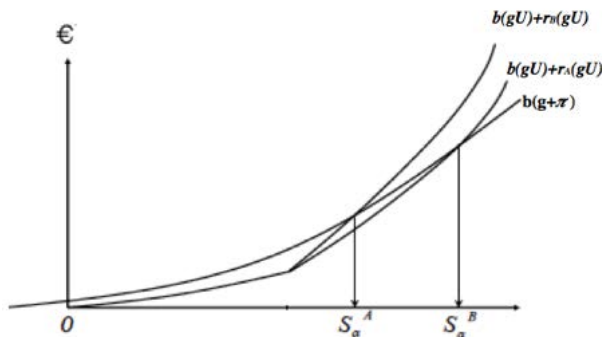Figure 2: Conditions for "rescue" equilibrium and "no rescue" equilibrium

Figure 3: Impacts of the change of r on the equilibriums

Conclusion

This paper goes some way towards addressing the lack of existing published economic analysis on the PPP mechanism in the Irish transport sector. Using economic theory, it has provided possible economic explanations as to why the Irish government has used the mechanism extensively in the transport sector. A consideration of the existing empirical evidence surrounding this issue however, suggests that that the Department of Finance's current case for the support of PPPs is suboptimal. The paper argues that the problems posed by the PPP design are really due to a non-consideration of transaction cost economics. When one considers this essential perspective, it becomes clear that further issues relating to renegotiation, opportunism and adverse selection are embodied in the PPP mechanism and need to be addressed and corrected before PPP strategies should be put into place. I propose that game

theory provides a useful framework for understanding such issues and have referenced a simple model, to describe the dynamic strategic behaviour of both parties, which may aid government in formulating effective management policy to solve the renegotiation problem.

References

Alexandersson, G. and Hultén, S., 2007. *Prospects and Pitfalls of Public Private Partnerships In the Transportation Sector- Theoretical Issues and Empirical Evidence*. Stockholm. Stockholm School of Economics.

Amaral, M., 2008. Auction procedures and competition in public services: The case of urban public transport in France and London. Utilities policy. 17, pp. 166-175. Paris. Elsevier.

Barrett, S., 2006. Evaluating Transport 21 - Some Economic Aspects. *Quarterly Economic Commentary*, Winter, 36-58.

Birch, M.H. and Haar, J., 2000. *The Impact of Privatization in the Americas*. Coral Gables: North-South Centre Press.

Chantrey Vellacott, 1999. DFK Economics report. London: Chantrey Vellacott Publishing.

Clements, C. and O'Mahony, M., 2005. Traditional or Public Private Partnerships for Procurement of Light Rail Transit? Dublin. Transportation Research Board of the National Academies.

Coase, R., 1937. The Nature of the Firm. *Economica* 4: 386-405.

Deloitte, 2007. Closing the Infrastructure Gap- the Role of Public Private Partnerships. London. Deloitte Research.

Demsetz, H., 1968. Why Regulate Utilities?. *Journal of Law and Economics,* 11(1), pp. 55-65

Department of Finance, 2001. Framework for Public Private Partnerships. Dublin: Stationary Office.

Department of Finance, 2010. What is a PPP? (online) Available at: http://ppp.gov.ie/glossaryfaqs/faqs/ [Accessed 1 December 2011]

Department of the Environment and Local Government, 2000. Introduction to Public Private Partnerships. Dublin: Stationary Office

Engel, E., 2009. On the efficient provision of roads. Mimeo. Working Paper

Engel, E., Fischer, R. and Galetovic, A., 2011. *Public-Private Partnerships to Revamp US Infrastructure.* New York. Yale University Press.

European Competitiveness Report, 2010. (online) Available at: http://easu.jrc. ec.europa/eas/downloads/dpf/JRC59169 [Accessed 2 December 2011]

European Investment Bank (EIB), 2005. Economic Pros And Cons Of PPPs: Concepts And Experience. Luxembourg. EIB Publications.

European Investment Bank (EIB), 2011. Productivity and Growth in Europe: Long term trends, current challenges and the role of economic dynamism. *EIB Papers* Vol. 16(1) pp. 66-94

Eurostat, 2004. New Decision of Eurostat on Deficit and Debt Treatment of Public Private Partnerships. Eurostat News Release, Luxembourg.

Ewers, H. J. and Tegner, H., 2000. Entwicklungschancen der privaten Real-isierung von Verkehrsinfrastruktur in Deutschland. Eine okonomische Anal-yse des Fernstrassenbauprivatfinanzierungsgesetzes, Berlin. Forschngsbericht

Irish Times, 2011. Can we fix it? PPPs can. 28 October (online). Available at at:http://www.irishtimes.com/newspaper/innova-tion/2011/1028/1224306356859.html [Accessed 24 November 2011]

Hodge, G., Greve, C. and Boardman, A., 2011. *International Handbook on Public-Private Partnerships.* Edward Elgar. Northampton, USA.

KPMG, 2011. What next for Infrastructure? – Infrastructure Insights for Ire-land. Dublin. KPMG Research

Ping, S., 2006. Model for Financial Renegotiation in Public-Private Partner-ship Projects and Its Policy Implications: Game Theoretic View. *Journal of Construction Engineering and Management,* 132 (7), July 2006, pp. 678-688

Pollock, A., 2005. NHS Plc: *The Privatisation of Our Health Care.* London. London University Press.

Ramusen, E., 2001. *Games and Information*. Blackwell Publisher Inc., Malden, Massachusetts.

Reeves, E., 2008. The Practice of Contracting in Public Private Partnerships: Transaction Costs and Relational Contracting in the Irish Schools Sector, *Public Administration*, Vol. 86, No. 3, pp. 969-986.

Reeves, E., 2011. The Only Game in Town: Public Private Partnerships in the Irish Water Services Sector, *The Economic and Social Review*, Vol. 42, No. 1, pp. 95-111.

Stacey, S., 1997. New capacities for old? Prospects for public private partnerships in service deliery in South Africa, Angola and Mozambique. Research report No. 61 commissioned by the International Development Research Centre of Canada.

Sutherland, D., 2010. Public Private Partnerships and Investment in Infrastructure, OECD Economics Department Working Papers, No. 803, OECD Publishing.

The Irish Examiner, 2011. Govt. to confirm postponement of major capital projects. November 10 (online). Available at: http://www.irishexaminer.com/breakingnews/ireland/govt-to-confirm-postponement-of-major-capital-projects-527757.html [Accessed 11 November 2011]

Williamson, O. E., 1981. The Economics of Organization: The Transaction Cost Approach, *The American Journal of Sociology*, 87(3), pp. 548-577.

Winston, V., 1991. "Efficient Transportation Infrastructure Policy", *Journal of Economic Perspectives*, 5(1), pp. 113-127.

Are the Foreign-Currency Official Reserves of Emerging Asia Excessive?

Irina Pechotski
Erasmus Student

In this paper, Irina Pechotski examines why countries hold foreign reserves and attempts to find a measure for the optimal level of reserves to hold. She applies this to the case of emerging Asian economies, who have rapidly increased their reserves over the past decade, concluding that the level is likely excessive, which may have helped insulate emerging Asia somewhat from the financial crisis.

Introduction

"There is no such thing as too much foreign reserves."
(Rhee Yeung Kyun, director-general of the International Department at South Korea's Central Bank)

"It is one thing to save for a rainy day, but one trillion dollars in reserve accumulation is more like building Noah's Ark."
(Ken Rogoff, former chief economist at IMF)

A development in the world economy which has attracted much attention is the rapid accumulation of foreign reserves in emerging Asia over the last decade. As the initial quotations may already indicate, there are differing attitudes towards holding reserves. It seems that emerging Asia cannot have enough foreign reserves, whereas the rest of the world is concerned and does not want to hold more reserves than necessary. This paper aims to explore the costs and benefits of foreign reserves and tries to analyse whether the level of emerging Asia's foreign reserves is excessive or not.

What are foreign-currency official reserves?

Foreign-currency official reserves are held by almost every country in the world and comprise "convertible foreign exchange, gold, special drawing rights (SDRs), and reserve position at the International Monetary Fund

(IMF)" (Bahmani-Oskooee and Brown, 2002).

They are an important national asset and a tool of the monetary and exchange rate policy. An important characteristic of foreign reserves is that they can be made readily available in order to fulfill the countries' demands. Countries have a need for holding reserves, although they may have different reasons for holding them. Reserves can be a formal backing for the domestic currency in order to provide confidence in the domestic currency. They are also used as a tool of exchange rate and monetary policy, especially in countries with a fixed exchange rate regime. These countries can maintain their fixed exchange rate by buying or selling the domestic currency in order to support the national currency. (Nugée, 2009; IMF, 2004)

Another two reasons often put forward are the mercantilist argument and the precautionary motive for holding reserves. The mercantilist view is that reserves are held for pursuing an export-based strategy. The basic idea behind this concept is the systematic undervaluation of exchange rates in order to promote own exports (Wyplosz, 2007). The precautionary motive regards reserves as a buffer to protect the economy against expensive output contractions, which may be triggered by sudden stops and capital flight (Aizenman and Lee, 2007), therefore reducing the likelihood and depth of a crisis (Edison, 2003).

But as reserves are not only advantageous; countries also incur costs by holding reserves, such as inflation, sterilization costs, opportunity costs, and central bank balance sheet losses (Green and Torgerson, 2007). Buying foreign currency leads to a rise in the monetary base. Therefore, the monetary authorities sterilize the reserve accumulation in order to offset the inflationary impact. By issuing bonds in exchange for currency in circulation, they reduce domestic liquidity (Park and Estrada, 2009). Hence, sterilization costs can be considered the "difference between what the central bank earns on international reserves and what it pays on domestic debt issued to sterilize the reserves" (Green and Torgerson, 2007). Opportunity cost means that the resources invested in foreign reserves could be used for alternative investments with higher returns because foreign reserves are usually held in "high-grade fairly liquid assets with low returns" (Wyplosz, 2007). Finally, the monetary authority has to face costs if a depreciation of the foreign reserves' value occurs. When the exchange rate appreciates and the foreign reserves are measured in domestic currency, the value of the foreign reserves falls (Green and Torgerson, 2007).

However, holding large amounts of foreign reserves does not only bear costs for the country itself, it influences the global economy too. Accumulating a high stockpile of foreign reserves leads to global imbalances that

can cause a global financial crisis such as the current one (Allen and Hong, 2011).

What is Emerging Asia and What Happened There?

In the last decade the world's foreign reserves have increased by an annually compounded rate of about 18% reaching US$10 trillion in 2011 (Figure 1). But an even faster accumulation of foreign reserves can be observed in emerging markets, with an annual growth rate of over 20%, namely from less than US$0.75 trillion to more than US$6 trillion between 2000 and 2010 (Independent Evaluation Office, 2011).

In particular, many emerging Asian countries[1] attracted attention. After the Asian crisis in 1997/1998 a high accumulation of foreign reserves was registered there (Figure 1).

■ Foreign Exchange Reserves of emerging Asia (billion of US$)

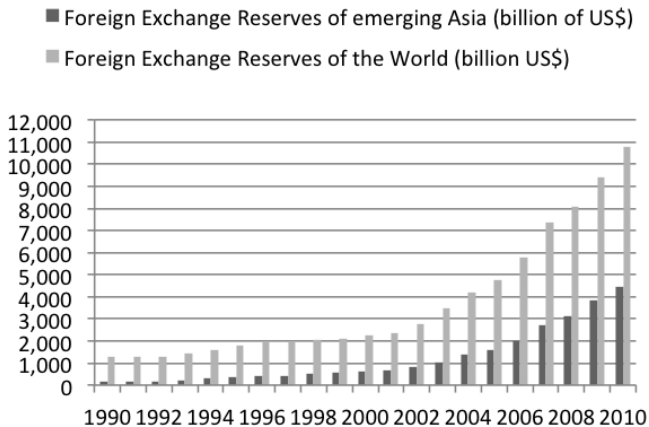■ Foreign Exchange Reserves of the World (billion US$)

Figure 1: Foreign Exchange Reserves including gold of Emerging Asia and the World, 1990-2010 (billion US$)

(World Bank Data)

China surpassed Japan as the largest reserve holder in the world in 2006 and among the twenty largest foreign reserves holders in the world in 2011, nine are emerging Asian countries: China, Taiwan, India, Korea, Hong Kong, Singapore, Thailand, Malaysia, and Indonesia.

1 Emerging Asia: China, Hong Kong SAR, India, Indonesia, Korea, Malaysia, the Philippines, Singapore, Taiwan Province of China, Thailand, and Vietnam as defined in the Regional Economic Outlook, Asia and Pacific (2010)

The Asian crisis hit large parts of Asia very hard. Before the crisis, the term "Asian miracle" was coined for this region of the world, where agriculture-based countries succeeded in turning into rapidly growing industrialised nations with "growth rates several times higher than advanced nations" (Krugman, 1994). But the crisis led to severe reductions in output, currency depreciations, business bankruptcies, political upheavals, and poverty. The IMF intervened with programs of economic stabilization and reforms in order to restore confidence in these countries again. However, these actions were seen as controversial and Stiglitz (2000) even believes "that the efforts of the IMF made the East Asian recession deeper, longer, and harder". Indonesia, South Korea and Thailand were the most affected countries (Bouchet). The costs of the crisis, e.g. in Indonesia, amounted to 55 % of GDP (Feenstra and Taylor, 2008). However, countries like Malaysia, who had forgone the help of the IMF, and economies with a higher level of international reserves like China, survived the East Asian financial crisis better than those with a lower level. Consequently, emerging Asia became distrustful of the IMF and the fundamental belief not to "rely on the IMF in the future even as a last resort lender" (Ito, 2007) was established. The self-insurance argument became crucial in the management of the foreign reserves to these countries.

What is optimal and what is excessive?

Econometric approach

To make a judgment about whether a country has excessive foreign reserves, there must be a concept of optimal reserves. Theoretical and empirical literature dealing with this topic emerged during the 1960s and 1970s. In the last decade a new interest in the topic arose, caused by the observation of the rapid accumulation of foreign reserves in emerging Asia. Many econometric models were developed trying to explain the level of foreign reserves. The determinants of reserve holdings used in these models can be divided into five groups, namely economic size, current account vulnerability, capital account vulnerability, exchange rate flexibility and opportunity cost (Gosselin and Parent, 2005).

Economic size, measured in GDP and GDP per capita, matters as it is assumed that international transactions rise with economic size. The higher the population and the GDP per capita are, the higher the amount of foreign reserves accumulated and needed will be (Edison, 2003). Current account vulnerability represented by trade openness and export volatility means that

a more open economy is more exposed to external shocks and should have more foreign reserves (Gosselin and Parent, 2005). Capital account vulnerability demands more foreign reserves, as with greater financial openness and larger potential for resident-based capital flight from domestic currency crisis, vulnerability rises (Edison, 2003). Exchange rate flexibility can be measured by the volatility of the exchange rate. In a fixed exchange rate regime more foreign reserves are needed because the central bank needs reserve holdings in order to maintain the fixed exchange rate. And a negative correlation can be found between opportunity cost and foreign reserve holdings. The higher the opportunity costs, the lower the foreign reserves will be (Edison, 2003).

   Edison (2003) developed an empirical model taking all of these determinants into consideration. At first, she estimates her model by using panel data for 122 emerging markets over the period of 1980 to 1996. As a result, the model can explain the level of foreign reserves depending on economic size, current account vulnerability, and exchange rate flexibility, whereas capital account vulnerability and opportunity cost are not found to be significant determinants.

| Year	Actual	Predicted	Difference
1990	202,14	187,59	14,55
1991	247,31	223,95	23,36
1992	253,36	263,48	-10,12
1993	295,45	312,38	-16,93
1994	371,18	366,86	4,32
1995	420,36	450,85	-30,49
1996	483,11	513,03	-29,92
1997	506,37	564,09	-57,72
1998	569,94	571,68	-1,74
1999	648,23	671,65	-23,42
2000	698,7	819,78	-121,08
2001	775,85	854,95	-79,1
2002	951,62	964,84	-13,22
2003	1217,13	1122,54	94,59
2004	1580,56	1363,65	216,91
2005	1827,08	1599,61	227,47
2006	2222,24	1900,56	321,68
2007	2916,75	2229,78	686,97

Table 2: Actual versus Predicted Nominal Reserves of Emerging Asia and Difference between Actual and Predicted Nominal Reserves, 1990-2007 (billion US$)[2]

2 See Park and Estrada (2009) expanded by the column "Difference"

Park and Estrada (2009) re-estimate Edison's model with panel data for 130 emerging economies from 1980 to 2004. Using this model Park and Estrada compute the predicted levels for emerging Asia's top ten reserve holders[3] and compare them to the actual levels for the period from 1990 to 2007 (Table 2, Figure 2).

Figure 2 shows that in the period from 1990 to 2002, the actual foreign reserves match the predicted values, although there are several years where the predictions even surpass the actual figures. The highest negative deviation can be found in 2000, where the prediction overestimates the foreign reserves in emerging Asia about US$120 billion. From 2003, however, the gap between actual and predicted foreign reserves grows rapidly (Park & Estrada, 2009). There is also a trend of increasing foreign reserves in the predictions. But whereas the predictions grow by 19 % on average from 2003 on, the actual reserves grow by 25 %. In 2003 the gap is US$95 billion and four years later it amounts to US$687 billion. That makes an increase of 623 % in four years.

Figure 2: Actual versus Predicted Reserves of Emerging Asia, 1990-2007 (billion US$) [4]
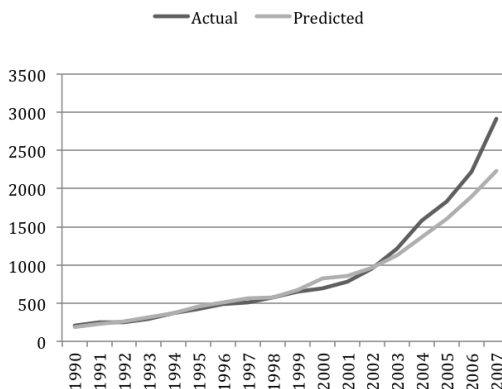
According to this model, it can be stated that the foreign reserves in emerging Asia are excessive as they are clearly above the levels that are predicted considering economic fundamentals.

However, there is also other literature which say that the foreign reserves are not excessive, but rather fit to the predictions and the strategy of

3 Note that these countries are the same as defined who is emerging Asia except for Vietnam.
4 See Park and Estrada (2009)

self-insurance. For example, Ruiz-Arranz and Zavadjil (2008), applying the model of Jeanne from 2007, state that the foreign reserves in this region of the world are in line with what can be predicted by "a simple model of optimal reserves applied to specific country and regional characteristics" (Ruiz-Arranz, Zavadjil, 2008). But nevertheless, they find that foreign reserve levels are close to the optimal level and therefore a slowdown in the accumulation is desirable.

Rules of Thumb

Another measure of the adequacy of foreign reserves are the so called "rules of thumb" like the reserves to import ratio or the reserves to short-term debt ratio. These ratios also support the judgment of excessive foreign reserves in emerging Asia. Although these measures "are based on general economic intuition rather than derived rigorously from formal theory" (Park and Estrada, 2009), they are often used in order to analyse the adequacy of foreign reserves.

Figure 4: Reserves to Import Ratio in Emerging Asia, 1990-201, (number of months)
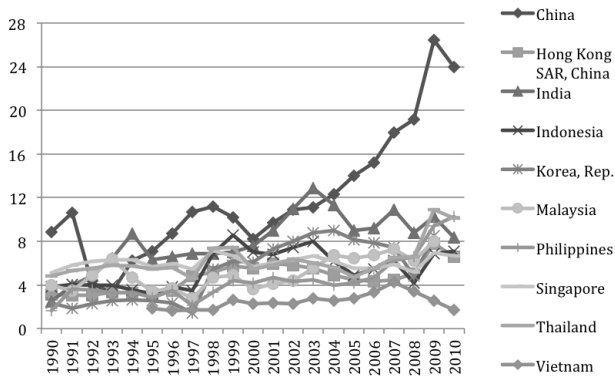
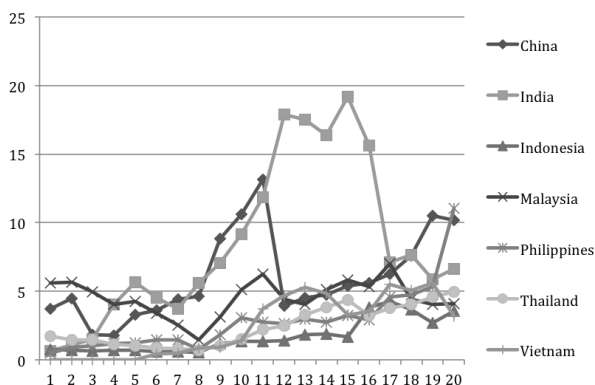(World Bank Data)

The reserves to import ratio (Figure 4) measures the number of months a country is able to finance in case that all money inflows like export revenues and external financing stop. It serves as a proxy for a country's current account vulnerability. In general, a three or four month benchmark is seen as adequate (Gosselin and Parent, 2005). Although this ratio and its benchmark

are criticized in terms of a lack of significance and empirical evidence, the IMF has been using this ratio as a measure for the adequacy of reserves since 1950.

Figure 4 shows that most countries have accumulated foreign reserves above the benchmark. Although some of the countries do not meet the benchmark even before the Asian crisis a wide gap arises afterwards. China and India in particular have the most striking reserve to import ratios, as they have the largest deviation from the benchmark. India has its peak in 2003 with approximately 13 months and China has had a steadily increasing ratio since 2000, with a peak in 2009 of 25 months.

The reserves to short-term external debt ratio (Figure 5) is also known as the Greenspan-Guidotti rule. In 1990 Alan Greenspan and Pablo Guidotti developed it, saying that countries should be able to serve at least all their external short-term debt without additional foreign borrowing for up to one year (Cifarelli and Paladino, 2008). It was also empirically supported by Bussière and Mulder (1999), and Jeanne and Rancière (2005) show the close relation of this ratio to the likelihood and depth of crises. The benchmark of this ratio is one, where foreign reserves equal short-term debt. (Edison, 2003)

Figure 5: Ratio of Foreign Exchange Reserves to Short-Term External Debt in Emerging Asia, 1990-2009

(World Bank Data)

So a benchmark above one shows that the country is able to serve its debt and can face the risk of a financial crisis, while a ratio below one can hint at a vulnerable capital account (Polterovich and Popov, 2003).

Figure 5 indicates that most countries are well above the Greenspan-Guidotti rule. The majority already had a higher benchmark before the Asian crisis. But after the crisis there were significant increases. India and China stand out again. China has its peak in 2000 with a ratio of 14 and India in 2004 with 19.

Current development

The global financial crisis hit emerging Asia hard, although it is far from the epicentre of the crisis. Emerging Asia, excluding China and India, lost 15 % of GDP on a seasonally adjusted annualised basis (IMF, 2009). As already examined, an important reason for holding reserves in emerging Asia is the self-insurance argument that has become crucial since the Asian crisis. If this were true, the global financial crisis could be seen as the "ultimate vindication for that strategy" (Dominguez, Hashimoto and Ito, 2011). Although they were able to recover faster from the crisis and experienced higher growth rates after the crisis again (Didier, Hevia and Schmuckler, 2011), there are beliefs that the reserves did not play an important role as a buffer in the crisis (Dominguez, Hashimoto and Ito, 2011). Looking at the development of the reserves, it can be concluded that the reserves were hardly used. There is no evidence of depletion. Only in India, Indonesia and Korea can reductions in their holdings from 2007 to 2008 be observed. In the other countries and in these three countries after 2008, a further increase is observable.

But nevertheless, Frankel and Saravelos (2010) find countries with higher foreign reserve holdings to have suffered less during the crisis. And De Gregorio (2011) concludes that the recent crisis showed that the reserves played "a rather deterrent role, in the sense that the sole act of having them reduces financial vulnerability", and not in the actual sense of the self-insurance motive, where foreign reserves act as a buffer stock of liquidity during times of sudden stops.

Conclusion

The examination of the adequacy of foreign reserves in emerging Asia proves to be a difficult and contentious topic. Edison's econometric analysis based on some economic fundamentals shows that the foreign reserves of emerging Asia are well above an adequate level. An informal investigation with some often cited "rules of thumb" concludes the same. But nevertheless this cannot be a definitive conclusion as Edison's model is just one among many models and as the "rules of thumb" have low empirical support. The problem is that there is no universally valid model determining the optimal level of reserves

for every country. Different models make different approaches with different assumptions.

What most economists agree on is the rapid accumulation of foreign reserves in emerging Asia, which has never happened to such an extent in any other region of the world before, and that the self-insurance argument is the most important explanation for the build-up. After the painful experience during the Asian crisis and with on-going financial integration, emerging Asian countries want to rely only on themselves, in terms of staying liquid during a crisis. So they accept the cost associated with holding reserves in order to prevent incurring the costs of a crisis.

But the costs of holding reserves are high not only for the country holding them, but also for the world, causing global imbalances which can result in a crisis such as the current one. The global financial crisis, which also spread to emerging Asia, showed that emerging Asian countries did not use their stockpile of foreign reserves; instead the accumulation went on. These countries survived the crisis better and recovered faster. But the question remains whether this was due to the foreign reserves and in particular, whether it was due to these high levels of foreign reserves. Would a lower level of reserves not have had the same deterrent effect?

Hence, foreign reserves can be both a curse and a blessing, as they can both cause and smooth a crisis. As long as no adequate level of foreign reserves can be determined, the discussion will go on.

References

Aizenman J. and Lee J., 2007. International Reserves: Precautionary Versus Mercantilist Views, Theory and Evidence. *Open Economies Review* 18:191-214

Allen F. and Hong J.Y., 2011. Why Are There Global Imbalances? The Case of South Korea.

Bahmani-Oskooee M. and Brown F., 2002. Demand for international reserves: a review article. *Applied Economics* 34:1209-1226

Bouchet, Dr. H. Asian Crisis and Consequences. SM71.9005 Country Risk Management.

Cifarelli G. and Paladino G., 2008. Reserve overstocking in a highly integrated world. New evidence from Asia and Latin America. *The European Journal of Finance* 4:4:315-336.

De Gregorio, J., 2011. *International Reserve Hoarding in Emerging Economies*. Economic Policy Papers, Central Bank of Chile.

Didier T., Hevia C. and Schmukler S. L., 2011. How Resilient Were Emerging Economies to the Global Crisis? Policy Research Working Paper No. 5637

Dominguez K. M. E., Hashimoto Y. and Ito T., 2011. International Reserves and the Global Financial Crisis. NBER Working Paper No. 17362

Edison, H., 2003. Are Foreign Reserves in Asia too High? World Economic Outlook.

Feenstra R.C. and Taylor A. M., 2008. *International Economics*. New York: Worth Publishers.

Frankel J. A. and Saravelos G., 2010. Are Leading Indicators of Financial Crises Useful For Assessing Country Vulnerability? Evidence from the 2008-09 Global Crisis. NBER Working Paper No. 16047.

Gosselin M.-A. and Parent N., 2005. An Empirical Analysis of Foreign Ex-

change Reserves in Emerging Asia. Bank of Canada Working Paper 2005-38.

Green R. and Torgerson T., 2007. Are High Foreign Exchange Reserves in Emerging Markets a Blessing or a Burden? Occasional Paper No. 6.

Independent Evaluation Office 2011. International Reserves: IMF Advice and Country Perspectives. Draft Issues Paper.

International Monetary Fund, 2011. Assessing Reserve Adequacy. IMF Policy Papers

International Monetary Fund, 2009. Global Crisis: the Asian Context. Regional Economic Outlook : Asia and Pacific.

International Monetary Fund, 2004. Guidelines for Foreign Exchange Reserve Management.

Ito, T., 2007. Asian Currency Crisis and the International Monetary Fund, 10 Years Later: Overview. *Asian Economic Policy Review* 2:16-49.

Jeanne O. and Rancière R., 2006. The Optimal Level of International Reserves for Emerging Market Countries: Formulas and Applications. IMF Working Paper No. 06/229.

Krugman, P., 1994. The Myth of Asia's Miracle. Foreign Affairs 73:6:62-79.

Nugée, J., 2009. Foreign Exchange Reserves Management. Handbooks in Central Banking No. 19.

Park D. and Estrada G. B., 2009. Are Developing Asia's Foreign Exchange Reserves Excessive? An Empirical Examination. ADB Economics Working Paper Series No. 170.

Polterovich V. and Popov V., 2003. Accumulation of Foreign Exchange Reserves and Long Term Growth. MPRA Paper No. 20069.

Ruiz-Arranz M. and Zavadjil M. 2008. 'Are Emerging Asia's Reserves Really Too High?'. IMF Working Paper No. 08/192

Stiglitz, J. 2000. 'What I Learned at the World Economic Crisis'. New Republic

World Bank Data Available at: http://data.worldbank.org/

Wyplosz, C., 2007. The Foreign Exchange Reserves Buildup: Business as Usual? Graduate Institute of International Studies and CEPR.

The Role of Inheritance within Capitalism

Sean Tong
Senior Freshman

Individuals often interpret the right to bequeath wealth as being part of the capitalist system. Sean Tong takes issue with this interpretation, arguing that no such association exists, and that the curtailment of inheritance can further the capitalist ideal of economic growth and consequently foster greater societal welfare.

Introduction

There can be little doubt that the practice of inheritance contributes to the existence of wealth inequality in capitalist societies: 31% of wealth in the United States can be attributed to bequests, in addition to the 20% gained through inter vivos transfers (Gale and Scholz, 1994). As will be shown in the following section, the distribution of inheritances is highly asymmetric. A small minority of the population collect large amounts while the vast majority receive little or nothing.

It will be argued in this essay that, far from being a central tenet of capitalism, the practice of inheritance actually runs counter to its principles and undermines the meritocratic ideals we claim to espouse. The case for a significant curtailment of inheritance will be presented from a theoretical perspective, before common objections are considered. It is worthwhile first to examine the nature of inheritance in Ireland, and to show the role it plays in perpetuating social stratification.

Inheritance in Ireland

Gifts and inheritances in Ireland are currently subject to capital acquisitions tax at a rate of 30% on amounts above a certain threshold. The value of this threshold is dependent upon the relationship between the two parties, and is set at €250,000 for transfers from parents to children[1].

Unfortunately, there is a distinct lack of quantitative data regarding inheritance in the Irish economy, at least of late. Nolan (1992) presents the

---

1 http://www.revenue.ie/en/tax/cat/

most recent source, the 1987 ESRI 'Survey on Income Distribution, Poverty and Usage of State Services'. This section will discuss some of the more salient trends found in the data, which examined the value of real property and businesses inherited by different households.

The distribution of inheritances received in the sample was highly asymmetric. Figure 1 (below) shows the mean value of inheritance received as a function of social class. There is a clear correlation between the two, with the value of the former rising steadily as we move from unskilled manual labourers (£33,216) to higher professionals (£215,732). Not only is the distribution of inheritances asymmetric, it is clearly skewed towards the more affluent classes.

Figure 1 : Mean Inheritance (£) by Social Class
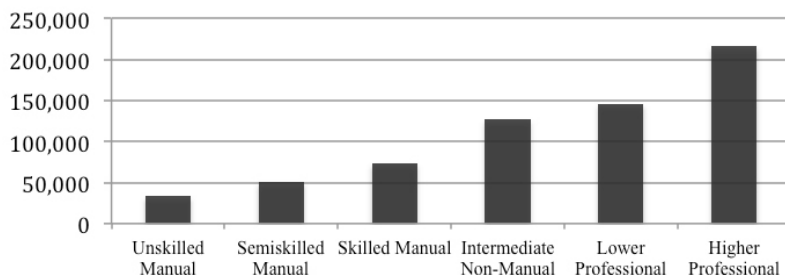
Further evidence of this comes from the link between current total wealth and inheritance. Figure 2 (below) shows that households in the top wealth deciles were far more likely to have received an inheritance. The majority (60.8%) of those in the top 10% had received an inheritance, while almost none (1.6%) of those in the bottom 20% had.

Figure 2 : Beneficiaries (%) by Current Total Wealth
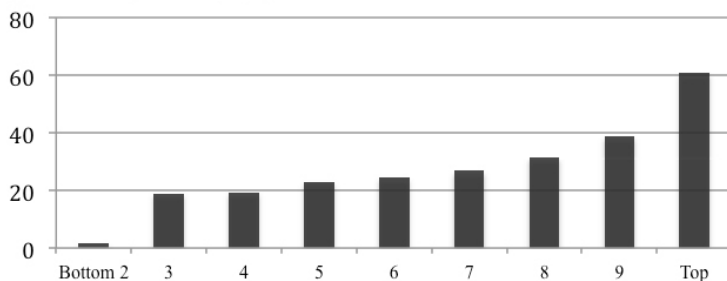
As well as demonstrating that the wealthy are more likely to have received inheritance, the data regarding current total wealth reflect those of Figure 1 in that those in the higher deciles also received larger mean inheritances. So while the mean value of inheritances received by those in the bottom 20% was just £25,984, the corresponding value for the top decile was £199,957.

It would appear that the practice of inheritance in Irish society, as elsewhere, serves to exacerbate wealth inequality and favour the rich. Those who have least tend to inherit least, while those who have more tend to receive more. It is this inequality that serves as the basis for the argument against inheritance.

The General Proposal

The exact form that a curtailment of inheritance (if desirable) should take is open to debate, and is not the focus of this essay. In order that arguments for and against a restriction may be considered more meaningfully, however, a general scheme will be outlined here that draws upon and incorporates suggestions from a number of different sources.

Ascher (1990) proposes full abolition of inheritance, with the property of the individual being auctioned at death and the proceeds accruing to the State. He includes a number of exemptions in this, of which I believe transfers to spouses and dependent lineal descendants have merit. Similar provision regarding inter vivos transfers would, of course, be necessary to prevent the use of gifts to circumvent the new legislation.

A suggestion made by Haslett (1986) may also be worth incorporating, as he includes a provision that would allow individuals to write a 'will' of sorts. Rather than specifying who should inherit what, this would give allocated individuals the opportunity to purchase specific items of the deceased's property at an independently-appraised price, in lieu of a free auction. This would allow for the safeguarding of family heirlooms, businesses and the like. The Government may provide loans to assist people in this regard.

This very brief outline should serve only as a reference point when considering the arguments for and against. It is clearly not the only means of curtailing inheritance: Haslett (1997), for example, believes that allowing a lifetime quota for gifts and inheritances would strike a more sensible compromise between distributive justice and cultural norms than complete abolition. It is worth stressing once more that the precise form that a restriction of inheritance should take is not the focus of this essay; rather, it is to establish whether a restriction of some sort is desirable in theory.

The Argument for Curtailment

The abolition or significant curtailment of inheritance can be justified on four primary grounds, which will be presented in this section. In brief, it will be argued that such a proposal would bring us closer to both our capitalist and meritocratic ideals, reduce wealth inequality, and provide a significant source of government revenue.

Capitalist ideals

In his lectures on jurisprudence, Adam Smith himself said that "there is no point more difficult to account for than the right we conceive men to have to dispose of their goods after death" (Fleischacker, 2004: 198). This is an entirely fabricated social right, and one that runs counter to capitalist ideals. The free enterprise system is predicated upon rewarding and, consequently, incentivising production. Remuneration is allocated to those who provide the factors of production, and in proportion to that provision. Inheritance stands in stark contrast to this principle: it is not earned, but comes about entirely as the result of a parental lottery. From a purely economic perspective, the random allocation of resources cannot possibly lead to an efficient outcome. To allow such a state of affairs is to undermine the working of the 'invisible hand' by which market activities are guided.

Chester (1976) rightly points out that the present arrangement of taxing incomes more heavily than inheritances is completely at odds with capitalist values, and serves to dampen productivity. A return to these values would necessitate a shift from the taxation of achievement (income) to ascription (inheritance).

Though it is often branded as a Communist proposal, the restriction of inheritance is entirely compatible with, and indeed demanded by, the central tenets of the capitalist system. Contrary to popular perception, the practice of inheritance and the advocacy of capitalism do not go hand in hand. To allow the former is to undermine the latter, and so the curtailment of inheritance must be seen as a means of returning to capitalist ideals.

Meritocratic ideals

While the practice of inheritance runs counter to our economic system, it is also in conflict with one of our social goals: that of meritocracy. If we are to expect individuals to compete in the market economy, it is imperative that the rules governing this are fair and uniformly imposed. Inheritance undermines this by significantly reducing equality of opportunity.

To be clear: the concept of equality of opportunity is entirely dis-

tinct from equality of outcome. Income disparities arise naturally in a capitalist economy and are a reward for productivity. Unequal opportunity, on the other hand, runs counter to the meritocratic ideals that we claim to espouse. Inheritance clearly contributes to this, as those born to wealthy families inherit more than those born to poor ones. Wealth is opportunity, and to receive more of the former is to have more of the latter.

Aside from the ethical implications of this, there are clear economic disadvantages to an unequal distribution of opportunity. As Stiglitz points out: "whenever we diminish equality of opportunity, it means that we are not using some of our most valuable assets—our people—in the most productive way possible" (Stiglitz, 2011). Buffett puts it more bluntly, comparing it to "choosing the 2020 Olympic team by picking the eldest sons of the gold-medal winners in the 2000 Olympics" (Sahadi, 2007).

A simple analogy, presented by Haslett (1986), makes this point easier to appreciate. We consider a hypothetical race between two competitors: in one case, the two start at the same point along the track, whereas in the second case one of the competitors is given a significant head start. It does not seem contentious to state that it is in the first of these races, which represents equal opportunities, that the effort exerted by the runners will be maximised. In the second case, poor motivation results from complacency and hopelessness, from the perspectives of the advantaged and the disadvantaged respectively.

Insofar as the capitalist system is predicated upon competition, the current state of affairs represents a rigged game. The rewards of production cannot be distributed accurately if outcomes are achieved through initial luck rather than present contribution. The practice of inheritance essentially exempts the children of the wealthy from the economic competition that others are subject to and that all have agreed upon. It is socially and economically corrosive, and for this reason its curtailment can be justified.

Wealth inequality

Though much focus has been placed on the disparity between high and low income earners, a much greater inequality is found with regards to wealth (Cagetti and De Nardi, 2006). In Ireland, the wealthiest 1% control 20% of all wealth, and the top 5% hold 40% of wealth (Bank of Ireland, 2007). While the income inequalities that arise naturally from market competition represent rewards for productivity, the wealth inequality that arises from inheritance does not. Rather, it is a means of perpetuating financial nobility, concentrating unearned wealth in the hands of an undeserving few.

By fragmenting large familial fortunes, a restriction upon inheri-

tance would bring about a number of social benefits. The economic and political power that concentrated wealth commands, and which may lead to distortions in the marketplace, would be reduced. In addition, as alluded to earlier, the incentive to work would be restored in the children of wealthy families. The reduction in wealth inequality that restrained inheritance would necessitate could thus serve a number of social and economic purposes.

Government revenue

The final, obvious benefit of this proposal is the significant government revenue that would be raised. While precise data are lacking in an Irish context, there can be no doubt that the heavy taxation of intergenerational transfers would represent a large, much-needed increase in the State's income.

What distinguishes this form of revenue from others is that it is a remarkably painless one. Jeremy Bentham (1795) considered it to be "absolutely the best" form of taxation, given that it affects prospective rather than current owners of property. Although affected heirs would doubtless feel aggrieved, this stems entirely from current expectations. Once the status quo is changed so that children no longer expect large transfers of unearned wealth, the disappointment caused by the proposal would dissipate to triviality.

The revenue raised from increased taxation on inheritance could be put to innumerable uses, some of which will be briefly referred to throughout this essay. It could, for example, be used to counteract some of the supposed negative externalities that a restriction on inheritance would bring, or to reduce rates of income tax to even further reward achievement over ascription. While the exact use of the revenue is beyond the remit of this essay, it should be clear that a curtailment of inheritance would be a significant contributor in this regard.

Response to Criticisms

In examining the literature, six primary objections to the curtailment of inheritance are seen to predominate. These will now be considered and evaluated in turn.

Violation of a natural right

It is often claimed that the testator has a 'right' to dispose of their property as they see fit, and that inheritance tax represents a violation of this right. In reality, however, inheritance is not a natural right but a civil one granted by society. This is illustrated most clearly by the heavy state involvement that is required to actually facilitate the practice of inheritance (Trout and Buttar,

2001).

The State acts as an intermediary between the original owner of property, whose claim ceases with his or her death, and the heir, whose ownership only comes into effect once the State legitimates the transfer. In the interim, assets are frozen, creditors are repaid, and competing claims are assessed. It is thus only through State involvement that the 'right' to inherit can even come about, which renders untenable any claim that such a right is a natural one.

As Haslett (1986) demonstrates, it is equally fallacious to claim that inheritance taxes are violations of property rights. To do so is to be guilty of begging the question. The right to property is already subject to a number of qualifications and exceptions, including income tax. The debate is whether or not inheritance should be considered a part of this qualified right to property. It is circular reasoning to simply claim outright that it ought to be.

Reduced incentive to work
It is argued that restricting inheritance would severely reduce the individual's incentive to work, under the assumption that the motive to bequeath is a significant contributor to the overall desire to produce. There are, however, clearly a multitude of reasons that people work, with bequests being just one of these.

Amongst these other motives are the desires to fund a good lifestyle, to achieve self-actualisation, and to guard against uncertainty. Haslett (1986) illustrates the insignificance of this effect by considering childless couples and professional athletes. There is no empirical evidence to suggest that the former group are less productive than couples with children (McClelland, 1961), and it seems risible to suggest that the latter would become less competitive should more stringent inheritance legislation be introduced. Empirical studies by both Hurd (1989) and Gan et al. (2004) estimated a very small bequest motive, and the latter concluded that most bequests are made unintentionally.

While it is unlikely that a limitation on inheritance would lead to a reduction in the incentive to work, it is quite likely that it would produce the opposite effect: that there would be an overall increase in this very incentive. Returning to the race analogy used earlier, it seems probable that an increase in equality of opportunity would increase the effort exerted by both the rich (out of newfound necessity) and the poor (out of newfound ambition).

Reduced incentive to save and invest
In a similar vein to the previous objection, it is claimed that a reduction in savings, and consequently investment, would result from this proposal (Tull-

ock, 1971). It is worth remembering at the outset that corporate and State savings would be entirely unaffected by changes in inheritance legislation, so any influence must be restricted to private savings.

As with the incentive to work, other motives clearly exist to save. Amongst these are the desires to provide for retirement and to guard against uncertainty in old age (Kessler and Masson, 1989). Given the inherent unpredictability of death, frivolous consumption is unlikely to predominate in the twilight years, even with well-developed markets for annuities. If this were not the case, the revenue raised from this proposal could easily be used to redress reductions in savings and investment, whether directly or through subsidies and the like.

The existence of other inequalities

While there is no doubt that true equality of opportunity can never be achieved, and certainly not without resorting to extreme measures, the presence of other inequalities in our society can in no way excuse the damage done by inheritance. While many forms of inequity are beyond our control, inheritance, as a socially constructed privilege, is not one of these.

The most remarkable thing about this argument, given its obvious flaws, is that so many find it convincing. Milton Friedman (1962) points out that significant individual differences exist in the allocation of natural abilities, and claims that the unequal distribution of inheritances should be similarly accepted, as there is no ethical difference between the two.

This reasoning is flawed on two accounts. As already stated, it is absurd to claim that the presence of one inequity should justify the existence of another. In addition, Friedman's suggestion that the two forms of inequality are comparable is untenable. What he fails to identify is that individual differences do not, in and of themselves, lead to income or wealth. It is, rather, the utilisation of these abilities in a manner that is valued by society that brings such a reward. It is this distinction that makes one justified, and the other not. A reform of inheritance legislation neither claims nor seeks to establish perfect equality; rather, it aims to restore some equality of opportunity in a domain over which we have complete control.

Practicalities

For the sake of brevity, this essay considers the restriction of inheritance from a purely theoretical perspective. For this reason, common practical objections such as the ineptitude of governments, administrative costs, and likely public outrage will not be considered here. While there is merit to some of these, and it is likely true that the unilateral implementation of radical measures would

be problematic, these difficulties are certainly not insurmountable. A more in-depth treatment can be seen in, for example, Ascher (1990).

Conclusion

This essay has articulated the argument for the abolition or significant curtailment of inheritance. It has been shown that it is in conflict with our economic and social ideals, and that to tax it more heavily would bring a number of ancillary benefits. Furthermore, the arguments in support of inheritance tend to be fallacious or inconsequential.

Though the practicalities of implementing a proposal of this sort, and indeed the specifics of the proposal itself, cannot be adequately considered in this essay, the conclusion remains: if we are to be consistent with our ideals, we must take measures to limit the role that inheritance plays in influencing individual outcomes.

References

Ascher, M.L., 1990. Curtailing inherited wealth. *Michigan Law Review*, 89:1:69-151.

Bank of Ireland, 2007. *The Wealth of the Nation*.

Bentham, J., 1795. *Supply without burden*.

Cagetti, M. and De Nardi, M., 2006. Wealth inequality: data and models. Working Paper 12550, Cambridge: National Bureau of Economic Research.

Chester, C.R., 1976. Inheritance and wealth taxation in a just society. *Rutgers Law Review*, 30:62-101

Fleischacker, S., 2004. *On Adam Smith's "Wealth of Nations": A Philosophical Companion*. New Jersey: Princeton University Press.

Friedman, M., 1962. *Capitalism and Freedom*. Chicago: The University of Chicago Press.

Gale, W., and Scholz, J., 1994. Intergenerational transfers and the accumulation of wealth. *Journal of Economic Perspectives*, 8:4:145-160.

Gan, L., Gong, G., Hurd, M. and McFadden, D., 2004. Subjective mortality risk and bequests. Working Paper 10789, Cambridge: National Bureau of Economic Research.

Haslett, D.W., 1986. Is inheritance justified?. *Philosophy and Public Affairs,* 15:2:122-155.

Haslett, D.W., 1997. 'Distributive Justice and Inheritance' in G. Erreygers and T. Vandevelde (eds.) *Is Inheritance Legitimate? Ethical and Economic Aspects of Wealth Transfers*. Berlin: Springer.

Hurd, M.D., 1989. Mortality risk and bequests. *Econometrica*, 57:4:779-813.

Kessler, D. and Masson, A., 1989. Bequest and wealth accumulation: are some pieces of the puzzle missing?. *The Journal of Economic Perspectives*, 3:3:141-

152.

McClelland, D.C., 1961. *The Achieving Society*. Princeton: Van Nostrand, pp. 234-237.

Nolan, B., 1992. The Pattern of Inheritance in Ireland: Household Survey Evidence. Working Paper 35, Dublin: The Economic and Social Research Institute.
W
Sahadi, J., 2007. Buffet: tax my kin, please. [Online] Available at: http://money.cnn.com/2007/11/13/pf/taxes/buffett_estate_tax/index.htm [Accessed 18 February 2012].

Stiglitz, J., 2011. Of the 1%, by the 1%, for the 1%. [Online] Available at: http://www.vanityfair.com/society/features/2011/05/top-one-percent-201105 [Accessed 18 February 2012].

Trout, J.D. and Buttar, S.A., 2001. Resurrecting "Death Taxes: inheritance, redistribution, and the science of happiness. *Journal of Law and Politics*, 16:765-841.

Tullock, G., 1971. Inheritance justified. *Journal of Law and Economics*, 14:2:465-474.

Should Policy Makers Limit the Size of Current Account Imbalances?

Cian Ruane
Senior Sophister

While the focus of talk about the global economic crisis has been on the financial sector and its problems, current account imbalances have also contributed to the instability the world is facing. Cian Ruane analyses these imbalances, why they exist and how they have impacted the world economy, concluding that it is necessary for policy makers to limit their size.

Introduction

Global current account imbalances have widened dramatically since the mid-1990s. In the 2000s a disorderly unwinding of global imbalances was seen as the greatest threat to global economic stability. The contribution of imbalances to the global financial crisis makes their persistence worrying for policy makers. Indeed, many of the structural issues that created these imbalances are still in place. This has led policy makers to consider whether international regulation is required. US Secretary of the Treasury Timothy Geithner proposed at the 2011 G20 meeting in Washington a limit on current account[1] (CA) surpluses and deficits of 4% of GDP (Cartapanis, 2011).

In this paper I argue that global imbalances are currently on an unsustainable path. Despite these imbalances being largely due to market distortions in developing countries, international cooperation and regulation will be required to stabilize the global economy in the medium-term.

In section one I discuss the framework of growth theory and examine the behaviour of global imbalances over the last 30 years. In section two I examine the argument that global imbalances were the underlying cause to the global financial crisis of 2008. In section three I evaluate the sustainability of the current path of global imbalances, and what risks they hold for the

---

1 In an open economy, the current account (CA) measures international transactions in goods, services and income. A CA deficit implies that a country is a net borrower, or that national savings don't meet investment demand.

global and European economies. Finally, in section four I detail the various mechanisms through which imbalances can correct themselves and consider the question of when policy makers should intervene to rebalance global current accounts.

Global Imbalances in Theory and in Practice

Classical growth theories provide a framework that can explain how global imbalances arise as a result of the normal functioning of the market. The prediction is that rational investors will invest in developing markets, in search of higher risk-adjusted returns (Feenstra and Taylor, 2011). The influx of capital into poor countries will lead to higher growth expectations, inducing consumers to consume more today. The increase in imports will lead to a CA deficit. What theory predicts, therefore, is a world where developing countries run CA deficits and advanced economies run CA surpluses. One should then see a convergence effect as developing economies grow quickly, lowering the marginal product of capital to the point that capital inflows start automatically rebalancing. This convergence effect should continue to take place so long as the Long-Run Budget Constraint is satisfied[2]. Countries can therefore run deficits for substantial periods of time, as their enhanced growth prospects will allow them to run CA surpluses in the future to pay back their debt. This was particularly evident in the European periphery from the period 1995-2008 (Lane and Pels, 2011).

Blanchard and Milesi-Ferretti (2010) differentiate between 'good' and 'bad' imbalances. For example, aging economies require more saving and hence will run CA surpluses. Also, capital will flow to where investment opportunities are best and also to where there are very liquid and deep financial markets. In this way, imbalances contribute to the efficient allocation of resources. Looking at the period 2000-2008, a significant proportion of surplus imbalances were run by oil-exporting countries that were responding to a higher price of oil and needed to hedge against future price volatility (indeed the price of oil did collapse at the beginning of the 2008 crisis).

We can therefore see that large CA imbalances can occur due to 'good' economic fundamentals and that global imbalances are not inherently 'bad' as is often suggested. Meddling in international balances can therefore have negative effects, not allowing countries to smooth consumption as they need and leading to inefficient investment. In general, policy makers need to be cautious when deciding whether imbalances need 'correcting' or not.

However, despite the evidence for convergence from Europe, glob-

2 According to the LRBC, an open economy's constraint differs from a closed economy's in that the net present value (NPV) of expenditure must equal the NPV of disposable income.

ally what we have seen over the last 20-30 years is an uphill flow of capital from poor to rich countries. This was famously brought to the fore in Lucas' paper 'Why Doesn't Capital Flow from Rich to Poor Countries' (1990). The predicted imbalances are therefore in reverse, with many advanced economies such as the US running consistent CA deficits while many emerging economies (particularly in emerging Asia) run CA surpluses.

Many explanations have been put forward to explain this. Lucas hypothesized that lower levels of human capital in developing countries, less advanced technology, worse institutions and political risk might all be contributing factors (1990). The current direction of imbalances can also be explained by distortions in emerging markets and advanced markets alike (Blanchard and Milesi-Ferretti, 2010). An exceedingly high savings rate in China is often explained by a lack of social insurance. Similarly, exceedingly high growth expectations in the US and in other advanced economies were seen to drive CA deficits in these countries by keeping savings rates low (Lane and Pels, 2011, Engel and Rogers, 2006). Another explanation is that developing countries, having suffered a sequence of currency crises in the 1980s and 1990s and built up large international reserves[3], are now focusing on export-led growth, artificially keeping their exchange rate depreciated and CA in surplus. Though these kinds of policy manipulations can be rational for an individual nation, they create systemic distortions and imbalances when a large group of countries behave in this way (Blanchard and Milesi-Ferretti, 2010).

Indeed what this state of the world illustrates is the difficulty in understanding the myriad of factors that determine global current accounts. Though policy makers manipulating their CA can be rational on an individual level, the imbalances created need not be 'good' imbalances and can create instability in the global economy. Indeed, we observe in the next section how these may have contributed to the crisis.

The Contribution of Global Imbalances to the Financial Crisis and its Aftermath
The Financial Crisis
Throughout the 2000s it was possible to divide the world into deficit and surplus countries. The main surplus countries included the oil-exporting countries, emerging Asia (China especially from mid-2000s on), and Germany.

---

3 These currency crises rendered countries very vulnerable to 'sudden stops' of capital. The results of these sudden stops were CA reversals that proved damaging to growth (Edwards, 2005). Capital movements and the CA are linked through the financial account, which is a measure of net changes of capital ownership. It is important to note that the presence of international reserves can result in the CA not being affected by capital flows in certain situations.

The deficit countries included the US and many European economies. The US deficit consistently increased throughout the 1990s and 2000s, from 4.3% in 2000 to 6% in 2006. This was of primary concern to policy makers, the main fear being a disorderly unwinding of global imbalances, causing a sudden depreciation of the dollar which would have massive repercussions for the global economy. Countries pegged to the dollar or holding large dollar reserves would be particularly affected.

However the crisis came from elsewhere, originating in the financial sector and in the 'shadow banking system'. Economists still debate the role of global imbalances in creating the crisis. Many had argued during the mid-2000s that the large US deficit was not actually a risk to global stability. Indeed, the large deficit could be explained by the backward financial system in many emerging markets, leading them to invest in the highly liquid US market (Dooley, Folkerts-Landau and Garber, 2005). This was a "win-win" situation which was completely harmless to the world economy. Others however claim that global imbalances were the underlying cause of the financial crisis (Portes, 2009).

As argued by Obstfeld and Rogoff (2009), I would agree that the global crisis was in itself a financial crisis. It originated in a highly leveraged financial sector that was kept in place by a financial asset bubble. Given the internationalisation of the banking sector, the crisis spread through financial channels, triggering banking and liquidity crises in many countries. That is not to say that global imbalances were not a contributing factor or that limits on global imbalances might not have reduced the size of the crisis or prevented it altogether.

An important factor in the crisis was how low global real rates were in the 2000s. This led to an underpricing of risk and cheap credit, fuelling asset bubbles and increasing leverage. Whether low global rates were due to high levels of saving in emerging Asia and China or the "Global Savings Glut" (GSG)[4], as argued respectively by Former US Treasury Secretary Henry Paulson and Chairman of the Fed Ben Bernanke, or due to the collapse of the technology bubble in the early 2000s (Obstfeld and Rogoff, 2009), the capital flows that helped finance the bubble in the US (causing a CA deficit) were due to CA surpluses in emerging Asia being refinanced into US debt.

Also significant were the low postponement costs in China and the US of correcting CA imbalances (Obstfeld and Rogoff, 2009). Indeed, as a surplus country it was not difficult for China to continue the sterilization pro-

---

4 This GSG was partially due to market distortions in emerging Asian economies, but also due to aging populations in more advanced economies such as Japan or Germany where there weren't sufficient domestic investment opportunities at home.

cess, and as a global reserve currency the US had the "exorbitant privilege" of being able to sustain levels of debt that would lead to investor panic for any emerging market (Roubini, 2010). Due to this, the US was not forced to rectify its budget deficits throughout the 2000s, and was not forced to take counter-cyclical fiscal policies that could have slowed the growth of the bubble. In short, whatever the 'underlying causes' of the crisis, I would argue that had CA limits been in place or attempts made to reduce imbalances the crisis may well have been of a lesser magnitude.

The Period 2008-2011

Global imbalances played an important role in determining how various countries were affected by and responded to the global crisis. As expected, the growth effects of the crisis were most severe for countries running large CA deficits[5] (Lane and Milesi-Ferretti, 2011). The global CA imbalances also played a role as a propagation mechanism for the crisis (BIR, 2011). The dramatic adjustment in the US CA (6% to 3%) meant that there was a sudden collapse in demand for goods from surplus countries. Emerging Asian and other surplus economies that were more or less insulated from the financial transmission mechanism of the crisis were therefore affected through the balance of payments mechanism.

However the most dramatic result that the crisis had was in accentuating the imbalances in the euro area. Throughout the 2000s, though the euro area as a whole tended towards a balanced CA, there were large surplus and deficit countries (Lane and Pels, 2010). There was therefore an asymmetry in terms of the shocks that both types of countries were subjected to when the financial crisis of 2008 hit. On the one hand, countries like Greece, Ireland and Spain that had not been running significant surpluses until then were forced to run pro-cyclical policies. The heterogeneity of the euro area meant that the ECB could not simultaneously keep rates at the right levels for core and periphery countries. Since then the sovereign debt crisis has led to voiced scepticism of the euro area's sustainability. Had CA balances been in check during this period it is likely that the euro area's crisis would not have become as severe as it currently is. Looking into the future, the euro area now knows that it will need either some form of fiscal union or a macro-prudential European supervisory authority if it is to remain a currency union.

The Sustainability of Current Imbalances and Risks for the Global

5 They also note that most of the adjustment was through expenditure compression as opposed to switching; the crisis hit overall demand, not just demand for foreign goods.

Economy

A lot of CA auto-adjustments have already taken place over the last three years. The US CA deficit shrank from 6% of GDP in 2006 to 3% following the crisis. Exchange rates have moved slowly, with a slight appreciation of the Renminbi contributing to this adjustment process (BIR, 2011). Equally, in Europe, pressures to pursue fiscal adjustments have led to a compression of current account imbalances (Lane and Milesi-Ferretti, 2011). However China is still running a very large CA surplus and a lot of structural problems remain; namely high savings in China, low savings in the US and currency manipulation for export-led growth in emerging markets (Blanchard and Milesi-Ferretti, 2010).

I agree with Roubini (2010) and Blanchard and Milesi-Ferretti (2010) that there are large risks for the future should global imbalances not correct themselves further. Firstly, the longer CA imbalances remain, the greater the risk of these leading to trade protectionism (BIR, 2011). Indeed, were Chinese exports to stay cheap for the US it might seem in US policy makers interest to introduce policies restricting the imports of Chinese goods. However, as was seen during the Great Depression, the effects of trade wars and protectionism can be devastating for the world economy. Another fear is that of a 'sudden stop' of capital flows to the US as investors lose faith in the USD. Obstfeld and Rogoff (2009) argue that investor faith in the US can no longer be taken for granted, and while the US has not seen any 'sudden stops' of capital, the S&P downgrade of US debt in summer 2011 is a worrying sign. They also argue that, while the US benefited from positive valuation effects on its external wealth prior to 2008, recent evidence has shown that these effects may well be going into reverse. If this were to happen it would further hurt US creditworthiness and increase the importance for the US to be able to run consistent surpluses.

The question is whether the path that current account imbalances will take without further policy maker intervention is sustainable. Indeed to ensure that none of the aforementioned dangers come to pass, we need to look at whether the auto-adjustments that are taking place will be sufficient to bring imbalances back onto a sustainable path.

There are many adjustment mechanisms that have already taken effect and may continue to do so. Continued but slow exchange rate appreciations in emerging Asia are likely, oil prices may stabilize reducing the surpluses run by oil exporting countries. Higher private saving in the US will continue as households try to deleverage, investment will fall, cross border premiums will rise and home bias will increase (Blanchard and Milesi-Fer-

retti, 2010). Though these may well stabilize current account balances at their current levels they are unlikely to cause dramatic corrections of imbalances in the short to medium term.

Even more dangerously, it can be argued that, while CA deficits pre-2008 were a sign of overheating, the scale of the compression of CA imbalances is due to undershooting (Lane and Milesi-Ferretti, 2011). In this case, these effects are cyclical and therefore temporary, meaning that further increases in imbalances are likely.

In light of the risks posed by global imbalances, and the uncertain nature of auto-correcting mechanisms functioning in the short to medium term, it is necessary for policy makers to intervene and find ways of correcting global imbalances.

Mechanisms of Rebalancing and Conclusion

From a national policy making perspective, the recent crisis has emphasized the importance of conducting monetary policy in a way that acknowledges the current account (Obstfeld and Rogoff, 2011). Had this been done in the US during the 2000s, as argued above the catastrophic consequences of the financial crisis might have been lessened. However national policy making that is effective in controlling global imbalances need to be done in both surplus and deficit countries for these changes to have significant effects. Not only that but there is an asymmetry between deficit and surplus countries' ability to control their own CA balances in the current environment (Cartapanis, 2011, Obstfeld and Rogoff, 2009). With the US in a recession, introducing policies that would reduce the CA deficit would also risk hurting recovery. However, it is not as difficult for surplus countries such as China to actively allow their currencies to appreciate, even though it would entail a slowdown in economic growth. The problem here, however, is that on a national level China faces short-term incentives to keep its surplus position as the world's largest exporter. Without some form of international agreement and negotiations it is unlikely that China will pursue this kind of policy independently. A trade-off with the US could be that the US commits to running tighter monetary policy, which would lessen the negative effects of China's exchange rate appreciation.

The CA surpluses in emerging Asian economies are another side of the imbalance that needs to be addressed. These countries have developed strong external positions in order to protect themselves from the risk of currency crises or speculation. Blanchard and Milesi-Ferretti (2010) argue that these states should stop basing their economy on export-led growth and look

at more domestic investment strategies. They recommend an extension of liquidity provisions for these countries, potentially through the IMF's Flexible Credit Line.

The final issue to consider is the timing of macro-prudential regulation. Indeed, while Blanchard and Milesi-Ferretti have emphasized that a "failure to act on the remaining domestic and systemic distortions that caused imbalances would threaten the nascent recovery", Cartapanis has questioned the wisdom of introducing macro-prudential regulations at this stage of a recovery citing studies by Allen ('Global Imbalances: Causes Consequences and Possible Solutions', 2011) that these kind of reversals are often associated with lower growth. Both agree, however, that corrections in global imbalances will need to begin with the surplus nations.

Though the current path of global imbalances is unsustainable, there is not necessarily a need for current account limits to be formally laid down as long as there is an informal agreement between all major economies to behave in such a way as to not exacerbate global imbalances. However, as we have seen in a lot of cases, the incentives faced by individual nations diverge from what is required for global stability. In my opinion a formal limit of 4% of GDP, as suggested by Timothy Geithner, is required to ensure that global imbalances begin to correct themselves. How this sort of limit should be enforced goes beyond the scope of this paper, however one consideration one would be the existence of a sunset clause, whereby the limits expire after 5/10 years. While the global economy may require CA limits now, these may well become a hindrance in the future and political pressure will not be enough to maintain them indefinitely.

References

Bank of International Settlements (BIS), 2011. 'Annual Report'. Available On-line at: http://www.bis.org/publ/arpdf/ar2011e.pdf

Blanchard, O.J., Milesi-Ferretti, G.M., 2010. 'Global Imbalances: In Mid-stream?', CEPR Discussion Paper, No. 7693.

Cartapanis, A., 2011. 'Is it Really a Good Idea to Want to Reduce Global Im-balances During a Crisis' tr., Associaton D'Economie Francaise.

Dooley, Folkerts-Landau and Garber, 2005. 'International Financial Stabiliity: Asia, Interest Rates and the Dollar', Deutsche Bank Global Research, NY. Edwards, S., 2005. 'Capital Controls, Sudden Stops and Current Account Re-versals', NBER Working Paper, 11170.

Fattouh, B., 2005. 'Capital Mobility and Sustainability: Evidence from U.S. Current Account Data', Empirical Economics, 30:245-253

Feenstra, R.C., Taylor, A.M., 2011. 'International Economics'. 2nd edition, Worth Publishers, USA.

Lane, P.R., Milesi-Ferretti, G.M., 2002. ' External Wealth, the Trade Balance, and the Real Exchange Rate', IMF Working Paper.

Lane, P.R., Milesi-Ferretti, G.M., 2006. 'The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970-2004', CEPR Discussion Paper, No. 5644

Lane, P. R., Pels, B., 2011. 'Current Account Imbalances in Europe', XXIVth Moneda y Credito Symposium, Madrid.

Lucas, R.E., 1990. 'Why Doesn't Capital Flow from Rich to Poor Countries?', The American Economic Review, 80, 2.

Obstfeld, M., Rogoff, K., 2009. 'Global Imbalances and the Financial Crisis: Products of Common Causes', CEPR Discussion paper, No.7606

Portes, R., 2009. 'Global Imbalances', Macroeconomic Stability and Financial Regulation: Key Issues for the G20, CEPR, London.

Roubini, N., Mihm, S., 2010. 'Crisis Economics'. Penguin Group, New York, USA.

Optimisation Heuristics in Paternalistic Public Policy

Tony O'Connor
Junior Sophister

In this paper, Tony O'Connor examines the motivations of paternalistic public policy. In doing so, he uncovers the subtle role of heuristics, or mental shortcuts, on policy formulation. This understanding is then applied to common paternalistic public policies, such as banned markets, Pigovian taxes, and choice architecture, identifying those based on heuristics, along with those based on more rational desires to enhance current and future societal happiness.

Introduction

"De gustibus non est disputandum."
"Over tastes, there can be no dispute".

Many consider the above quote praiseworthy, primarily because it helps an individual avoid the inane and irresolvable argument that seeks to establish the primacy of certain tastes over others, particular regarding music, food and other simple goods. Thus the quote is both useful and enlightening when a difference in preferences arises between two individuals: the choice of one does not impede upon the happiness of another.

Alas, public policy is quite different in this respect. In the face of varying preferences, one who judges policy faces a sort of intellectual peril, for he never knows whether his judgement on policy is based on an objective truth regarding societal well-being, or simply his own preference for the shape and form of society.

Concretely, we shall see that while in many cases the motivation for policy stems from a desire to increase the happiness of both present and future generations, other policies arise from a very specific definition of 'good' that the individual concerned has become accustomed to over the course of their life and learning. It is this latter heuristic, or rule-of-thumb, disguised as a preference, that the policymaker must be vigilant of.

Optimiser Type	Dominant Preference
Present Utilitarian	Happiness of Current Generations
Future Utilitarian	Happiness of Future Generations
Libertarian	Individual Freedom
Classical Socialist	Equality of Outcome
Traditionalist/Reactionary	Social Stasis and Reversion
Doctor	Maximum Average Life Expectancy

Figure 1: A Brief Taxonomy of Optimiser Types and Corresponding Preferences

Variance in the Sources of Preferences

In Figure 1, a brief taxonomy of different preferences are outlined. For each preference, a specific type of optimiser has been labelled who values that preference highly.

It is assumed here that those expressing preferences wish only to improve the present and future of human society. It is thus trivial to distinguish between a rational preference and a heuristic preference, by asking ourselves a question. If the single societal state that corresponds to the optimiser's preferences was enacted, is there likely to be greater happiness than in the state arising from a flexible preference (that of increasing the sum of happiness regardless of what must be done in the process)?

The answer here should only ever be a neutral or negative one, for if a specific state results in the maximisation of the sum of happiness over time, then we would hope that a rational, open society, with broad criteria for a good policy, would adopt this state, in order to achieve the maximum level of happiness possible.

The Libertarian believes that the optimal state maximises freedom, and thus freedom is the heuristic he uses when evaluating policy. In the past, when the government was quite uninformed about the choices and behaviour of individuals, then this view would have correlated almost perfectly with those who simply wished to maximise happiness. I say almost, as it was lucidly demonstrated by Sen that in certain situations where individuals face a list of choices, and the happiness of each person is dependent on the action of the other, then it may be impossible to achieve a Pareto-efficient outcome, defined as that where the happiness of one person cannot be raised without

the happiness of the other person falling (Sen, 1970).

An interesting addition to the list is the Doctor. He values life expectancy above all else, as 'good' results are associated with a healthier patient. Having become accustomed to this interpretation of goodness, it soon becomes a heuristic, which he applies to the world in general, outside of the context in which it developed. There are many worlds where individuals could live longer by refusing to partake in certain activities, such as the consumption of alcohol and certain types of food, and dangerous, thrill-seeking activities. However, as individuals routinely partake in such activities, in full knowledge of the potential consequences, we can infer that people do make a trade-off between length of life and the happiness derived from it, thereby negating the validity of the Doctor's heuristic, in policy terms.

A heuristic similarly possesses the mind of the Reactionary. Such a person is greatly distressed by the rapid change inherent in society, and comes to view it as inimical to his interests, and somewhat detrimental to the moral fibre of society. He thus comes to believe that 'what is old is good', and supports policies that arrest societal change, or that attempt to revert society to a previous state.

Policy Analysis

We have thus concluded that a rational policy, in an open society, will have as its only preference the increasing of the living standards and happiness of future generations. We now analyse paternalistic public policy, dividing the concomitant policies into policies that are rationally motivated and those that are heuristically motivated.

Undesirable and Illegal Markets

Certain markets are discouraged, and sometimes outlawed, by society, as they are perceived to sell goods that are damaging to the physical or moral health of the populace. Inherited societal values often play a prominent role here; a good example being the manner in which certain religions, such as Islam, prohibit alcohol. Where a large majority of the populace share these social mores, they are often enshrined in law.

Thus, as Hillman said, we can conclude that the markets that a society prohibits are expressions of community values, and are therefore heuristically motivated (Hillman, 2009). Contemporary examples in Western countries include prohibitions on markets for marijuana, prostitution, and organs.

The Market for Drugs: A Policy Based upon Community Values
Given that we know that such policies may be heuristically motivated, we must investigate whether they contribute to the greater good.

When banning drugs such as marijuana, the legislator may firstly consider the deleterious impact of drugs on the health and mental well-being of the addict and general consumer of drugs. Invoking a revealed preference analysis of the personal choice involved, it is most doubtful that an addict to drugs would say that if they were given the chance to go back in time, they would choose to become addicted once more. We can thus tentatively conclude that banning drugs does enhance the levels of happiness in the society in which it is implemented, specifically for potential users of drugs who would otherwise have become addicted.

However, the law is rarely successful in eliminating the market for drugs; rather, they push it into the underworld, and disadvantaged areas, where the strength of the law is tentative at best, nonexistent at worst. This creates an externality in that certain areas, usually disadvantaged ones, become havens of crime; this severely impacts on the well-being of the residents of those areas, who are not involved in such forbidden industries, but must tolerate the criminals and their lack of respect for both the law and the rights of other citizens.

A second externality occurs in the countries in which the drugs are cultivated. Money accumulated from the illegal sale of goods in the domestic country is used to finance illegal drug trafficking in the cultivating country; and using these revenues, drug cartels purchase weapons, political power and labour, thereby eroding the political system. The ongoing Drugs War in Mexico is a salient example of this externality.

Thus, although the overall effect on current and future happiness is uncertain and difficult to measure, it seems probable that the displacement effect of the externalities affects far more individuals from those who are 'saved' from the policy. As the ban benefits a small, specific group of individuals who would have been victims of their own choices, but disrupts the lives of many more who would not otherwise have been disrupted, it is unlikely that this policy is morally or practically justifiable.

The Market for Organs: A Rational Paternalistic Policy Based Upon An Evidence-based Analysis
Organ markets are outlawed as there are doubts over whether a market in organs would increase general welfare.

Firstly, if there were to be a market in organs and blood, there is a chance that the blood and organs supplied would be of a lower quality and

lower overall quantity. We may conclude this as the altruistic motivation for donating organs and blood would disappear. In addition, in the event that low income is correlated with poor health, then those with the greatest incentive to sell would be in possession of lower quality organs and blood (Hillman, 2009). However, this analysis ignores the fact that individuals may still donate their organs, and donate the money received to a fund to aid individuals in the need of organs, who cannot afford said organs.

The policymaker may also worry that, were organ markets to be legalised, then poor people would be quite willing to sell their organs for the immediate pleasure they would derive from the sale price of the organ. As would have discounted any future possible costs to their health, they would have concluded irrationally that the immediate pleasure is greater than the larger but distant and probabilistic cost, in the future, that could come from a lack of organs. This behaviour is known as hyperbolic discounting, and it implies that as time increases, so does the rate at which the future costs and benefits are discounted. A person 'suffering' from hyperbolic discounting will always undervalue future costs and benefits, and is thus not a very shrewd long-term decision maker (Hillman, 2009).

A striking example of this occurred in China where a teenager sold his kidney, in an illegal market, for the paltry sum of £2,000, which he subsequently spent on an iPad2 (Foster, 2011). It is thus not an unreasonable conclusion that the causes of welfare may be advanced by precluding market forces from the procurement and allocation of organs.

A Transient Analysis of the Policy of Compulsory Spending and Pigovian Taxes

In the preceding analysis, it was implied that some individuals cannot judge objective probabilities of harm to themselves, as they may conclude from previous experience, in which they were unharmed, that there is little chance of injury or disease. This assumption is often used in the justification of compulsory spending, whereby the government requires individuals to purchase, directly and indirectly, safety equipment such as helmets for bicycles, seat belts for cars, and safety equipment in the workplace.

It may also be used to justify a Pigovian tax, which is a tax on unhealthy goods, intended to dissuade an individual from their consumption. It should be evident that both of the preceding policies would be encouraged by a Doctor-type optimiser, who cares primarily about the length of life, and less about the quality of it.

However, in a lucid paper, Viscusi provided statistical evidence that cast doubt on the earlier assumption that was used to justify both these poli-

cies. Concretely, he found that individuals overestimated the risk of lung cancer from smoking. He further mentioned that the overestimation of risks is associated with risks that are well-publicised, in this case lung cancer (Viscusi, 1990).

This undermines the case somewhat for paternalistic policy regarding individuals and risks, and also for Pigovian taxes, along with providing further evidence for the dubious nature of heuristically motivated paternalism.

Libertarian Paternalism: A Rational Policy

Libertarian paternalism refers partly to constructing choices is such a way that inertia on the part of the individual will still result in optimal societal happiness. It is libertarian in that the individual is free to choose, and is paternalistic in that the choice is already made for you and stands if you take no action to change it. It is popularly known as 'nudging'.

A concrete example can be seen the case of organ donation. Thaler and Sunstein (1999) argued that to increase organ donations, a state could implement a policy of presumed consent. Under this, all citizens are presumed to be consenting donors. However, they could easily opt out if they wish. Thaler and Sunstein cited a study which demonstrated that when the default choice was changed from an opt-in to an opt-out, or presumed dissent to presumed consent, the number of people who consented to become organ donors increased from 42 percent to 82 percent. Where people were given no default, but simply asked to choose, the percentage choosing to donate organs was 79 percent (Thaler and Sunstein, 1999).

Thus, the policy of presumed consent resulted in a rate that closely approximated the preferences of the individuals concerned. The default would be that the organs would be donated. If the individual did not wish to see their organs donated, then they must opt-out.

The above example can be generalised to many areas where human inertia results in sub-optimal decisions, such as pensions, where a person intends to start contributing, but has never exerted action to that effect, due to inertia.

As can be seen, such a policy can do much to raise levels of current and future happiness; and it achieves this by taking no freedom from the individual, and evidently not based on any of the heuristics we have previously identified.

Conclusion

To conclude, we can say that policies based on optimisation heuristics tend to be sub-optimal. The particular danger they pose is that they are liable to be interpreted to be an end in themselves, rather than simply as an avenue to ever greater happiness, which should always be the intended aim. Examples of such policies included the ban on the sale and distribution of drugs such as marijuana, Pigovian taxes, and compulsory purchasing.

However, though we pointed out that a number of paternalistic policies were motivated by heuristics, and were in this manner sub-optimal, we were careful to point out that paternalistic policies, in and of themselves, could be optimal policies, which contribute to the sum of human happiness both now and in the future. Examples here included libertarian paternalism and the ban on a market for organs.

References

Foster, P., 2011. Chinese Teen Sells His Kidney For An iPad 2. *The Telegraph*. Available at: http://tgr.ph/ig8wbj [Accessed Jan 10, 2012]

Hillman, A.L., 2009. *Public Finance and Public Policy*. 2nd Ed. Cambridge: Cambridge University Press.

Sen, A.K., 1970. The Impossibility of a Paretian Liberal. *Journal of Political Economy,* 78, pp.152–57.

Thaler, R.H. and Sunstein, C.R., 2009. *Nudge.* London: Penguin Books

Viscusi, W.K., 1990. Do smokers underestimate risks? *Journal of Political Economy*, 98(12), pp. 53-69

When Senior Sophister Played the Ultimatum Game: Investigating the Theory of Rational Choice

Iain Snoddy, Harry Clancy, Niall Connolly, Graeme O'Meara, Charles Pearmund, Alan Reynolds, Ronan Sandford
Senior Sophister

A core assumption in many economic models is that humans are rational and seek only to maximise personal utility. In this paper, through the prism of the Ultimatum Game, Iain Snoddy et al examine this assumption and find that it breaks down among Senior Sophister students, coming to the conclusion that what we truly care for is perhaps slightly more complex.

Introduction

This paper attempts to determine whether Senior Sophister economics students adhere to the concept of the rational utility maximizer. To answer this question we played the Ultimatum Game with a subset of our Senior Sophister Economic Theory class. In the first section, literature regarding the Ultimatum Game is reviewed. The second section explains our experimental methodology and in the third section some selected results are presented. The final section shall contain some concluding remarks.

Background and Literature Review

The Ultimatum Game is just one example of an extensive game with perfect information. It is closely related to the Dictator Game and differs only in that it incorporates a sequential element; unlike the dictator game the second player is not passive but can respond to the action of player one. The most basic form of the Ultimatum Game is played out between two participants, player one and player two. Player one is given an initial sum of money and is told to divide this sum between himself and player two where player two may be anonymous or non-anonymous, depending on the structure of the game. On receiving this

proposal from player one, player two must decide whether to accept or reject the offer. The key feature of the ultimatum game is as follows: if player two accepts the offer, the sum is divided between the players according to the proposal of player one, but if player two rejects the offer both players receive nothing.

Given the sequential structure of the Ultimatum Game, the concept of a Nash Equilibrium is not robust and rather the concept of a subgame perfect equilibrium is used to determine the static equilibrium of the game. The subgame perfect equilibrium is defined as "a strategy profile with the property that in no subgame can any player i do better by choosing a strategy different from [it], given that every other player j adheres to" (Osborne, 2004:165). In the ultimatum game with indivisible units[1], there exist two perfect subgame equilibria. The first occurs where player two accepts all offers made by player one, and as a result player one offers 0. The second is found when player two accepts all offers from player one except 0 and as a result player one will offer the smallest possible amount to player two. If player two is acting rationally they would accept this amount no matter how small as to reject it would lead to a payoff of 0 (Thaler, 1988). These are the equilibrium results we would expect if both players adhere to the theory of rational choice.
 However, experimental evidence finds that the subgame perfect equilibrium is seldom found in practice. The first experimental study of the Ultimatum Game was conducted at the University of Cologne by Güth, Schmittberger and Schwarze (1982). The game they conducted was a relatively simple Ultimatum Game played with a small sample of university students (n=21) with various sums between 4 and 10 DM given to player 1 for division. The authors found that the mean offer of player one amounted to 37% of the initial sum while the modal division was an equal part split of 50%.

A more rigorous experiment was conducted by Kahneman, Knetsch and Thaler (1986). In this paper students not only played a simple version of the Ultimatum Game, but they were also asked which allocations of the sum of $10 they would consider a fair offer, with the value changing in 50c increments. A sample of 115 students was used in this game with the authors taking careful steps to ensure all participants fully understood the rules of the game. They found that the mean offer of player one ranged from $4.21 to $4.76, while the mean minimum acceptable offer ranged from $2.00 to $2.59 (1986). The authors extended the Ultimatum Game in a subsequent experiment. In this experiment if player 2 rejected the initial offer she would become the proposer in another game. Player 2 would then be given two choices; either allocate $10 equally between herself and a fair allocator (some-

---

1 This means the initial sum given to player one is not infinitely divisible. For example, the smallest possible unit of €10 is €0.01. This sum cannot be further divided.

one who offered $10 out of $20 in round 1) or divide $12 equally between herself and to an unfair allocator in round 1. In this round 74% of participants opted to divide $10 with the fair allocator at a personal cost of $1, in other words participants were willing to 'pay to punish' unfairness (1986).

Further experiments have made great alterations to the Ultimatum Game but have found much the same result: people typically do not act in accordance with the theory of rational choice. An experiment conducted by Ochs and Roth (1989) extended the simple Ultimatum Game to be played over multiple stages. If player two rejected the offer made by player one, then the game would be replayed with player 2 acting as the proposer, but with the initial sum to be allocated diminished. Again they found that participants failed to act in the most rational manner as player 2 often made 'disadvantageous counter-proposals' where they kept less in round two than they were offered by player one in the first round (1989). We shall now turn to an explanation of the methodology used to conduct our in-class experiment of the Ultimatum Game.

Methodology

In our experimental analysis we conducted 4 separate rounds of the Ultimatum Game, with slight alterations in the game structure between each round.

Round 1: In round one each player was given a playing slip and told he had an initial sum of €10 to divide between himself and player two, where player two would be randomly assigned. The anonymity of the game was emphasised. All offers were then collected and randomly distributed back to the class so that each player both made an offer as player one, and then received an offer as player two. As player two, the participants now had the option to accept or reject the offer. At this point the rules of the Ultimatum Game were strongly emphasised: if player two accepts the offer, the sum is divided between the players according to the proposal of player one, but if player two rejects the offer both players receive nothing.

Round 2: The game played in round 2 was exactly the same as that conducted in round 1. However, between rounds the results of round 1 were put into a Microsoft excel spreadsheet and shown to the class. The spreadsheet showed the average offer made, the numbers of accepted and rejected offers, the frequency of offers made and the conditional rejection frequency amongst other information. A game was also selected at random and a real payoff (delicious chocolate eggs) was made to the two participants. All of this was done to show players the results of their actions and to increase their knowledge of the game's structure.

Round 3: This round was played non-anonymously. Player one was

given a play slip and told to write down their name, assigned number and favourite colour. They then has to pass their form to the player on their immediate left; this player would now act as player one and divide the initial sum of €10. Player one in this case now knows the name of the player to whom they are making an offer, and some trivial information regarding the player. Player one was instructed to make an offer and also to write their name and assigned number on the play slip. The play slips were now passed back to player two who decided whether to accept or reject the offer. Note that player two now also knows the name of the person making the offer.

Round 4: The game in this round was not only played non-anonymously but was also conducted in public. In this case we paired players and assigned them as player one and player two. Those playing the game were then instructed to stand and face each other and play the game as before, with player one instructed to make an offer and player two deciding whether to accept or reject the offer.

It is important to note that, before and during each round, the rules of the game were stressed to participants, namely that if player two accepts the offer, the sum is divided between the players according to the proposal of player one, but if player two rejects the offer both players receive nothing.

Experimental Results

General Results

In looking at the initial results presented in Table 1 below it is already clear that the participants in our experiment did not act in accordance with the theory of rational choice. The most common offer in all rounds was a 50-50 split of the €10. Furthermore, the lowest average offer in any round was €3.80, made in rounds 1 and 2; this is a far cry from the €1 offer we would expect to see, given completely rational actors. We would also expect to see all positive offers accepted by player 2, as a rejected offer leads to a zero payoff. However, our experiment also fails to observe this result: the lowest average rejected offer in any round was €2.10, suggesting that responders typically rejected small but positive offers.

Observing the pattern of the results presented in Table 1 is also interesting. It suggests that not only are all offers significantly greater than the rational optimum, but that offers, on average, increased in the non-anonymous game and increased even further in the publicly played non-anonymous game in round 4. In fact in round 4 the average offer was greater than a perfectly equal 50-50 split. The average rejected offer also increased in the non-anonymous game. This result is interesting as it suggests that when some detail is known about the proposer, player 2 expects the offer to be more eq-

uitable. The large difference between round 3 and round 4 further suggests that players expect individuals whom they have seen and with whom they have interacted verbally to act much more generously and are indeed willing to 'punish' those individuals more severely through rejecting the offer.

Table 1: Overview of Results

	Round 1	Round 2	Round 3	Round 4
Average Offer	3.8	3.8	4.8	5.1
Modal Offer	5.0	5.0	5.0	5.0
Average Rejected Offer	2.3	2.1	2.6	3.8

Figure 1 presents the frequency of offers made in each round. What is most noticeable is the relative lack of very low offers and of very high offers. When the shift is made from anonymous to non-anonymous there is a huge jump in what we may call 'ludicrously irrational' offers, those greater than a 50-50 payoff of €5. In all rounds there is sufficient proof to suggest that our Senior Sophister economics class are motivated by factors other than rational utility maximisation as only 10% of students chose the rational option of 1 in rounds 1 and 3, and only 5% acted rationally in rounds 2 and 4. Turning to the actions of player 2, Table 2 shows the frequency with which offers were rejected in each round. Typically low offers were rejected, with offers of 1 being rejected in all rounds, and 2 being rejected in all rounds except round 1. Surprisingly, all offers of 0 were accepted in round 3, although this accounts for only one observation. Conversely, few 50-50 allocations were rejected with only 14% rejecting 50-50 divisions in round 3, while such divisions were accepted on all occasions in the remaining rounds. The rather high rejection rate observed for an offer of 4 in all rounds is interesting, and suggests that many participants consider an offer of 5 as the sole equitable possibility.

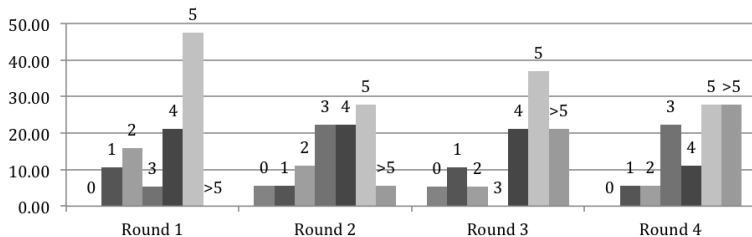Table 2: Conditional Rejection Frequencies

Offer	Round 1	Round 2	Round 3	Round 4
0	*	100	0	*
1	100	100	100	100
2	67	100	100	100
3	0	50	*	50
4	50	25	25	50

| 5 | 0 | 0 | 14 | 0 |
| >5 | * | 0 | 0 | 20 |

*No offers were made at this value

Diagram 1: Offer Frequencies

Given these descriptive results, it is clear thus far that Senior Sophister economics students are not representative of Homo Economicus; they fail to adhere to the theory of rational choice. In the case of player 1 (the proposer), offers are significantly greater than the optimal offer of 1. Similar conclusions can be made about participants in the role of player 2 (the responder) as non-zero offers are frequently rejected. As such it must be concluded that these participants are concerned with factors other than utility maximisation. One such factor may be distributional fairness (Thaler, 1988). In particular individuals may feel it is unfair to keep most of a sum that they did nothing to earn. Similarly students may feel they ought to punish individuals who keep a large portion of this sum for themselves. Asking proposers to allocate a sum for which they have exerted effort could therefore greatly change the results of the game. A further explanation is that proposers are neither altruistic nor concerned about fairness, but rather they have a 'fear of rejection' and so make offers that are typically much higher than what is rationally optimal. As explained by Thaler (1988), a proposer may fear rejection if she does not believe in the rationality of other participants: she does not believe that other participants will realise that accepting all non-zero offers is their optimal action.

Selected Results

By using data gathered on participants we were able to make some further statistical observations. As is shown in Figure 2 there is very little variation in the average offer made by different gender groups. Both groupings offered results much higher than would be expected under assumptions of perfect rationality. Figure 3, however, shows that on average, males were more likely

to accept a low offer than were females, suggesting perhaps that females have a heightened sense of fairness.

Diagram 2: Average Offer Made by Gender

Diagram 3: Average Offer Rejected by Gender
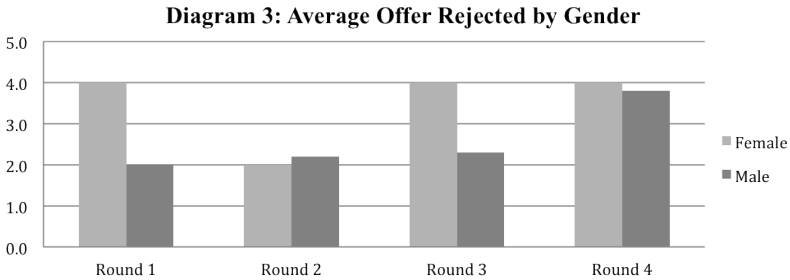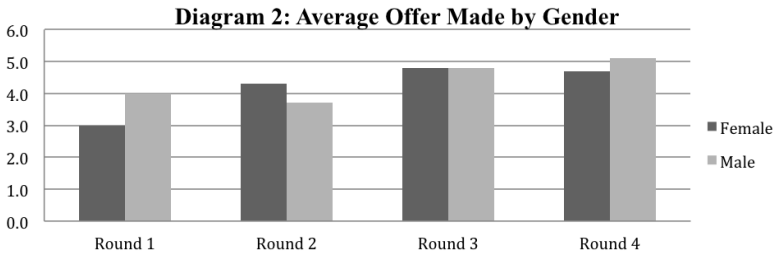
Looking at Figures 4 and 5, we can also observe the difference between the actions of students studying single honours economics, and those taking economics as part of a joint degree. Perhaps surprisingly, single honours economics students offered higher amounts on average in each round. In could therefore be concluded that joint honours economics students are more likely to act as rational optimisers than are single honours students[2]. Figure 5, however, calls this result into question as joint honours students reject higher offers than do single honours students. Nevertheless, the variation in the actions of both groups is very slight, and as a result no firm conclusions can be drawn. What is most perplexing is the observed trend reversal found in round 4 in both diagrams.

---

2 One cannot but wonder if this greater rationality influenced their choice of joint rather than single honours economics.

Diagram 4: Average Offer Made by Degree Programme
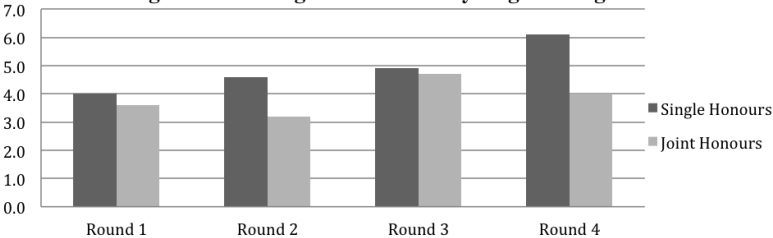
Diagram 5: Average Offer Rejected by Degree Programme
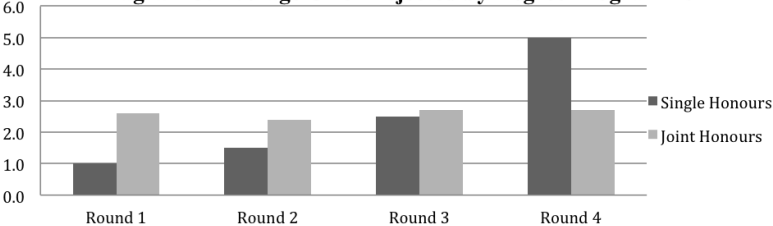
Conclusion

In conducting a relatively simple experiment using the Ultimatum Game we have found that our Senior Sophister class of economic students do not readily adhere to the theory of rational choice. This result is supported by various other experiments using the Ultimatum game, found in the economic literature. Our finding that Senior Sophister economic students care about factors other than rational utility maximisation will come as a surprise to some, and a relief to many.

References

Güth,W., Schmittberger, R. and Schwarze, B., 1982. An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior and Organization, 3:367-388.

Kahneman, D., Knetsch, J.L. and Thaler, R.H., 1986. Fairness and the Assumptions of Economics. The Journal of Business, 59:4:285-300.

Ochs, J. and Roth, A.E., 1989. An Experimental Study of Sequential Bargaining. American Economic Review, 79:3:355-384.

Osborne, M.J., 2004. An Introduction to Game Theory. Oxford: Oxford University Press.

Thaler, R.H., 1988. Anomalies: The Ultimatum Game. Journal Of Economic Perspectives. 2:4:195-206.

AN ECONOMETRIC INVESTIGATION INTO THE NATURE OF THE RELATIONSHIP BETWEEN UNEMPLOYMENT AND SUICIDE

BULAT KUBEYEV
Senior Sophister

With this comprehensive econometric study, Bulat Kubeyev investigates the effect of unemployment levels on suicide rates at different levels of GDP/capita, using data from 88 countries over the last decade. Interestingly, it is found that a country's income per person has a significant effect on the direction of the relationship between suicide rates and unemployment.

Introduction
'You take my life, when you do take the means whereby I live.'
-The Merchant of Venice (4.1.371-72)

Dorland's Illustrated Medical Dictionary defines suicide as 'the act of intentional taking of one's own life'. It can be triggered by a wide range of social and economic factors, and while the study of suicide sometimes presents a challenge to evaluate different aspects of societal behavior (e.g. bullying, loneliness, etc.), suicide rates, like most economic situations, can be proxied by unemployment rates. The effects of unemployment on suicide have undergone some major quantitative analyses? for Sweden, Germany and the United States, while mostly qualitative analysis has been performed for Ireland.

But does unemployment really affect suicide rates? Previous research in this area has not provided a definite answer, but most macroeconomists would agree that unemployment results in a loss of social and financial stability. This generally leads to severe frustration or even depression, because 'unemployed people not only lose materially, they also potentially lose access to social networks, self-esteem, self-confidence, a scheduled life structure, a sense of identity and possibly a purpose for their lives' (Neumayer, 2004,

p.1038). Naturally, depressed persons are more prone to suicide than those with no underlying mental health problems.

A factor that was previously overlooked in most research papers is the link between unemployment and income. However, though the effect of unemployment on suicide rates seems to be intuitively quite straightforward, when it is interacted with personal income the results become somewhat ambiguous. This paper is examining whether the effect that unemployment has on suicide rates varies consistently with the level of real GDP per capita.

It is worth mentioning at this stage that the study presented in this paper is not the first of its kind and hopefully not the last. Still, one of its unique features is its geographical scale, while its main contribution is that it allows us to understand the effects of unemployment on suicide in countries with various income levels, which are not necessarily similar[1].

Background/ Motivation

The idea that there is an association between unemployment and suicide is hardly new. Much literature has been written on the subject and it can certainly be grouped by methods of analysis, countries of interest and even by the hypothesis being tested. Nevertheless, practically all significant research in this area has one common feature – it was conducted using either time-series or cross-sectional data analysis. This research proves to be invaluable as it is generally thorough and definitive, but it is mostly conflicting. For instance, Gerdtham and Johannesson (2003) use time series analysis for Sweden and end up with the positive relationship between unemployment and suicide rates. However, when Neumayer (2004) applies the same method of time series analysis to Germany, this relationship is negative[2].

Generally speaking, most recent papers, such as the one by Andres (2005), use ordinary least squares (OLS) as a method of analysis and then go on to report the positive and significant effect that unemployment has on aggregate suicide rates, but in the end fail to find any significant association between income and suicide. On the other hand, Chuang and Huang (1997) find this relationship to be negative, if only across Taiwan.

Similarly, at the initial stages of research, it was assumed that this

1 The idea of "similar income levels" is abstract in this context: EU and OECD member states all have different income levels, but they are also all developed and can be thought of as being "similar". Hence, including third world countries and CIS member states can improve the transparency of the results.

2 It may be argued that the discrepancy in the results of similar analyses for Sweden and Germany is caused by the differences between The Nordic Model and The Rhineland Model respectively, but such an argument would be mere speculation. Other factors (e.g. amount of sun light per year) affect one's mental health just as much.

paper would focus solely on Ireland (including the Republic of Ireland and Northern Ireland), partially due to the scope of the information that was available in the public domain. A national study conducted in 2001 on behalf of the Irish Health Board provides a fully qualitative assessment of suicide trends in Ireland complete with the profiles and social standings of the deceased. However, with a population of approximately 6.2 million people and only 32 counties, it soon became clear that Ireland was a poor candidate for this type of study. Furthermore, even if a time-series analysis were to be conducted, collecting annual figures related to suicide (complete with the profiles of the deceased) would present a significant challenge, as the Republic's national police force, an Garda Síochána, can only disclose details with the permission of the families of the deceased.

It immediately becomes evident from the existing research that once the effects of unemployment or income have been estimated, they largely depend on both the country and the time period in question. Furthermore, one can expect a certain degree of interaction between these variables, which results in biased estimators. According to Wooldridge (2003), ignoring a significant interaction term usually produces higher variances and thus, reduces efficiency of the regression model. One possible solution to this problem is to introduce an interaction term between two main control variables and see what effect it has on suicide rates, i.e. to include two effects at the same time. It makes sense to further control for certain demographic and social factors, which might influence aggregate suicide rates.

In general, this paper will try to estimate a fixed effects model, by way of panel-data analysis, for 88 countries with different development levels (including member states of the Commonwealth of Nations, EU, OECD, UN-ASUR, CIS and EFTA) during ten consecutive years (2000-2009). The decision to use panel-data analysis, instead of a cross-section or a time-series, was made with a view to controlling for potential bias associated with omitting a variable or failing to include a time invariant factor that is unique to a particular country (e.g. low levels of yearly sunlight in northern countries quite possibly affects suicide rates and yet remains unobserved).

Empirical Approach

The objective of this paper is to find out how unemployment affects suicide rates depending on income. To do so, it analyzes the relationship between suicide rates and a number of social and economic factors in 88 countries during 2000-2009. Consequently, it compensates for the problem of omitted variable bias.

72

However, before constructing an empirical model, it is vital to make a decision as to which variables should be included and which should be omitted altogether from the final equation. The decision process is mostly concerned with the independent variables, also known as control variables, since the only possible choice for the dependent variable is the aggregate suicide rate, "$srate_{i,t}$".

As regards the explanatory variables, it is logical to include unemployment rates "$unemp_{i,t}$" and real GDP per capita "$gdppc_{i,t}$", so that we are able to test our hypothesis. Furthermore, as suggested by Wooldridge (2003), an interaction term "$interact_{i,t}$" is introduced in the equation to examine whether the relationship between unemployment and suicide varies for countries with different income levels. This term is generated using STATA 11 and thus, is not automatically present in the dataset.

Now, even though unemployment rates are themselves sufficient to control for the fluctuations in the economy, it is advised to include another variable that serves the same purpose. According to Andres (2005), this variable is the real GDP growth rate, "$grgdp_{i,t}$".

The remaining three variables are introduced in order to control for social factors. These variables tend to change over time and are thought to have a direct impact on the aggregate suicide rates. Using the idea put forward by Chuang and Huang (1997) with reference to Durkheim (1897) that societal suicide rates are influenced by social integration and social regulation, both fertility rate "$fertility_{i,t}$" and family formation rate "$ffrate_{i,t}$" have been included for each country. Finally, alcohol consumption has been previously used as a control variable by Andres (2005). Nevertheless, some experts remain sceptical on the subject of alcohol's relationship to depression, and consequently suicide. The decision to include "$alcohol_{i,t}$" was based on the idea put forward by Chaloupka, Grossman and Saffer (2002), that alcohol consumption causes depression and not vice versa.

Since the study presented in this paper is conducted using panel-data analysis, it is vital to include a full set of year dummies (and to leave one out to compensate for the intercept) in order to control for aggregate time effects. Thus, the baseline equation to be estimated is:

$$srate_{i,t} = \beta_0 + \beta_1 unemp_{i,t} + \beta 2 gdppc_{i,t} + \beta IN interact_{i,t} + Y_{i,t}\lambda + \alpha_i + \gamma_t + \varepsilon_{i,t}$$

(1)

Where $Y_{i,t}$ is a vector, which denotes the following control variables: "$grgdp_{i,t}$", "$fertility_{i,t}$", "$ffrate_{i,t}$" and "$alcohol_{i,t}$". Subscripts i and t in equation (1) index

a country and a time period respectively; β_0, β_1, β_2, β_{IN} are estimated coefficients; λ denotes an estimated coefficient vector; α_i and γ_t are dummy variables whose purpose is to capture unobserved country and time specific effects respectively; $\varepsilon_{i,t}$ is an error term (and we further assume that $\varepsilon_{i,t}$ is independently and identically distributed with mean 0 and variance σ^2 for all i and t).

It is worth noting that if the interaction term is significant, then the partial effect of "$unemp_{i,t}$" on "$srate_{i,t}$" will depend on the level of "$gdppc_{i,t}$" (per Wooldridge, 2003:190):

$$\Delta E(srate_{i,t})\Delta unemp_{i,t} = \beta_1 + \beta_{IN} gdppc_{i,t} \qquad (2)$$

Consequently, all else being equal, the marginal effect of unemployment varies with different levels of real GDP per capita. Similarly, the marginal effect of real GDP per capita changes with various unemployment rates when everything else is held constant:

$$\Delta E(srate_{i,t})\Delta gdppc_{i,t} = \beta_2 + \beta_{IN} unemp_{i,t} \qquad (3)$$

The primary focus is on the relationship described by equation (2).

Dataset

The dataset, comprised of annual suicide rates and socio-economic figures for 88 countries during 2000-2009, has been compiled from the following three sources: WHO Mental Health Data 2011, World Bank Database 2011 and the United Nations Population Division 2011.

Although this dataset is quite large geographically, it has been carefully selected. It may be argued that by including countries with different development levels, one compromises both data transparency and quality of indicators. However, all 88 countries in the dataset have their own well-established national statistics agencies, which have been consistently supplying quality data to their international counterparts. Countries with unique social features have been excluded in order to achieve "truer" coefficients (e.g. Maghreb states consume very little alcohol compared to the rest of the world – they were omitted).

All descriptive statistics for the variables used in equation (1) are listed in Table 1, along with their respective meanings (generated using STATA 11).

Table 1: Variables and general statistics

Variable	Meaning	Observ.	Mean	St. Dev.
srate$_{i,t}$	Suicide rate (per 100,000)	880	11.07	8.89
unemp$_{i,t}$	Unemployment, total (% of total labour force)	880	9.28	6.20
gdppc$_{i,t}$	GDP per capita (current US $)	880	13924.37	16893.88
grgdp$_{i,t}$	Growth rate of GDP per capita (%)	880	3.22	4.65
alcohol$_{i,t}$	Adult consumption of alcohol, per capita (litres)	880	7.06	3.85
ffrate$_{i,t}$	Family formation rate (per 1000)	880	5.51	1.82
fertility$_{i,t}$	Fertility rate, total (births per woman)	880	2.05	0.72

Sources: WHO 2011; World Bank 2011; UN 2011.

General trends associated with suicide rates in four major economic and political entities are presented in Figure 1.

Figure 1: Suicide rates (per 100,000) during 2000-2009
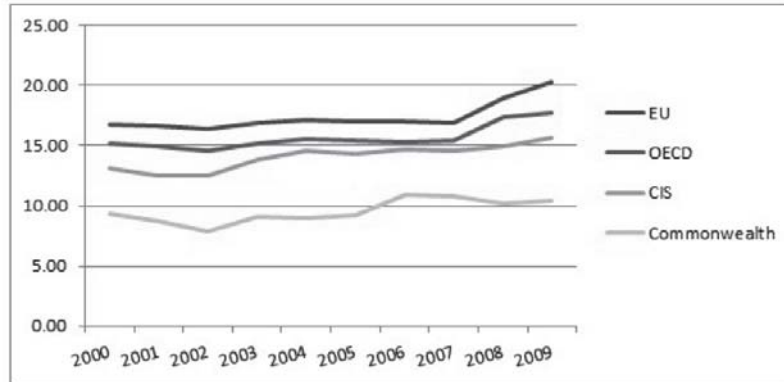
Dataset Issues

In his paper, Andres (2005) states:

> "The effect of socioeconomic predictors of suicide differs across age groups. Thus, the use of age-specific suicide data definitely does make sense, and as demonstrated in this study, it may result in helpful guidance for health policy makers"
>
> (Andres, 2005, p.449)

Due to the lack of age- and gender-specific statistics, the dataset used here is not as comprehensive as was initially intended. However, it is not the objective of this paper to investigate the effects of unemployment on suicide rates for different age groups and genders. On the contrary, it aims to study the aggregate effect of unemployment on suicide rates with respect to different income levels. Therefore, omitting age-specific statistics does not constitute a major drawback in this type of research.

Empirical Results

$$srate_{i,t} = 17.032 - 0.331unemp_{i,t} + 0.0007gdppc_{i,t} + 0.0008interact_{i,t}$$
$$(1.926) \quad (0.055) \quad\quad (0.00002) \quad\quad (0.0004)$$

$$- 0.187grgdp_{i,t} + 0.399alcohol_{i,t} - 0.472\,ffrate_{i,t} - 2.435fertility_{i,t}$$
$$(0.026) \quad\quad (0.089) \quad\quad (0.181) \quad\quad (0.807)$$

$$N = 880 \quad R\text{-sq} = 0.5455 \text{ (within)}$$

The intercept term is 17.032, which represents the suicide rate if all independent variables were zero simultaneously – not a meaningful situation in itself. The directions of the relationships, indicated by the signs of the coefficients of the independent variables and the intercept, are largely as anticipated. Specifically, all else being equal, a one percent increase in alcohol consumption corresponds to a 0.399 percent increase in suicide rate. On the other hand, a one percent increase in family formation rate corresponds to a 0.472 percent decrease in suicide rate, ceteris paribus. Similarly, holding all else constant, a one percent increase in fertility rate results in a 2.435 percent drop in suicide rate. The coefficient on economic growth, although significant by intuition, is statistically insignificant with a p-value of 0.346. This is an odd occurrence, considering that it was previously used in the study of suicide by other re-

searchers.

Most importantly, in the presence of both "$unemp_{i,t}$" and "$interact_{i,t}$", the negative impact of unemployment on suicide is reduced by income, since β1<0 while βIN>0. On the other hand, in the presence of "$gdppc_{i,t}$" and "$interact_{i,t}$", the effect of income on suicide rates does not change. The beta value for the interaction term is statistically different from zero at the 1% significance level. Hence, the partial effect of "$unemp_{i,t}$" on "$srate_{i,t}$" depends on the level of "$gdppc_{i,t}$". Consequently, it was decided to run the original regression again, this time including a dummy variable that takes on the value of one for the countries with high levels of real GDP per capita and zero otherwise. Note that "$grgdp_{i,t}$" was omitted. The following regression was obtained:

$$srate_{i,t} = 16.534 - 0.364unemp_{i,t} + 0.0006gdppc_{i,t} + 0.001interact_{i,t}$$
$$(1.973) \quad (0.058) \quad\quad (0.00002) \quad\quad (0.00038)$$

$$+ 0.386alcohol_{i,t} - 0.49\ ffrate_{i,t} - 2.341fertility_{i,t}$$
$$(0.09) \quad\quad (0.182) \quad\quad (0.811)$$

$$N = 880 \quad R\text{-sq} = 0.6472 \text{ (within)}$$

After removing the interaction term, the resulting regression is:

$$srate_{i,t} = 16.073 + 0.358unemp_{i,t} + 0.0004gdppc_{i,t} + 0.274alcohol_{i,t}$$
$$(1.964) \quad (0.053) \quad\quad (0.00001) \quad\quad (0.0901)$$

$$-0.470ffrate_{i,t} - 2.369fertility_{i,t}$$
$$(0.182) \quad\quad (0.812)$$

$$N = 880 \quad R\text{-sq} = 0.6420 \text{ (within)}$$

Therefore, in richer countries an increase of one percent in unemployment rate corresponds to a ceteris paribus 0.36 percent rise in suicide rate.

The R-squared value for the regression with a dummy variable and an omitted variable is greater than its value for the original regression. Arguably, this indicates that the latter model has better predictive powers than the former.

However, the most important result obtained was that for countries with higher income levels the effect of unemployment on suicide was indeed positive and significant. Conversely, for countries with lower incomes, this

effect is negative.[3]

Possible Extensions

The results of the empirical analysis seem to be conclusive. Nevertheless, a further study of the effect that each control variable has on suicide rates is certainly possible. Naturally, in order to take full advantage of the panel-data analysis, the proposed study should cover a much longer time period. Furthermore, it would be interesting to see if unemployment has a different effect on males vs. females, youth vs. senior citizens, service industry vs. manufacturing etc. Unfortunately, the feasibility of such a study is questionable due to the lack of necessary international data.

Summary/Conclusions

The study initially set out to find a link between unemployment and suicide using a number of other socio-economic variables. As such, the results were more or less anticipated, with a possible exception: evidence was found to suggest that economic growth does not help to explain suicide rates, even during shorter periods of time. This contradicts a couple of research papers, including those of Andres (2005) and Neumayer (2004). This contradiction possibly stems from the difference in approaches used across the board.

What was not expected is the way income level affects the relationship between unemployment and suicide. Investigation suggests that the notion of suicide being positively dependent on unemployment is not exactly true. According to the analysis, unemployment has a positive significant effect on suicide rates in richer countries, while the opposite is true for poorer countries. It only goes to highlight that "what you've never had you never miss". Losing employment in the richer countries is much more distressing than in poorer countries.

The fact that suicide rates are sensitive to income levels contradicts the study of 15 European countries by Andres (2005). However, as was mentioned before, the effect of income on suicide rates would possibly be insignificant when the cross-section of the panel data consists of countries with relatively 'similar' income levels.

Overall, this paper should act as a guide for the researcher wishing to undertake a more comprehensive study of different factors affecting suicide rates (possibly when data are more abundant) and as a point of reference for the policy maker wishing to implement new health policies.

---

3 By trial and error, the threshold between high income and low income was found to be around US$ 7500.

References

Andres, A. R., 2005. Income inequality, unemployment, and suicide: A panel data analysis of 15 European countries. *Applied Economics*, 37, pp.440–450.

Chaloupka, F. J., Grossman, M., and Saffer, H., 2002. The Effects of Price on Alcohol Consumption and Alcohol-Related Problems. *Alcohol Research & Health*, 26(1), pp.22–34.

Chuang, H-L. and Huang, W-C., 1997. Economic and social correlates of regional suicide rates: A pooled cross-section and time-series analysis. *Journal of Socio-Economics*, 26, pp.276–290.

Department of Public Health on Behalf of the Chief Executive Officers of the Health Board, 2001. Suicide in Ireland. A National Study. Dublin: Department of Public Health.

Dorland's Illustrated Medical Dictionary, 32nd ed. 2011. Philadelphia, Pa.: Elsevier/Saunders, c2012.

Gerdtham, U-G. and Johannesson, M., 2003. A note on the effect of unemployment on mortality. *Journal of Health Economics*, 22, pp.500–520.

Neumayer, E., 2004. Recessions lower (some) mortality rates: Evidence from Germany. *Social Science and Medicine*, 58, pp.1035–1050.

United Nations, Department of Economic and Social Affairs: Population Division, Population Estimates and Projections Section Database, 2011. [Online] Available at: http://www.un.org/esa/population/ [Accessed 23 February 2011].

United Nations, Department of Economic and Social Affairs: Population Division, 2011. World Fertility and Marriage Database 2008 (incl. estimates). [CD-ROM] New York: UN.

World Bank Database, 2011. [Online] Available at: http://databank.worldbank.org/ddp/home.do [Accessed 23 February 2011].

Wooldridge, J. M., 2003. *Introductory econometrics: A modern approach.* 2nd

ed. USA: Thomson.

World Health Organization, Department of Mental Health and Substance Abuse, 2011. Global Status Report on Alcohol and Health 2010: Country profiles. [CD-ROM] Geneva: WHO.

World Health Organization, Mental Health Statistics, 2011. [Online] Available at: http://www.who.int/topics/mental_health/en/ [Accessed 23 February 2011].

An Investigation into the Proportion of Women in National Parliament in a Cross-Country Sample

Debbie Blair
Junior Sophister

In national parliaments worldwide, under-representation of women has been a persistent problem. In this rigorous econometric analysis, Debbie Blair seeks to explain the proportion of women in a national parliament, using a wide variety of socio-economic, political and cultural variables. The study draws some interesting conclusions for policy makers looking to address this gender imbalance.

Introduction

'In politics, if you want anything said ask a man, if you want anything done ask a woman.'

The above quote can be attributed to Margaret Thatcher, British politician and Prime Minister from 1979 to 1990. December 2011 saw the release of The Iron Lady, a film which portrayed her influential role in politics. Since the start of the 20th century there have been immense social changes in women's rights and roles, with the aim of achieving equality of the sexes. however,, despite the plethora of advances that have been made women still remain underrepresented in national parliaments. Worldwide, women constitute only 19.9% of parliaments, despite composing of just over 50% of the population. This paper intends to examine the factors that may be responsible for determining the proportion of women in parliament. It will focus on three main variable groupings namely socio-economic, political and cultural factors. This research aims to shed light on their respective contributions, in order to inform policies as to how to equalise political representation and thus create a more equitable society.

Background

Although women constitute just over half the population, there is a significantly smaller percentage of women in national parliaments. This topic has received much attention in the literature, with researchers attempting to explain both the existance of this discrepancy and cross country variation. Past research has prominently focused on the socio-economic and institutional factors that may impact the engagement of women in parliament. however,, recent papers have highlighted the importance of cultural factors that act as barriers to female involvement (Ruedin, 2010).

Often, socio-economic factors are suggested as contributing factors that affect the supply of qualified women. Educational attainment, labour force participation and GDP per capita are the prime candidates for influence. Previous research has found mixed results for the role of education in predicting parliamentary participation (Rule, 1987). A woman may achieve high levels of education but be reluctant to pursue a role in public life, such as in parliament. Higher female labour force participation rates have been more commonly associated with higher participation in parliament (Rule, 1987). It is also likely that GDP per capita may play a role; countries with higher levels of per capita income typically more equitable societies.

The type of electoral system and the use of quotas are frequently cited as explanatory factors. Electoral systems are typically broken down into four major subtypes; Proportional Representation (PR), Majority Representation (MR), Mixed and Other. PR systems are generally associated with a higher proportion of women in parliament (Tripp & Kang, 2008). This is likely to be due to representative nature of the PR system. Gender quotas are an area of much debate and controversy. Many of their opponents believe that positive discrimination is itself inequitable and may prevent the election of highly qualified candidates, who just happen to be male. Despite this and other criticisms, a number of countries have adopted various forms of gender quota. It remains to be shown whether one type of quota is more effective than the other. however,, studies report that quotas are associated with higher female representation (Wangnerud, 2009).

Nevertheless, models which focus solely on the above categories lack explanatory power. It seems that there are substantial cultural barriers which hinder female participation. Cultural variables are hard to measure, and research typically uses proxy variables such as region and religion (Inglehart Norris, 2003) to capture the effects of cultural and attitudinal differences. Recent work by Reudin (2010) found that more direct measures of attitudes towards women held higher predictive power. However, there remain diffi-

culties in obtaining reliable direct measures and those that are available are limited to a small number of countries.

This research attempts to account for the proportion of women in parliament, as predicted by the factors that have been proposed in the literature. It hypothesises that a combination of socio-economic, political and cultural factors is likely to best explain the variation. It builds on prior research by attempting to identify if one grouping of variables makes an overarching contribution. The researcher tentatively suggests that cultural variables are likely to hold substantial explanatory power. Despite recent advancements in socio-economic status of women and the acknowledgment of the importance of representative political structures, there has not been a worldwide increase in number of women in parliament; this suggests that significant cultural and psychological barriers may be responsible for low representation.

Early research focused primarily on OECD countries but modern research includes worldwide statistics. This research focuses on a worldwide sample of countries at one point in time. While it may be useful to see the trends across time, this research has been restricted to cross-sectional data in order to avoid confounds such as years since suffrage, changes in electoral systems, and adoption of gender quotas at different times.

Empirical Approach

This study uses Ordinary Least Squares estimation in order to investigate the determinants of the proportion of women in national parliament as of 2009 in a cross-sectional sample of countries. The model will estimate equation 3.1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \varepsilon$$

(3.1)

Dependent Variable

Y (Proportion of Women in Parliament) : The dependent variable of interest is the proportion of women in national parliament as of 2009.

Independent Variables

Socio-economic Variables

X_1 (gdppc): The first explanatory variable is the level of Gross Domestic Product per capita in thousands of dollars, adjusted for purchasing power using 2005 as the base year. It is hypothesised that this will have a positive effect on the proportion of women in parliament.

X_2 (Lffm) : The second independent variable is the female labour force participation rate. This represents the proportion of women aged 15 and older that is active in the labour force. It is hypothesised to have a positive effect.

X_3 (Total3ed) : This third explanatory variable refers to the percentage of the total population of the same age bracket participating in third level education. This variable serves as a benchmark for the education level of the total population. It is hypothesised to be positively related to the dependent variable.

X_4 (Fem3ed) : This variable represents the proportion of the gross female population that is enrolled in third level education. It is expected to have a positive impact on involvement in parliament.

X_5 (FemPop): This independent variable is the last of the five socio-economic variables included in this analysis. It represents the percentage of the population that is female. It is expected to have a slight positive effect on participation rates.

Political Variables

X_6 (quota) : This is a dummy variable indicating the presence of a gender quota. A value of one was assigned if a gender quota was present in the electoral process and a value of zero was assigned if there was no quota. It is hypothesised that presence of a gender quota would have a positive effect on the dependent variable. The coefficient on this dummy is denoted α_1.

X_7 (electsys) :This variable refers to the type of electoral system, being classified into one of four subtypes (MR, PR, Mixed or Other). In the regression model analysed, three dummies were included for this variable in order to capture the individual effect of each electoral system. It is believed that using the PR system will have a positive effect on the proportion of women in a national parliament. The coefficients are denoted as follows: δ_1 (MR), δ_2 (PR), δ_3 (Mixed) and δ_4 (Other) .

Cultural Variables

X_8 (religion) : This is a dummy variable indicating the predominant religion in the country. It was subdivided into three classifications; Catholic, Muslim and Other. It is believed that both
Catholic and Muslim will have a negative impact on proportion of women in parliament. The coefficients on these are denoted γ_1 (Catholic), γ_2 (Muslim) and γ_3 (Misc) respectively.

X_9 (region) : This final explanatory variable consists of a number of dummy variables intending to capture the attitudes of people from particular regions. It is hypothesised that a country's being Nordic is likely to have a positive effect; a country's being Middle Eastern is expected to have a negative effect.

There were no further hypotheses concerning the other regions. The coefficients on each region are denoted as follows: ϕ_1 (AsiaPac), ϕ_2 (LatAmCar), ϕ_3 (Western), ϕ_4 (EastEur), ϕ_5 (Nordic), ϕ_6 (Africa) and ϕ_7 (Mide).

Error Term
ε (error term): The error term is believed to capture the unobservable factors that could not be controlled for, such as family history in politics and direct attitudes to women as politicians and any other variables. The error term is assumed to have zero mean, constant variance and to follow a normal distribution.

The full model, thus, is shown in regression equation 3.2.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$
$$+ \alpha_1 X_6$$
$$+ \delta_1 X_{71} + \delta_2 X_{71} + \delta_3 X_{71}$$
$$+ \gamma_1 X_{81} + \gamma_2 X_{82}$$
$$+ \phi_1 X_{91} + \phi_2 X_{92} + \phi_4 X_{94} + \phi_5 X_{95} + \phi_6 X_{96} + \phi_7 X_{97}$$
$$+\varepsilon$$

(3.2)

Data
The majority of data was sourced from the World Bank (2009). Recent data were available for all variables of interest for 2009 and this year was chosen for analysis. Data on the type of electoral system in each country were downloaded from the Inter-Parliamentary Union (2009) and a list of countries with gender quotas was obtained from the Global Database of Quotas for Women (2009). Religion statistics were acquired from the CIA World Factbook (2009). Each country was classiffied into one of three categories, depending on its predominant religion.
Only Muslim- and Catholic-dominated countries were considered independently, as it has been consistently shown that the preponderance of these religions influences the proportion of women in parliament. All other countries were grouped together as miscellaneous (miscrel). While this may mask other influential patterns from other faiths, this step was taken in order to preserve the degrees of freedom of the model. Countries were divided into regions according to United Nations Statistics Division country classifications. These are: Nordic (Nord), Western Europe and North America (Western), Eastern European (EastEur), Latin America and the Caribbean (LatAmCar), Middle

Eastern countries (Mideast), African countries (Africa) and countries in Asia and the Pacific (AsiaPc).

The initial sample consisted of 216 countries. However, countries that were missing values on 4 or more of the explanatory variables were excluded. The final sample consisted of 180 countries.

Each variable was given a specific label that will be referred to throughout this paper. These are shown in Table 4.1.

Table 4.1 Variable Labels

Summary statistics for the numerical variables were generated; the mean,

Variable Name	Label
Proportion of Women in Parliament	Parliament
GDP per capita PPP (constant 2005 international $)	gdppc
Labour participation rate, female (% of female population ages 15+)	lffm
School enrolment, tertiary (% gross)	Total3ed
School enrolment, tertiary, female (% gross)	Fem3ed
Population, female (% of total)	fmpop
Presence of Gender Quota	Quota
Type of Electoral System	Electsyst
Predominant Religion	Religion
Region	Region

standard deviation and the ranges, where appropriate, are displayed in Table 4.2.

Table 4.2: Descriptive Statistics

Measure	Parliament(%)	Gdppc (000s $)	Lffm (%)	Total3ed (%)
Mean	17.3767	11916.5	2.8871	40.62
Standard Deviation	10.9395	13088.8	14.7864	27.6799
Range	[0,56.3]	[298.297,73196.383]	[13.8, 91]	n/a

Measure	Fem3ed (%)	Fmpop(%)
Mean	45.67	49.986
Standard Deviation	32.90	3.02

Range	n/a	[24.96686, 54.0225]

Each nominal variable contained a number of classifications. The count of each subgroup is shown in Tables 4.3, 4.4, 4.5 and 4.6.

Table 4.3: Quota

	Count
Yes	103
No	78

Table 4.4: Type of Electoral System

	Count
MR	71
PR	68
Mixed	34
Other	8

Table 4.5: Religion

	Count
Catholic	48
Muslim	44
Misc	89

Table 4.6: Region

	Count
AsiaPac	34
LatAmCar	31
Western	22
EastEur	21
Nord	5
Africa	52
Mideast	14

Results

Diagnostic Checks

In order to determine the appropriate method of estimation, some diagnostic checks were performed. The main concern was that the data would exhibit heteroscedasticity, which would violate the assumptions of OLS. Scatterplots of the residuals revealed no obvious patterns and both the Breusch-Pagan and the White test (for heteroscedasticity) were insignificant for all of the regressions. Each regression was also tested for misspecification of the data using, the Ramsey RESET Test. This test was insignificant for all but regression one which involved only socio-economic variables. Multi-collinearity was assessed using the Variance Inflation Factor (VIF). Whilst there were somewhat high VIF values for the education variables, the mean VIF was still a reasonably low value. It was concluded that the Gauss-Markov assumptions were upheld. Thus OLS was considered appropriate and was applied in all of the analysis that follows.

Model

A number of models were analysed, in order to distinguish between the effects of various groupings of factors. The estimated coefficients, standard errors (in parentheses) and significance are shown in table 5.1.

The first regression model included only the socio-economic variables. In this model, GDP per capita, female labour force participation rate and percentage of population that are female were all significant at the 1% level. However, the practical effect of GDP is small, as the coefficient is negligible. In comparison, it would seem that an increase of 1% in the labour force participation rate leads to a .24% increase in the proportion of women in parliament. This is in line with previous literature (Rudein, 2010, Rule, 1987). An increase of 1% in the proportion of women in the population leads to an increase of 1.45% in parliament. However, the practical importance of this finding is questionable, as there is not much scope to increase the proportion of women in the population beyond its present level.

The second model included only political factors. Here, the presence of gender quota was significant at the 1% level. Having a quota increased the proportion of women in parliament by 6.2%. There was no affect found for any of the different types of electoral systems, contrary to the expectation that having proportional representation would positively influence the numbers of women in parliament.

The third regression model included only the factors being used as proxies for cultural factors. The Muslim variable was significant at the 10%

level (p=0.075). This indicated that, compared to the 'other religions' group, having a Muslim-dominated country led to a decrease of 3.77 percentage points in the proportion of women in parliament. The regional variables gave rise to three significant effects. The Nordic countries compared to Western countries had a strong significant positive effect (p=0.000). Nordic countries had a 21.07% increase in the proportion of women in parliament. In comparison, AsiaPac had a significant negative effect (p= 0.037) when compared to Western. Similarly, Middle Eastern countries had a negative effect at the 10% level (p=.099).

Table 5.1: Regression Results

Variable	1	2	3	4
Cons_	-72.976 (24.987)	10.742 (3.819)	20.726 (2.259)	-22.311 (32.760)
GDPpc	.000 *** (.000)			.000 (.000)
Lffm	.259*** (.077)			.257*** (.093)
Total 3ed	-.007 (.209)			.029 (.221)
Fem3ed	-.033 (.170)			-.059 (.182)
FMpop	1.454 ***			.655 (.626)
QuotaY		6.202 *** (1.632)		5.058*** (1.937)
Electmixed		2.332 (4.227)		-3.086 (4.975)
Electmr		.327 (1.988)		-2.924 (2.415)
Electpr		2.106 (1.349)		-.230 (1.567)
Muslim			-3.774* (2.108)	-.720 (2.916)
Catholic			1.574 (1.995)	-1.382 (2.478)
AsiaPac			-5.9822** (2.822)	-6.375 (4.818)
Nord			21.075*** (4.938)	14.008*** (4.992)

			-3.079	-8.362*
LatAmCar			(2.731)	(4.337)
EastEur			-3.021 (2.948)	-5.92 (3.889)
Africa			-3.190 (2.643)	-9.489* (4.984)
Mide			-6.464 (3.902)	-7.021 (5.678)
Observations	99	180	180	99
Adjusted R2	.2476	.1596	.2944	.4136

*** P<.01, **P<.05, * P<.1

However, all of the first three models had relatively low explanatory power. Each model does not control for the other possible factors that are likely to influence the proportion of women in parliament. This leads to biased estimates and violations of the assumptions of OLS. Thus, a model including all the proposed explanatory variables was run. This model was then tested with a number of restrictions, in order to delineate whether any one grouping had an overarching influence on proportion of women or whether it is best predicted by an amalgamation of different influences.

This model confirmed expectations that having a gender quota present increased the proportion of women in parliament. A significant positive effect was found (p=.011) in which having a quota resulted in a 5.058% increase. This is in line with prior findings. It makes sense logically that gender quotas in the political environment are likely to lead to increased numbers of women in parliament, when compared to having no such quota. Labour force participation rate was, notably, the only socio-economic factor found to have a significant effect on parliament participation rates (p=.007). An increase of 1% led to a .257% increase. This suggests that the more women that are engaged in working world the more likely they are to be represented in parliament. This is in line with prior findings. However, it could be argued that an environment that promotes active participation in the labour force also encourages participation in parliament, i.e. both variables could be explained by a third factor, not a direct causal link between the two.

Contrary to expectations and past research, there were no significant effects found for the religious variables. It may be that they are a poor proxy of the attitudes of a country. Alternatively, the broad classifications of Catholic, Muslim and other, used in this model, left out important differences. Out of the regional dummy variables there were three significant effects. In line with

the hypothesis it seemed that a country's being Nordic had a positive effect on the proportion of women in parliament. Interestingly both Latin American Caribbean countries and African countries displayed negative significant effects on proportion of women in parliament, although significance was only found at the 10% level. It is worth noting that the regional variables were being compared to the Western classification of countries. This was chosen as the benchmark group for a number of reasons, namely that it is this region that has typically endorsed women's rights and that the countries within the region are generally perceived as relatively democratic.

The Adjusted R-Squared in this model was .4136. This suggests that 41.36% of the variation in the proportion of women in parliament is explained by the model. The p-value of the F-test of overall significance was less than .01 which supports the overall significance of the model. Tests of multiple restrictions indicated that each grouping of variables adds significant explanatory power to the model, which justifies their inclusion in the final model.

It seems that a variety of factors contributes to explaining the proportion of women in parliament. Policy makers should focus on a variety of mechanisms in order to increase the numbers of women in parliament. Certainly some of the factors found to be significant in this model are modifiable. The implementation of gender quotas and policies that aim at increasing female labour force participation rates may be useful mechanisms of change.

Limitations and Extensions

This research is not without its limitations, some of which have been alluded to throughout this paper.

Firstly, the appropriateness of religion and region variables as proxy variables for attitudes is questionable. It is postulated that a direct measure of attitudes would have significantly more explanatory power. Rudein (2010) found that using direct measures of attitudes from the world value survey had significantly larger effects and his model had higher explanatory power. However, the sample size of countries with direct measures is small at present and more research is needed in order to broaden the spectrum of countries covered. Attitudes are difficult to measure and certainly the development of a satisfactory scale would be a significant advancement for many fields. There is no such tool available now and research is limited by using proxy variables that may or may not reflect the cultural values towards women in parliament.

Secondly, the variable of quota indicated the presence of a quota at some stage in the political system. However, there are many different forms of quota in place around the world, including reserved seats, legal candidate

quotas and political party quotas. Future research should consider whether one type of quota is more effective at increasing the number of women in parliament.

Thirdly, there are explanatory variables, which have not been included in this analysis, which are likely to be included in the error term. These may be biasing the results. One likely candidate is a family history of involvement in politics, which has been shown to influence involvement in politics by both genders. The length of time since women were given voting rights in a country may also have an impact. Certainly there are likely to be different effects, depending on the length of time since they were allowed to vote. In some circumstances, this would result in fewer women involved as it is against the norm and there is a lack of role models. In other circumstances, a newly acquired voice may be valued more highly and encourage more women to engage in politics. The omission of these and other variables could be addressed in future research.

Finally, this analysis looked at a cross-country sample at one point in time. It would be interesting to use panel data in order to track longitudinally the evolution of the rates of female involvement in parliament across countries.

Conclusion

This research has demonstrated that the proportion of women in parliament can be predicted using OLS estimation. It seems that a model involving a mix of socio-economic, political and cultural factors holds the most explanatory power. Certainly, there is no simple relationship between any one variable or set of variables and the proportion of women in parliament. However, this should not deter policymakers, researchers and the public from endeavouring to increase participation rates. On Forbes' List of World's Most Powerful People 2011, Angela Merkel was the only woman to feature in the top 10. It is hoped that recognition of the causal influences of under-representation in parliament will allow barriers to be broken down, and could lead to more equal proportion of the men and women, not only in parliaments, but in all of the worlds forums.

References

Central Intelligence Agency, 2009. [online] Available at https://www.cia.gov/library/publications/the-world-factbook/index.html. [Accessed 15 February 2012]

Global Database of Quotas for Women, 2009. [online] Available at http://www.quotaproject.org/index.cfm. [Accessed 15 February 2012]

Inglehart R. Norris, P., 2003. *Rising Tide, Gender Equality and Cultural Change Around the World*. Cambridge UK Cambridge University Press.

Internatonal Parliamentary Union, 2009. [online] Available at http://www.ipu.org/parline-e/mod-electoral.asp. [Accessed 15 February 2012]

Ruedin, D., 2010. The Representation of Women in National Parliaments: A Cross-National Comparison. *European Sociological Review*. Vol.0 Issue 0. 1-14.

Rule, W., 1987. Electoral Systems, Contextual Factors and Womens Opportunities for Election to Parliament in 23 Democracies. *Western Political Quarterly*. Vol. 40. 477-498.

Tripp, A.M., A.M. Kang, A., 2008. The Global Impact of Quotas : On the Fast Track to Increased Female Legislative Representation. *Faculty Publications: Political Science*. Vol.41. 338-364.

United Nations Statistics Division [online] Available at http://unstats.un.org/unsd/methods/m49/m49regin.htm. [Accessed 15 February 2012]

Wangnerud, L., 2009. Women in Parliaments : Descriptive and Substantive Representation. *Annual Review of Political Science*. Vol.12.51-69.

World Bank, 2009. [online] Available at http://data.worldbank.org/. [Accessed 15 February 2012]

Estimating the Impact of Changes in the Federal Funds Target Rate on Market Interest Rates from the 1980s to the Present Day

Donal O'Cofaigh
Senior Sophister

In this paper, Donal O'Cofaigh quantifies the influence the U.S. Federal Reserve possesses over market interest rates, through their tool of choice, the Federal Funds Rate. Through detailed econometric analysis, he finds that though the Fed still retains significant influence over market interest rates, this influence has weakened over time. In addition, he finds evidence that increases in the Federal Funds Rate have a greater impact on bonds, the longer the maturity.

Introduction

In 1988 Timothy Cook and Thomas Hahn (henceforth referred to as C&H) published a paper entitled "The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s". The authors note that previous attempts to distinguish whether the Federal Reserve (Fed) can influence interest rates had found little support for the theory, which, C&H argued, conflicted with the views among financial markets. This premise is based on the following three principles.

Firstly, the Fed's policy instrument is the Federal Funds Rate (FFR). Secondly, the Fed uses the FFR in order to fit its policies. Thirdly, the expectations theory hypothesis stipulates that long run interest rates are what the markets expect short run interest rates to be at that point in time (removing the risk premium). C&H argue that these three principles imply that "the Fed influences market rates through its control of the current funds rate and its influence on expected future values of the funds rate". The authors then test

46

this theory by regressing the changes in the FFR on changes in bond yields for a number of different maturities. It must be acknowledged that the authors stipulate that the time period over which the tests are conducted was somewhat unique. September 1974 to September 1979 was a period in which the Fed monitored rates particularly closely, so closely in fact, that market participants could identify most target changes on the day they were implemented. C&H found that changes in the FFR were followed by large movements in the same direction in short rates, moderate movements in intermediate rates, and small but significant movements in long-term rates.

I have attempted to reconstruct these tests from the period 1980 to present. My theory is fundamentally the same, in that prior to testing the data, I expect that changes in the FFR will result in corresponding changes in the same direction in bond yields. I expect this impact to have been somewhat diluted since the '70s as the Fed has sought to increase its transparency when conveying information to the markets in terms of its stance on monetary policy. This has culminated recently in the Fed announcing at the January Federal Open Market Committee meeting that it anticipates that rates will stay low into 2014. This was followed by a flattening of the yield curve as long rates decreased further.

Goodfriend (1990) noted that a particularly interesting aspect of C&H's findings was that T-Bill rates tended to move by only half of the change in the target rate, which would have suggested that the market has already built in the other 50 percent of the move into the market rates on the day of the change. What will become obvious from my findings below is that the markets have moved to price a significantly larger portion of anticipated rate changes prior to the announcement. If the FOMC announced at a March meeting that it was not raising rates at the present time, but was moving to an asymmetric bias towards raising the rate in the future, you would see a shift upwards in the yield curve today, despite the fact that rates are unchanged. What is evident is that, were the Fed to imply an increasingly hawkish stance, the market would anticipate future rate increases and would price that in to the market that day. Consequently when it comes to the day of the actual rate change, the market has fully anticipated this change, and yields do not change to the same extent that they might have done in the past. Forward curves are particularly indicative of the market's anticipation of the Fed's movements. For example, the 1y1y forward is the one-year rate in one year's time. If this is significantly above the current one year rate, then this would provide evidence that the market anticipates rate increases over the coming 12 months.

The paper continues as follows. In Section I, I review the expectations hypothesis for the term structure of interest rates, and then I introduce

the initial regressions of changes in the FFR on bond yields. Section II reviews a number of potential scenarios where a change in the target rate could potentially have a greater impact on yields. Section III briefly focuses on the potential impact that changes in the target rate may have on inflation expectations.

Section I

Rational expectations implies that the interest rate on a longer-term security can be expressed as a weighted average of the current and expected future short-term interest rates over the life of the longer term security, plus a risk premium (Simon, 1990). C&H's findings, along with those held by market participants, constitute strong evidence that expectations of future level of the funds rate influence current market rates. Campbell and Shiller (1990) note that the expectations theory of the term structure implies that the spread is a constant risk premium, plus an optimal forecast of changes in future interest rates.

I have regressed changes in the FFR on changes in bond yields, beginning in October 1979, up to the most recent rate changes.[1] The Fed adjusted the target rate approximately one hundred and fifty nine times over the course of the sample.[2] The path of bond yields and the FFR from 1970 to now is best summarized in Figure 1.

Since the late 1970s, the yield on Treasuries has trended downwards, closely following the trend in the Fed's target rate. In order to see this more clearly, Figurea 2 and 3 depict the progressions of both the 3-Month T-bill yield and 30-Year Treasury from the late '70s up to the present day, with the target rate. The 30-Year Treasury has followed a similar trend, but quite notably does not follow quite as closely.

The results to the following regression are found in Table 1.

$$\Delta i = \beta_0 + \beta_1 \Delta FFR + u$$

Where Δi denotes the change in yield, and β_1 measures the market response to the unanticipated portion of the target change.

---

1 The data were sourced from Bloomberg. Close of day yields were extracted. It must be noted that 12-Month T-Bills were omitted because the data on Bloomberg for 12-Month yields were inconsistent.

2 A number of the sample changes were reported in the data as a change on a specific day when in reality they were spread out over the course of a number of days. In particular, Bloomberg has on record an increase of 8.5 percent in the FFR on the 01/04/1980, which I omitted from the sample.

Table 1

Coefficient	3month	6months	2yr	5yr	10yr	30yr
Intercept	0.0140	0.0274	0.0299**	0.0301**	0.0275**	0.0193*
	(0.0190)	(0.0177)	(0.015)	(0.0137)	(0.0115)	(0.0109)
β_1	0.0847**	0.0608*	0.0544***	0.0302**	0.0203**	0.0245**
	(0.0384)	(0.0313)	(0.017)	(0.013)	(0.0096)	(0.0118)
Adjusted R Squared	0.1114	0.0660	0.0734	0.0249	0.0137	0.0260
Standard Error	0.2372	0.2226	0.1889	0.1723	0.1454	0.1369
D-W Statistic	2.1635	2.1476	2.2616	2.1394	2.0951	2.1386

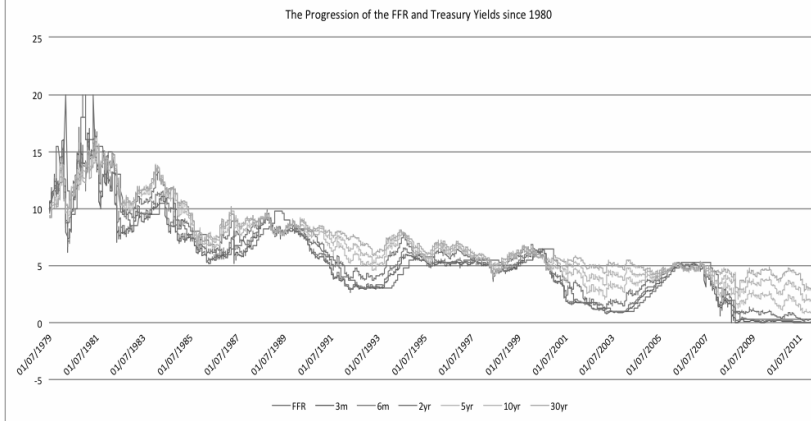Figure 1: The Progression of the FFR and Treasury Yields since 1980
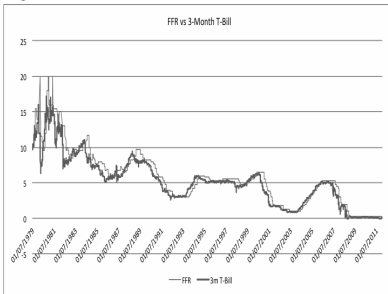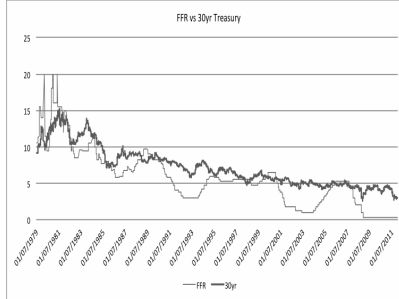

Figure 2: FFR vs 3-Month T-Bill

Figure 3: FFR vs 10-yr Treasury

Having failed to reject the null hypothesis of homoskedasticity it was not necessary that I included the robust standard errors, but I felt that it added weight

to the findings. Durbin-Watson test results for each regression are included in the findings. In each case the result is in excess of 2.0, so serial correlation is not an issue3. Adjusted R-Squared values indicate that up to 11 percent of the variation in the yields can be accounted for by variation in the FFR.

As we can see from the results, in each case a change in the FFR has a statistically significant impact on the change in bond yields. The results are interpreted as follows: for a 1 percent increase in the FFR, we would expect the 3-Month T-bill yield to increase by 8.5 basis points. The reaction of longer-term rates to changes in the FFR diminishes as the maturity increases, but the impact remains statistically significant. An interesting point to highlight here is the contrast between the results observed by C&H, and what I have here. So while the results remain statistically significant, C&H found the corresponding changes in bond yields to be far greater than the impact I have found. The theory for this was briefly highlighted above. As a consequence of the attempted increase in transparency from the Fed, it would appear that changes in the FFR become almost entirely anticipated prior to the actual change, with in excess of 90 percent of the move priced into market yields.

The correlation coefficients of different maturities are presented in Table 3. The figures here highlight the tendency for long-term rates to move with the T-Bill rates. One of the methods of testing the expectations hypothesis is to regress the changes in T-Bill yields on changes in Treasuries of a longer maturity. I ran the following regression, with the results presented in Table 2.

$$\Delta i = \beta_0 + \beta_1 \Delta FFR + \beta_2 \Delta 3Month + u$$

Where β_2 denotes the relationship between changes in the 3-Month T-Bill and bonds of longer maturity, and β_1 the deviation as a consequence of changes in the FFR. The results now depict statistically insignificant responses of bond yields of longer maturity to changes in the FFR, coupled with a statistically significant impact from the change in 3-Month bills. Thornton (1998) notes that it might be unwise interpreting these results in terms of the expectations theory. Mankiw and Miron (1986 cited in Goodfriend, 1990), noted that in the presence of a time-varying premium, the coefficient in such a regression tends to be biased downwards. An alternative theory posed by Thornton (1998) is the segmented market hypothesis, which asserts that individuals have a preference to either borrow or lend in one end of the market,

3 I tested for Serial correlation in each case, and found the easiest way to represent it in the findings was with the Durbin Watson Statistic, the premise of which is if D.W.>2, then there is no issue with serial correlation.

and a consequence of this is that bond yields of different maturities are not necessarily related. The statistically significant impact of changes in 3-Month yield does not provide concrete evidence of the expectations theory. Durbin-Watson statistics once again indicate no issues with serial correlation. The adjusted R-Square results highlight that a much greater extent of the variation of our bond yields is now caused by the independent variables.

Table 2

Coefficient	2yr	5yr	10yr	30yr
Intercept	0.0214**	0.0231**	0.0221**	0.0147*
	(0.0099)	(0.0098)	(0.0088)	(0.0086)
β_1	0.0031	-0.0121	-0.0121	-0.0035
	(0.0183)	(0.0179)	(0.0156)	(0.0153)
β_2	0.6052***	0.5006***	0.3831***	0.3306**
	(0.1214)	(0.1368)	(0.1246)	(0.1309)
Adjusted R Squared	0.6057	0.4845	0.3952	0.3413
Standard Error	0.1232	0.1253	0.1138	0.1126
D-W Statistic	2.2580	2.1008	2.1295	2.1529

Table 3: Correlation Coefficients

	3month	6month	2yr	5yr	10yr	30yr
3month	1.000					
6month	0.940	1.000				
2yr	0.781	0.876	1.000			
5yr	0.698	0.822	0.951	1.000		
10yr	0.630	0.766	0.891	0.972	1.000	
30yr	0.591	0.718	0.834	0.899	0.933	1.000

Section II

In order to test whether the market's reaction to a change in the Federal Funds target rate was significantly different for a larger change I created a dummy variable that assigned 1 to every change in the FFR that was greater than 25 basis points. There were eighty occasions from my sample of changes where

the change signaled by the Fed appeared to be greater than 25 basis points. The regression was as follows.

$$\Delta i = \beta_0 + \beta_1 \Delta FFR + \beta_2 \, Dummy + u$$

The results to this regression are presented in Table 4. The β_1 coefficient in each scenario remains statistically significant, and is almost identical to those findings in Table 1, but the β_2 coefficient accompanying the dummy variable is in no case statistically significant, and is inconsistent in sign across the maturities. What we can conclude from these findings is that large changes in the FFR do not have a statistically significantly greater impact than smaller changes. This could likely be attributed to the greater transparency the Fed has sought, and increasingly aggressive language and opinion towards changing rates would provide indicators to the market in advance of a rate change. Durbin-Watson statistics and Adjusted R-Squared values are similar to those found in Table 1.

Table 4

Coefficient	3month	6months	2yr	5yr	10yr	30yr
Intercept	0.0191**	0.0215**	0.0235**	0.0237**	0.0226***	0.0199***
	(0.0088)	(0.0103)	(0.0104)	(0.0103)	(0.0086)	(0.0069)
β_1	0.0848**	0.0606*	0.0542***	0.0301**	0.0202**	0.0245**
	(0.0384)	(0.0316)	(0.0171)	(0.0130)	(0.0097)	(0.0119)
Dummy	-0.0102	0.0118	0.0127	0.0126	0.0097	-0.0011
	(0.0381)	(0.0353)	(0.0299)	(0.0273)	(0.0230)	(0.0217)
Adjusted R Square	0.1061	0.0607	0.0686	0.0200	0.0085	0.0197
Standard Error	0.2379	0.2233	0.1894	0.1727	0.1457	0.1373
D.W. Stat	2.1689	2.1405	2.2554	2.1328	2.0885	2.1400

Dummy=1 if change in FFR was >25 basis points

Thornton (1998) argues that the Fed typically makes addition rate adjustments in the same direction, and that the market generally comes to anticipate this behaviour, so successive moves in the same direction are more likely to be priced into the market yields. This could also be argued to increase the effect that the Fed has on the market when it switches from a period of rate rises to the first decrease. The reversal could potentially be seen by the market as the first in a series, and so yields might move by a greater amount in antici-

pation of future changes. In order to assess the potential for this, I constructed another dummy variable that assigned 1 to each rate change that was not of the same sign as the one previous to it. There were 35 instances in my sample when the rate change did not follow previous changes in sign. The regression I ran was as follows, with the results depicted in Table 5.

$$\Delta i = \beta_0 + \beta_1 \text{Dummy} + \beta_2 \Delta \text{FFR} + u$$

Once again the β_2 shows little deviation from the findings in Table 1, with a similar case for D-W statistics and Adjusted R-Squared values. While the coefficients that accompany the dummy variable are consistent in sign across the board, they remain statistically insignificant. What the numbers indicate is if there is a rate change that is different in direction from the preceding change, the change tends to be larger, with the size of basis point move varying from 0 to 3.8. Unfortunately, the evidence is not concrete as the coefficients lack statistical significance.

Table 5

Coefficient	3month	6months	2yr	5yr	10yr	30yr
Intercept	0.0140	0.0213	0.0257	0.0216*	0.0204**	0.0139*
	(0.0192)	(0.0162)	(0.0136)	(0.0117)	(0.0096)	(0.0079)
Dummy	0.0000	0.0278	0.0193	0.0387	0.0324	0.0246
	(0.0543)	(0.0577)	(0.0504)	(0.0477)	(0.0409)	(0.0420)
β_2	0.0847**	0.0600*	0.0538***	0.0292**	0.0194**	0.0238*
	(0.0382)	(0.0314)	(0.0172)	(0.0131)	(0.0098)	(0.0124)
Adjusted R Square	0.1057	0.0626	0.0692	0.0272	0.0159	0.0252
Standard Error	0.2379	0.2230	0.1894	0.1721	0.1452	0.1369
D.W. Stat	2.1635	2.1243	2.2565	2.1336	2.0870	2.1279

Dummy=1 if Policy Changes Direction

Section III

The difference between the reaction of longer maturity bonds and that of shorter maturity bonds to changes in the FFR is seen as supportive of the interpretation that target changes affect the market's inflation outlook. The general assumption is that inflation can be reduced by increasing short-term interest rates. Alvarez et al (2001) note the evidence linking money growth, inflation and interest rates: "increases in average rates of money growth are

associated with equal increases in average inflation rates and interest rates". The control of inflation is vitally important, and is one of the primary mandates for central banks around the world. Bernanke (2007) discusses the importance of the control of inflation, which is deemed to inject noise into the price system, to make long-term planning more complex, and to increase uncertainty. The same author also notes that experience suggests that high and persistent inflation undermines public confidence in the economy and in the management of economic policy generally.

The US Treasury began issuing 10-Year Treasury Inflation Protected Securities (TIPS) in January 1997 and 30-Year TIPS in April 1998[4]. I have compiled the yield data for bond break-even levels from 1998 up to the present day and have run a regression similar to that from Table 1. In this case, measures the changes in breakeven on the day of change in the FFR.

$$\Delta i = \beta_0 + \beta_1 \Delta FFR + u$$

Bond break-even level is essentially calculated from the Fisher Equation.

$$i = r + \pi$$

It is calculated as the difference between the yield on nominal bonds (i) and the yield on an inflation-linked bond of the same maturity (r). It can be interpreted as the expected inflation over the period up to the point of maturity of the bond. There were forty-nine instances, in which the Fed changed the target rate, that were applicable to TIPS. The results of the above regression are found in Table 6. Because changes in monetary policy are unlikely to have an immediate effect on market inflation expectations, we would anticipate that changes in inflation expectations would be muted in the short end of the curve and would have some impact in the long end. This is consistent with the findings in Table 6. While changes in the FFR have the expected impact on breakeven levels, the data are only significant at the 30-Year maturity. The signs of the β_1 coefficients are the opposite to those in Table 1. This is what we would expect: an increase in the FFR would decrease the market's anticipated rate of inflation and so reduce bond break-even levels. Both the Adjusted R-Squared value and the D-W statistics are similar to those found in previous results.

It must be highlighted that the measures of inflation used above (bond break-even level), are influenced by changes in inflation risk premi-

4 Consequently, break-even data do not exist for maturities less than 10 years for an extended period.

ums and liquidity premiums, and analyses are constrained by the fact that these markets have only been in operation in the US for a short period of time. However, these measures are market determined, with investors backing their views with real money. This is in contrast to forecasts determined by economic models and surveys.

Table 6: Impact on Break-Even

Coefficient	10yr BE	30yr BE
Intercept	0.0090	0.0070
	(0.0079)	(0.0058)
β_1	-0.0286	-0.0389**
	(0.0246)	(0.0179)
Adjusted R Square	0.0068	0.0740
Standard Error	0.0639	0.0455
D.W. Stat	2.0460	2.1739

Summary and Conclusions

The results highlighted in the tables above all indicate the strength that the US Federal Reserve has over market yields, and long-term inflation expectations. Although this impact has been diluted over time, the statistically significant results still emphasize the importance of the role the Fed has to play. An important assumption underlying all of the above results is that movements in the target federal funds rate cause movements in other market rates and not the reverse. For reverse causation to be a valid concern, changes in the federal funds rate would have to be triggered by contemporaneous daily changes in market rates, which seems unlikely, given the highly volatile nature of interest rates. Meyer (2004) made reference to the fact that the Federal Reserve does not like to be pushed into making changes by the market. Although the Fed will strive to avoid "shocking" the market, they are not likely to make changes if they do not see fit, even if the market has moved to price in an anticipated change.

The Fed began operations in 1914. Prior to the creation of the Fed, there were a number of measures of short-term interest rates, which were often subject to periods of extreme volatility, with changes of over 10 percent occurring on 8 occasions during the Civil War period. Since the creation of the Fed, such extreme temporary spikes have been absent. Evidence over time suggests that the Fed has had strong influence in smoothing rates.

References

Alvarez, F., Lucas, R. and Weber, W., 2001. Interest Rates and Inflation. *The American Economic Review*, 91(2), pp.219-25.

Bernanke, B., 2007. Inflation Expectations and Inflation Forecasting. Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts.

Campbell, J. and Shiller, R., 1991. Yield Spreads and Interest Rate Movements: A Bird's Eye View. *The Review of Economic Studies*, 58, pp.495-514.

Cook, T. and Hahn, T., 1988. The effect of changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s. Richmond: Federal Reserve Bank of Richmond.

Gibson, W., 1972. Interest Rates and Inflationary Expectations, New Evidence. *The American Economic Review*, 62(5), pp.854-65.

Goodfriend, M., 1990. Interest Rates and the Conduct of Monetary Policy. Richmond: Federal Reserve Bank of Richmond.

Meyer, L., 2004. *A Term at the Fed: An Insider's View*.

Simon, D., 1990. Expectations and the Treasury Bill- Federal Funds Rate Spread over Recent Monetary Policy Regimes. *The Journal Of Finance*, 25(2). pp.567-77.

Thornton, D., 1998. Tests of the Market's Reaction to Federal Funds Rate Target Changes. *The Federal Reserve Bank of St. Louis Review*, November/December 1998. 25-36.

Political Aspects of the Bundesbank: Did the German Central Bank Have the Final Say on German Monetary Union and European Monetary Union?

Flannery Dyon
Senior Sophister

As the euro area's debt and banking problems increase, so has the role Germany plays in euro area policy-making, as the European Central Bank attempts to reassure markets. Flannery Dyon provides a fascinating insight into the influence of the Bundesbank, the German Central Bank, in shaping both German and European monetary policy.

Introduction

"There was just a moment of chaos when the D-mark took over East Germany on July 1st: a crowd of 10,000 gathered in East Berlin's Alexanderplatz and banged on the doors of Deutsche Bank, which as a publicity stunt chose to open at midnight. But the Prussian discipline prevailed. The Bundesbank had advised East Germans to save, not spend, their new notes. The East Germans obliged by drawing much less than the Bundesbank had expected. The feared spending boom was not to be."

In just a few sentences The Economist's Berlin correspondent (1990) manages to describe the mixture of excitement and apprehension that accompanied the first hours of the German Monetary Union (GMU), which was accomplished on July 1st 1990, preceding the formal German reunification that would take place on October 3rd of that same year.

The achievement of GMU was an essential step in completing formal German Reunification, which had been one of the main foreign policy goals of the Federal Republic of Germany since its creation in 1949. The East Germans also saw in the Deutsche Mark (DM) a symbol of the economic success and high standard of living West Germany had enjoyed, of which they also hoped to benefit from.

However, these hopes were equally matched by the uncertainty of unified Germany's economic future, as very little data and knowledge existed regarding the actual economic state of the German Democratic Republic (GDR), and also because there was huge debate with regards to how the 'Mark der DDR' (Mark of the GDR, or 'Ostmarks') should be converted.

As the monetary institution of the Federal Republic, and due to its independent status, the Bundesbank ('Federal Bank') played a key role in determining the modalities of GMU and in managing the monetary consequences. However, reunification also had international effects - the DM was tied to the European Monetary System (EMS) and Germany had been part of the project for a European Monetary Union (EMU).

The Bundesbank has always been driven to protect both price stability and its independence - two of its main characteristics -; and in order to do so it has played a political game with the federal government by using its influence on the German public and its international reputation to determine key aspects of both GMU and later on EMU. This was at times a risky game, as the independent status of the Bundesbank is determined by an all too revocable law - and modifications to this status have almost been ratified by both the Bundesrat ('Federal Council') and the Bundestag ('Federal Parliament'). We are interested in discussing the role the Bundesbank played in both GMU and EMU and how its philosophy and its Stabilitätskultur ('culture of stability') determined the modalities of these two unprecedented events. It is a remarkable occurrence that a Central Bank should exert such influence outside of its own realm of expertise, which is why we argue that the political dimension of the Bundesbank is what protected its Stabilitätskultur.

The Emergence of the Bundesbank and its Monetary Thought

Knowledge of the historical context is paramount to understanding the genesis and philosophy behind the Bundesbank's policies. In its official publication on its scope of tasks, legal framework and history ('Die deutsche Bundesbank: Aufgabenfelder, rechtlicher Rahmen, Geschichte'), the Bundesbank goes back extensively on the monetary institutions that preceded it.

Germany's monetary institution has been known as the Bundes-

bank since 1957 and has been the sole monetary policy agent in the Federal Republic of Germany (FRG) since then. However, the Reichsbank ('Imperial Bank') – the monetary institution from 1871 to 1945 – had a profound impact on monetary thought in Germany. Its board members were appointed by the Kaiser ('Emperor'), and later the Reichspräsident, so that it was not independent.

During the first half of the 20th century, Germany experienced two world wars, hyperinflation, high unemployment, devastation, fascism, and loss of sovereignty. Inflation is associated with those disastrous events, and that is why the law giving the Bundesbank its purpose speaks of protecting the currency. Indeed central bank independence means that monetary policies are not subject to a time-consistency problem.

In 1945, as the Americans administrated a quarter of post-WWII Germany, they were confronted with the volatility and weakness of the Reichsmark, thus initiating the Währungsreform ('currency reform') of 1948, which then spread to the zones administrated by the UK and France. Within each federal state a Landeszentralbank ('state central bank') was set up, and in March 1948 the Bank deutscher Länder (BdL) was created to govern them. This highly decentralized central bank resembled the Federal Reserve System. In June 1949, the DM became the official currency of what would, a few months later, become the FRG. The BdL remained subject to the directives of the Allies (excluding Russia). It was only in 1957 that the FRG founded an entirely independent central bank, the Bundesbank, and gave it one clear purpose: to protect the currency ["mit dem Ziel, die Währung zu sichern" (Bundesbank, 2006:26)]. 1957 is also the year that the Treaty of Rome was established, which founded the European Economic Community, thus a correlation link could be ascertained.

Nevertheless, the Bundesbank's independence could be challenged if the Bundesrat and the Bundestag revoked this law. This is an important element in order to understand the at times fragile relationship between the Bundesbank and the federal government.

However, since the Bundesbank has been established, the DM has been a strong, reliable and very stable currency, associated with the 'Wirtschaftswunder' ('economic miracle') of the 1950s and '60s in the minds of the German population. The stability that followed from 1957 to 1990 - with exception of the oil crises of 1973 and 1979 – sharply contrasts with the former volatility of the currency during the '20s and '30s.

The European Monetary System and the Build-up to the Reunification of 1990

Since 1949, the German federal government has been involved in two major foreign policy endeavours: the European Integration process and the 'Ostpolitik'.

The former led to participation in the EEC, the common market, and in various attempts at creating a monetary union. Although the Government was keen in asserting the FRG as part of Europe and in furthering peaceful relations with its neighbours, as well as in regaining a respectable place on the stage of world politics (Kaltenthaler, 2002), we can argue that a strong motivation for this openness to the West could have been the possibility to benefit from more trade and improve exports.

However, as talks of a European Monetary Union (EMU) surfaced throughout the 60s, the Bundesbank was strongly against it, for it involved letting go of monetary authority – and endangering price stability.
The FRG was already part of the Bretton Woods system until its collapse in 1971, which meant that the Bundesbank was struggling to keep inflation low internally whilst being tied to a fixed exchange rate system. It seemed that the Bundesbank did not want to take part in another fixed exchange rate system, whereas this was what the European integration process seemed to tend towards.

At this point we can identify two coalitions: a 'foreign policy coalition' and a 'monetary stability coalition'. The foreign policy coalition, which is made up of the Chancellery and the Foreign Ministry, pushes towards more European integration, whereas the monetary stability coalition, which safeguards the stability of the DM, is led by the Bundesbank and the Finance Ministry (Kaltenhalter, 2002).

In a sense they can both be thought of as lobbies, so the Bundesbank shaped negotiations and influenced decisions on a European level, for it had the German public opinion on its side, and other European leaders, namely France, could not consider an EMU without the German giant. There had been a European snake before, which imploded due to the oil crisis in the '70s and the impossibility for other European institutions to follow the Bundesbank's policies and the strong DM. In 1978, another attempt at an EMU was initiated. Although negotiations between government representatives started by considering a European 'basket of currencies', so that there would be symmetry, pressure from the Bundesbank and the monetary stability coalition led the European Monetary system (EMS) launched in 1978 to have the DM act as the anchor currency. (Indeed the German Dominance Hypothesis has

been proven even though it should be considered carefully (Kirchgässner and Wolters,1993; Gardner and Perraudin, 1993). The other European central banks had to ensure that their currency remained within a certain relation-value to the DM. Thus an asymmetric European fixed exchange rate system was created. "The burden of policy adjustment in the system was asymmetric; Germany led, the others had to adjust in order to follow" (Kaltenthaler, 2002:80)

While the FRG was integrating itself into the West, it was also opening itself up to the East. The second major foreign policy endeavor of the FRG was the Ostpolitik - the attempt to normalize and appease relations with the GDR. After decades of European integration and increased diplomatic relations to the east, the fall of the Berlin Wall on November 9th 1989 would not only mark the end of the cold war but also disrupt the course of German-German and German-European relations. Reunification was on the horizon, and the next logical step was German Economic and Monetary Union.

German Monetary Union and the Bundesbank's Stand

The GDR, although being the eastern block's country with the highest GDP per Capita, was a drastically different country from the FRG. Not only was its Ostmark considered to be 'funny money', with very low purchasing power, its banks offered none of the financial assets that were available in the West. So achieving monetary union meant overcoming these two challenges. Furthermore, the Bundesbank had access to very little data regarding the GDR and the unpredictability of events made the introduction of the DM in East Germany risky (Hagen, 1993).

Deciding on the conversion rate between the Ostmark and the DM was particularly problematic. Prior to GMU, East Germans could convert their Ostmarks for DMs at a 10:1 conversion rate on the black market, reflecting the vast gap between the two economies. A 1:1 parity rate would destroy all possible East German competitiveness, and the finance minister at the time, Theo Waigel, as well as the president of the Bundesbank, Karl Otto-Pöhl, warned that a hasty monetary union could lead to lasting damage to the East German economy. (Welt am Sonntag, 2004).

Regardless of this, politics took over. The Bundesbank had repeatedly managed to impose its own interests (as we've seen in the case of the EMS), but it had been able to do so because of its popularity with the German public. As Germans protested against the projected conversion rate of 2:1, they held signs saying "Wir sind ein Volk, 1:1" ('We are one people: 1:1'). The Federal Government and the Bundesbank could hardly go against the spirit of

reunification, proving that both are almost equally dependent on public approval. Especially concerning savings and social benefits, one could not make disparities so blatantly obvious in such newly formed union.

Ultimately it was decided that wages and rental payments would be converted at 1:1 rate, as well as savings deposits. Other savings would be converted at a 1:1 rate but only within a certain limit (up to 2000 DM for under 14 year olds, up to 4000 DM for the people between 14 and 58 years old, and up to 6000 DM for the people over 58). The general exchange rate was set at 2:1. (Collier Jr). Although the Bundesbank regarded the conversion rates as a political success, it did declare that the 'actual favourable savings rate was 1·8:1' (2006).

These politically influenced decisions had numerous implications; mainly there was the fear that there would be a huge surge in spending in East Germany, which would lead to inflation. The large amount of capital investment in East Germany also created some inflationary pressure. The Bundesbank anticipated this, and thus rose interest rates at this time (This would, in turn, have an effect on the EMS, as we will see later.).

Aside from the monetary issues, reunification demanded huge fiscal transfers (nearly 170 billion DM in 1992 alone, see Renzsch, 1998) and necessitated the privatisation of East German companies – which proved to be problematic due to the contention of ownerships. The government could not print money in order to finance reconstruction, and the Bundesbank would of course not let it do so. The influential German labour unions put East German and West German wages on par with one another, although productivity and the ability to work in an entirely different economy were vastly different. East German companies could not compete, with wages being too high and West German products invading the market (unemployment in East Germany remains particularly high nowadays).

The Ramifications of GMU on a European Scale

The consequences of German reunification were not only internal. In 1991, R. F. Owen already feared the consequences a DM appreciation might have on the EMS, and on the possible EMU.

The rise in interest rates by the Bundesbank, the high levels of both public and private demand for capital, the fiscal expansion and the increased attractiveness of the DM as an investment currency led to an appreciation of the DM (Sinn, 1996; and Hagen, 1993). This had severe implications for the EMS. Indeed an appreciation of the DM meant that the other countries in the EMS fell under a lot of stress, even more so than previously. If the anchor cur-

rency in a fixed exchange rate regime appreciates, the other currencies have to resist devaluation pressure and artificially increase interest rates and forgo economic growth (Sinn, 1996).

> "The rise in the German interest rate had important consequences for Europe, as it led to a crisis in the European Exchange Rate mechanism (ERM) that preceded European Monetary Union. The higher German interest rate meant that the Deutschmark required a revaluation with the ERM, or equivalently, the devaluation of other ERM currencies. France and other countries attempted to maintain the existing exchange rates, fearful of a loss of deflationary credibility. But in 1992 speculative attacks forced several countries to devalue, while the United Kingdom and Italy left the ERM."
>
> (Hunt, 2006:15)

This is what happened in France, for example (Kaltenthaler, 2002). Many could not keep up and had to exit the system and devalue their currencies. On Black Wednesday, 16 September 1992, the UK pound devalued, followed by the Italian lira. Even though the Bundesbank had previously offered to revalue the DM within the system, a political decision of defending the existing rates (Sinn, 1996) meant that there was no stopping the capital flows, thus leading to the breakdown of the system.

WThe asymmetry in the system had been the source of frustration for all participants, especially France. The Franco-German duo that seemed to have led the European integration process was put under strain due to the EMS, which was exacerbated by the German Reunification. A unified Germany could mean a further delay in talks of a symmetrical EMU. But the French were one of the four initial administrators of defeated Germany in 1945 – and still had some hold over German politics, and especially German borders. Their participation, as well as the USA's, the UK's and Russia's, was needed in order to finalize German unity. Article 7 of the Two-plus-Four Treaty clearly states:

> "(1) The French republic, the Union of Soviet Socialist Republics, the United Kingdom of Great Britain and Northern Ireland and the United States of America hereby terminate their rights and responsibilities relating to Berlin and to Germany as a whole. As a result, the corresponding, related quadripartite agreements, decisions and practices are terminated and all related Four Powers institutions are

dissolved.

(2)The united Germany shall have accordingly full sovereignty over its internal and external affairs."

The French threat of withholding this signature was to a certain extent what precipitated the intergovernmental conference on EMU (Sinn, 1996), as though, in a way, EMU had been bargained against German unity. There was also a belief that German unity could be more easily accomplished within an integrated Europe.

All these events precipitated EMU, and while it was the government that signed the official documents, the Bundesbank would not easily forgo its authority over price stability in Germany. An EMU could not exist without Germany, and the German public was siding with the Bundesbank. The government had to negotiate according to the terms the Bundesbank requested. They needed a guarantee that an EMU would not be influenced by the inflationary tendencies of other European countries, and so that the European Central Bank (ECB) remained independent.

> "The Kohl government faced a Bundesbank and Finance Ministry committed to shaping EMU according to their preferences. That meant an EMU that first and foremost had price stability as its goal. The best way to accomplish this was to replicate the institutions of the Bundesbank and its price stability at the European level. The Bundesbank's position was backed by the banking and industry organizations that also wanted to avoid EMU becoming a way to reinflate Europe. The Bundesbank had German public opinion on its side, which feared that EMU would replace the strong Deutschmark with a new weaker European currency. This meant that in the IGC negotiations, the German representatives pushed for and achieved an EMU which conformed, to a very large extent, to the Bundesbank and its allies' institutional preferences"
>
> (Kaltenthaler, 2002:81)

There is a vast array of literature that compares the Bundesbank and the actual ECB (notably Debrun, 2001; and Chortareas, 2003), pointing towards the fact that both institutions not only resemble each other, but are in fact 'related'. A counter-argument to the possibility of board members being influenced by their respective governments is the theory of a 'trickle down' effect, in which the independent status of the ECB seeps down to the national central banks,

making them adverse to influence from their government (The Economist, 7-13 July, 1990).

Because we have observed the ways in which the Bundesbank influenced the EMU negotiations, the convergence criteria for EMU and the design of the ECB seem to be more logical. The convergence criteria to achieve before the actual EMU as laid out in article 109j of the Maastricht treaty were as follows:

> "- the achievement of a high degree of price stability; this will be apparent from a rate of inflation which is close to that of, at most, the three best performing Member States in terms of price stability;
>
> - the sustainability of the government financial position; this will be apparent from having achieved a government budgetary position without a deficit that is excessive as determined in accordance with Article 104c(6);
>
> - the observance of the normal fluctuation margins provided for by the Exchange Rate Mechanism of the European Monetary System, for at least two years, without devaluing against the currency of any other Member State;
>
> - the durability of convergence achieved by the Member State and of its participation in the Exchange Rate Mechanism of the European Monetary System being reflected in the long-term interest rate levels."

These criteria almost all seem suited to the German economy and the objectives of the Bundesbank. However, it could be argued that as the exchange rates were fixed and the conversion to the Euro occurred in 1998, the DM was still suffering from the after effects of German unification (Sinn, 1996). After the creation of the Maastricht Treaty, which also laid out the mission and statute of the ECB, there was a fear amongst the monetary stability coalition in Germany that because other countries had a tendency for budget deficits, a future bail-out would be needed from the ECB; which is why the Bundesbank, along with the then finance minister, Theo Waigel, pressured the federal government into presenting the project for the Growth and Stability Pact, which was implemented in 1997. This event resonates with our current financial crisis, as the role of the ECB has come under scrutiny, and many have wondered whether it would become a lender of last resort.

Conclusion

By exploring the singular position and philosophy behind the Bundesbank's policies, we have in turn highlighted the political aspect of this institution. This political dimension has affected many decisions that were made with regards to GMU and EMU. We also indicated the similarities between the Bundesbank and the ECB, which could lead us to question whether the ECB has also had to manoeuvre the political environment it is in to defend its policies.

Moreover, we can also see that cultural and historical traumas affect economic thoughts and convictions profoundly, and monetary policy has not been sheltered from either politics or cultural specificities.

References

Bundesbank. Die Deutsche Bundesbank: Aufgabenfelder, Rechtlicher Rahmen, Geschichte. Edited and Published by the Deutsche Bundesbank, Frankfurt am Main, April 2006. (Available at http://www.bundesbank.de/info/info_veroeffentlichungen.php)

German Monetary Union: Save and Prosper. *The Economist*, July 7-13, 1990.

Baltensperger, E (et al.), 1998. *Fifty years of the Deutsche Mark: Central Bank and the currency in Germany since 1948*. Oxford: Oxford University Press.

Bernanke, B. S. and Mihov, I, 1996. What does the Bundesbank target? NBER Working Paper Series, Working Paper number 5764.

Chortareas, G., 2008. Monetary policy rules in the run-up to the EMU" *Metronomica*,59:4, pp. 687-712.

Collier, Jr and Irwin, L, 1991. On the First year of German Monetary, Economic and Social Union. *Journal of Economic Perspectives,* 5:4, pp. 179-186

Debrun, X, 2001. Bargaining Over EMU vs. EMS: Why Might the ECB Be the Twin Sister of the Bundesbank? *The Economic Journal*, 111: 473, pp. 566-590.

Eichengreen, B, 1993. "European Monetary Unification" *Journal of Economic Literature,* 31:3, pp. 1321-1357.

Faust, J., Rogers, J.H., and Wright, J.H. An empirical comparison of Bundesbank and ECF monetary policy rules. Board of Governors of the Federal Reserve Sytem, *International Finance Dissussion Papers,* No. 705.

Frowen, S.F and Hölscher, J. ed, 1997. *The German currency union of 1990: A critical assessment.* Basingstoke: Macmillian in association with Anglo-German Foundation for the Study of Industrial Society.

Gardner, E. H. and Perraudin, W. R. M., 1993. Asymmetry in the ERM: A Case Study of French and German Interest Rates Before and After German Unification. Staff Papers - International Monetary Fund, 40:2, pp. 427-450.

Gazioglu, S. and McCausland, D., 1997. German reunification: the budget deficit, sustainable growth and inflation. Aberdeen: University of Aberdeen, Dept of Economics.

Hagen, J., 1993. Monetary union, money demand, and money supply: a review of the German monetary union. *European Economic Review*, 37:4, pp. 803-827

Hunt, J., 2006. The Economics of German Reunification, Department of Economics, McGill University and NBER.

Kaltenthaler, K, 2002. German Interests in European Monetary Integration. *Journal of Common Market Studies*, 40:1, pp. 69-87.

Kirchgässner, G. and Wolters, J., 1993. Does the DM Dominate the Euro Market? An Empirical Investigation. *The Review of Economics and Statistics,* 75:4, pp. 773-778

Lohmann, S., 1998. Federalism and Central Bank Independence: The Politics of German Monetary Policy 1957-92. *World Politics*, 50:3, pp. 401-446.

Owen, R. F., 1991. The Challenges of German Unification for EC Policymaking and Performance. *The American Economic Review*, 81:2, pp.171-175.

Renzsch, W., 1998. Financing German Unity: Fiscal Conflict Resolution in a Complex Federation. *Publius*, 28:4, pp. 127-146.

Other Resources Available Online
The OECD's online statistical database, Available at <http://stats.oecd.org/Index.aspx>

Maastricht Treaty of 7/02/1992, Available at <www.eurotreaties.com/maastrichtec.pdf> (Accessed on 6 January 2012)

The two-plus-four Treaty of 12/09/1990, Political Archive of the Foreign Office, (online) Available at <http://www.auswaertiges-amt.de/EN/AAmt/PolitischesArchiv/EinblickeArchiv/ZweiPlusVier_node.html> (accessed on 20 January 2012)

Retrospection on Germany's Ostpolitik, Political Archive of the Foreign Office, (online) Available at <http://www.auswaertiges-amt.de/EN/AAmt/PolitischesArchiv/EinblickeArchiv/Ostpolitik.html
http://europedia.moussis.eu/books/Book_2/3/7/2/2/?all=1> (Accessed on 5 January 2012)

Reuth, R. G, 2004, "Chronik einer hastigen Währungsreunion", Die Welt am Sonntag, (online), available at <http://www.welt.de/print-wams/article115075/Chronik_einer_hastigen_Waehrungsunion.html> (Accessed on 17 January 2012)a

Land Bank Proposals 1650-1705

Charlie Landale
Senior Sophister

The land bank proposals of the late 17th century are often overlooked when the origins of monetary thought are discussed. Charlie Landale provides a clear and detailed analysis of a number of such proposals, convincingly arguing that the thinkers behind them had identified many of the ideas underlying contemporary theory on the role of money in an economy. He concludes that they deserve greater recognition for their contribution to monetary economics.

An Introduction

During the 1690s and early 1700s, economics was still in its embryonic stage. Relatively little consideration had been given to what the nature and functions of money were. This is unsurprising if we look at the world through the lens of the late 17th century. Money was made of metal, and there was therefore no scope for creating more money without finding new supplies of silver and gold. There were two types of wealthy individual: moneyed men and landed men.

The land bank proponents were early contributors to the economic debate. In their pamphlets the principal problem that they identified was the sluggish economy. They all agreed that the situation could be improved and saw the best means of improvement as an increase in the supply of money. Rather than doing this as the Spanish and Portuguese did by sailing to the new world and bringing back vast quantities of precious metals, they proposed using the banking model that had succeeded in Amsterdam and Venice. According to Schumpeter, they "fully realised the business potentialities of the discovery that money - and hence capital in the monetary sense of the term – can be manufactured or created"(1954: 321). Britain, which was not rich in terms of gold and silver, had plenty of potential in its land. Therefore, a land bank appeared to be a sensible suggestion.

None of the land banks that were set up succeeded. Their proponents, however, contributed important ideas to monetary economics and fuelled the early development of the subject. As Vickers observes, "theoretical discussion moved away from the nature and value of money to the consideration

of the function and importance of monetary circulation and money-holding and payment habits"(1968: 30). In this essay, I consider those ideas in their contemporary context. I will avoid being drawn into the mechanics of their business proposals and focus on the implications of their economic theories.

William Potter's Early Proposal

William Potter has been labelled as the "forerunner of the land-bank projectors" (Schumpeter, 1954: 292). Writing in 1650, he was some 40 years ahead of the other proponents in publishing the Key of Wealth and then The Tradesman's Jewel, a condensed version of the former. His proposal hinged on the idea of merchants issuing collateralised credit. Once the reader cuts through the tangle of his prose, his theory is modern. He makes the observation that "this revolution of commodity, is proportionable to the revolution of money, or that which passes for such" (1650: 5). This sentence elegantly contains three important ideas. First, that economic activity is linked to the volume of money in a system. Potter believed that this was a highly elastic relationship, an idea that is reinforced later when he says: "if money could be increased… it would be a means suddenly to ingrois all the trade and richs that the world could afford for money" (1650: 6). Secondly, his use of the word "revolution" implies that the flow of money is circular. Thirdly, that gold and silver are not the only forms of money. This is perhaps what gave rise to Schumpeter's assertion that Potter was antimetallist in an era where gold and silver were the dominant forms of currency (1954: 292). What constituted money was not important for Potter. He viewed it as a medium to facilitate trade.

One way that merchants were overcoming the problem of a lethargic economy was by issuing private credit. However, this was a relatively risky solution that did not fully solve the problems associated with a shortage of money (Potter, 1650). Murphy describes what Potter was identifying as market participants' inability "to express [their] notional demand as effective demand… there was a cash-in-advance requirement" (1997: 47). Potter believed that this problem could be overcome by increasing the supply of money. This need not cause rampant inflation. In fact, Potter believed that "where trade is extraordinarily quick, commodities may be afforded at much lesse rate, and yet tradesmen gain much more per annum than otherwise" (1650: 26). In short, greater economic activity would drive down prices because merchants could make the same profits from higher sales and lower prices. This is not his only use for the word "quick". He describes how trade will be increased and debts will be paid off more quickly by the "quick and thick revolution of bills" (1650: 14). Potter identified the velocity of circulation as a factor in determining economic activity. This fi-

nal contribution is not mentioned in any of the other land bank proposals. Although it is not fleshed out, it highlights Potter's understanding of the nature of money and puts him amongst the very first monetary economists.

Dr H. C.'s Obsession with Land

It was not until some years later that the land bank was discussed again. The most prolific author and vociferous advocate of such a bank was Dr Hugh Chamberlen, a man of medicine whose family was credited with the invention of gynaecological forceps (Murphy, 1997: 50). His critics would argue that he should not have erred from medicine into the realm of monetary economics. However, his proposals, the bulk of which he made in the 1690s, give some important insights into the workings of the 17th century mind. Horsfield best describes the main thread of Chamberlen's proposals: "to capitalise a 100 year annuity for 100 times its annual value, create bills for the capital sum" (1960: 157).

His reasons for opening a bank based on land credit are important. He says: "credit, rightly founded upon land, must evidently be more secure than any other sort of credit" (Chamberlen, 1696: 2). This statement is critical when we try to understand the proponents of the land bank. They believed that land had the most stable value of all assets. Credit was an important way of trading. The problem with credit was that merchants would only sell on credit to those they knew and trusted. Chamberlen, like Potter, proposed that this situation could be improved by having a universal credit based on something with a very stable value: land. The land bankers were a set of pre-physiocratic economists. Whilst land produces a steady income over a long period, however, it offers poor short term rewards. In a world where money was becoming increasingly important it is understandable that there should be some way for land owners to raise money based on the value of their land. However, due to the illiquid nature of the asset, the bank was bound to run into trouble. Anyone who took a loan from Chamberlen's bank secured with their land simply transferred their liquidity issues to the institution.

Chamberlen attracted plenty of contemporary criticism, particularly over his rather arbitrary calculations of security and rent. In A Bank Dialogue between Dr. H. C. and a Country Gentleman, the Country Gent says of the Doctor's arithmetic: "this is as plain as that 1 and 2 makes 5" (Anonymous, 1696: 3). Despite this abuse, and the failure of the Office of Land Credit in 1699, he continued to publish pamphlets on the subject in Ireland and Scotland, but never raised enough support to found another bank.

John Briscoe Lends to the Government

The next most prolific of the land bank projectors was John Briscoe. In many

respects, his proposal was similar to Chamberlen's, and appeared around the same time in 1694. The bank was once again based on the future rental income that could be expected from land. Rather than 100 years income, however, Briscoe only used 20. Unsurprisingly, he was accused by Chamberlen of plagiarism (Horsfield, 1960: 180). Despite the fact that Horsfield describes Briscoe's major contribution, A Discourse of the Late Funds, as "a lengthy and chaotic work" (1960: 181), it contains some modern theory that puts it ahead of any of Chamberlen's writings. Briscoe believed that the new investment opportunities offered in the forms of the Million Act, the Lottery Act, and most importantly the Bank of England were drawing money out of trade and into the government's accounts. In his opinion, his bank would allow trade to flourish and give the government the opportunity to borrow at a lower rate than they were able to from the Bank of England. The rate at which the government was borrowing money, he believed, would lead to the "ruin [of] the trade of the kingdom" (Briscoe, 1694: 3).

As with Potter and Chamberlen, Briscoe made the link between an increase in the supply of money and greater economic activity. According to him, the new bills would be "to all intents as useful as money; it will be (as it were) an introducing so many fresh-monied men into the Kingdom with several millions more than was before, for the supply of their majesties" (1694: 7). Motivated by finding a means for the government to borrow at a low rate of interest, he was simultaneously concerned with improving the capital of the land owners: "gentlemen will have an opportunity of improving their estates by building, planting, draining or watering their land" (1694: 8). It is not until we get to Nicholas Barbon that the idea of capital is explored in greater detail, but there is an implied understanding of the link between investment and increased productivity here. Briscoe managed to secure over £100,000 in land subscribed to his bank, but it ultimately failed after 1696 as a consequence of a lack of liquidity in much the same way that Chamberlen's did (1960: 194).

Asgill and Barbon: Money and Capital

The last of the land bank proponents to found a bank, albeit another unsuccessful one, were John Asgill and Nicholas Barbon. They briefly worked with Briscoe, and then, in the spirit of the age, they both accused him of plagiarism and were accused by him of copying his ideas (Schumpeter, 1954: 292). Their banking model was closer to a modern building society where a pool of savings was used to make advances on mortgages (Horsfield, 1960). Asgill goes some way in defining the characteristics of money in Several Assertions Proved. He specifically states five that would make land securities suitable as a form of money: they have value; they are durable and incor-

ruptible; they are divisible; the value of each can be certified by a stamp; and they are deliverable (1696: 18). He also offers further insight into the 17th century economist's relationship with land and its importance by stating "what we call commodities is nothing but land severed from the soil… man deals in nothing but earth" (1696: 21). It is easy to see why he considered land to be a suitable security for credit. Several Assertions Proved is a logical pamphlet, even if we now know some of the theory to be wayward. At the end, Asgill acknowledges the dangers of economics as a science:

> "To argue from a chain of positions, successively depending upon one another, which is the most dangerous way of arguing, because if one position happens to be false, it variates all that follows."
>
> (1696: 78)

Perhaps a great deal of the current difficulties that the financial world faces could have been avoided if decision makers had kept this in mind.

Nicholas Barbon was similarly clear in his writing. His major contribution, which Briscoe touched on briefly, was the concept of stock. In A Discourse of Trade, he states that "interest is the rent of stock, and is the same as the rent of land" (1690: 31). If the return from stock was the same as the return from land, then again it makes sense that Barbon should believe that land was a good asset to use as a basis of credit. Schumpeter states:

> "if the reader is to understand the history of interest theory during the nineteenth century, and some part of it even during the first four decades of the twentieth, it is absolutely necessary to realise fully what this means."
>
> (1954: 329-330)

Barbon realised that people did not want money for its own sake. They wanted what they could buy with money. Hence, he was saying that the monetary element was unimportant and paving the way for the 'real' analysis of the 19th century (Schumpeter, 1954: 330). For this reason, Arthur Monroe describes him as "the only downright supporter of the fiat theory" in the 1690s (1966: 115). Barbon also understood the importance of a high level of trade for governments, and made the link between a high level of consumption and a buoyant economy: "the chief causes that promote trade… are industry in the poor and liberality in the rich" (1690: 62). According to Vickers, making this link "places Barbon's work in the forefront of his time" (1968: 86-87). Sadly, this made no difference to the fortunes of the bank that he and Asgill set up,

and it failed in 1699 due to liquidity issues much like the others. The concepts of consumption and capital are at the centre of basic economics today. Barbon lays them out neatly in A Discourse, and he is perhaps rather overlooked in the study of the history of economics as a serious contributor to the field.

Scottish Synthesis

The last proponent of a land bank whom I wish to consider in this essay was described by Schumpeter as " in the front rank of monetary economists of all time" (1954: 295). He is referring to John Law. Where the others identified important economic problems in relative isolation, Law produced two works that were a synthesis and improvement of all of their theories. The first was an Essay on a Land Bank, discovered by Murphy relatively recently. He describes it "in many respects [as] a treatise on money and not a general theory of employment and money" (1997: 50). To this end, Law does not propose an expansion of the money supply in the Essay. Instead, he considers land backed securities as an alternative to metallic money. The major leap forward that Law makes is his definition of 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" (1994: 55). The use of the word 'by' rather than the word 'in' implies that Law views money as something distinct from commodity. Murphy sums up Law's identification of three functions of money from this definition as "(1) a measure of value, (2) a means of payment, and (3) a standard of deferred payments" (1997: 54). Law apparently leaves out the fourth function, that of the store of value. However, as the Essay is predominantly concerned with the stability of the value of money, and one function that is identified is the deferred payment function, Murphy (1997) suggests that Law surely implied this. Another addition to the definition of money that Law made was a second category of money. This had a less stable value, and performed the function of exchange or payment at a point in time, but was unable to be used in the long run as a measure of value or means of deferred payment. According to Law (1994), the reason that land would be a good security to back money is that, unlike silver, the supply of land is fixed. As this is the case, the supply of money is predictable and hence the value of money is too.

Law's Essay on a Land Bank was a relatively simple precursor to Money and Trade, his proposal to the Scottish parliament for a land bank. As the title suggests, this pamphlet links economic activity and money in no uncertain terms:

"Domestick trade depends on the money. A greater quantity employs more people than a lesser quantity. A weak'ned sum can only set a number of people to work proportion'd to it, and 'tis with little success laws are made, for employing the poor or idle in countries where money is scarce."

(Law, 1705: 13)

Law was demonstrating his macroeconomic abilities here, identifying a clear link between employment, and in a broader sense trade, and money supply. This was one of three objectives of this pamphlet. The nature of money, a second objective, was considered to a certain extent in An Essay on a Land Bank. The final objective was "the policy issue of how to produce a new monetary structure capable of expanding the money supply" (Murphy 1997: 80). His analysis of banks' effect on the supply of money gives us part of the answer to this question. He says: "the use of banks has been the best method yet practis'd for the increase of money" (1705: 36). As the quantity of money drives domestic and foreign trade, the bank performs a key role in Law's economic model. Indeed, there were many more issues that Law touched upon, such as the interventionist nature of his proposal to fiddle with foreign exchange rates. This lead Schumpeter (1954) to comment that Law gave birth to the idea of managed currency. Whether or not this is true, he certainly made the greatest contribution to monetary economics of all of the Land bank proponents.

Conclusion

None of the land banks that were proposed were embraced by the respective parliaments in the same way that the Bank of England or the Bank of Scotland were. Perhaps the politicians of the day understood the great flaw of all of the scheme: that land can fall in value just like any other asset. Notwithstanding, the contribution that the land bank projectors made to the understanding of money was significant. They began a debate on the definition, characteristics and functions of money, identifying the importance of money with respect to economic activity, and this in turn with the wealth and power of government. They linked money to the macro economy and issues such as employment. They paved the way for the conception of money as capital and opened the door to currency management.

All of these issues are relevant in the post-financial-crisis world, where the definition of money is contracting (Coggan, 2011). The debate now rages around what has real value. The creation of money in response to the credit crisis was unprecedented in terms of scale, and went some way to rein-

stalling confidence. However, quantitative easing cannot continue forever. Perhaps we will see a return of the security backed credit that the land bank projectors proposed. Whatever happens, their contribution to monetary thought is an important one, which is rather overlooked in the study of the subject.

References

Anonymous, 1696. *A Bank Dialogue between Dr H. C. Chamberlen and a Country Gentleman*. London.

Asgill, John, 1696. *Several Assertions Proved in Order to Create another Species of Money than Gold and Silver*. London.

Barbon, Nicholas, 1690. *Discourse of Trade*. London.

Briscoe, John, 1694. *A Discourse on the Late Funds of the Million-Act, Lottery-Act and Bank of England*. London.

Chamberlen, Hugh, 1696. *The Constitution of the Office of Land-Credit Declared in a Deed*. London.

Coggan, Philip, 2011. *Paper Promises*. London: Allen Lane.

Horsfield, J. Keith, 1960. *British Monetary Experiments 1650-1710*. London: London School of Economics.

Law, John, 1994. *John Law's Essay on a Land Bank*, ed. Antoin E. Murphy. London: Aeon.

Law, John, 1705. *Money and Trade Consider'd with a Proposal for Supllying the Nation with Money*. Edinburgh.

Monroe, Arthur, 1966. *Monetary Theory before Adam Smith*. New York: Augustus M. Kelly.

Murphy, Antoin E., 1997. *John Law*. Oxford: Oxford University Press.

Potter, William, 1650. *The Key of Wealth*. London.

Potter, William, 1650. *The Trades-Man's Jewel*. London.

Schumpeter, Joseph, 1954. *History of Economic Analysis*. New York: Oxford University Press.

Vickers, Douglas, 1968. *Studies in the Theory of Money 1690-1776*. New York: Augustus M. Kelly.

David Hume: Monetarist theorist?

Marcus Spray

Senior Sophister

In this insightful and rigorous study, Marcus Spray uses the prism of Monetarism to discuss the work of influential thinker and theorist David Hume. He finds much evidence for the notion of Hume as monetarist in his views on many issues, but raises questions over his views on inflation and free markets. As a result, he concludes that while Hume was undoubtedly influential in the monetarist sphere, he was perhaps not fully monetarist himself.

Introduction

In this essay, I aim to establish whether David Hume can truly be seen as a forerunner to modern monetarist thinking.

David Hume has traditionally been seen as a father to monetarist theory and has indeed been a great influence on many monetarist theorists such as Milton Friedman (Friedman, 1968; Murphy, 2009; Meyer, 1980; Wennerlind & Schabas, 2008). However, certain aspects of Hume's work such as his apparent fondness for inflation have caused academics to question this view.

Among his variety of works, Hume only published one book on economics, Political Discourses. The work is split into 13[1] different essays and they predominantly deal with the topic of macroeconomics[2]. The use of the word 'discourses' is important as Hume applies a philosophical method to his analysis of economics (Murphy, 2009, p.100). For instance, Hume poses questions such as 'is money wealth', or 'is the accumulation of money beneficial (Ibid, p.100)? However, this ambitious philosophical approach may be the reason behind many equivocations and hesitancies found in the text. Hume, the perennial sceptic, would naturally doubt and question any apparent assumptions or assertions. This inconsistency in his work has thus caused difficulty for economic historians in assigning Hume a definite place within monetarism.

1 The 13th essay – 'Of the Jealousy of Trade' – was added 6 years after the first publication.
2 Seven out of the thirteen essays deal with macroeconomic issues.

In a very narrow sense, monetarism describes the view that 'changes in the money stock are the predominate factor in explaining changes in money income'[3] (Mayer, 1978 p.1). But in a broader sense, the term covers a wide variety of propositions apart from the basic quantity theory (Ibid, 1978). Nevertheless, not all monetarists would subscribe to the same set of propositions and certainly some propositions would carry more weight than others. In analysing whether David Hume is a forerunner to monetarist theory, it is pointless to apply a modern day list in the effort to see whether the theories and ideas match up. Many monetarist ideas that are important today were not invented or were utterly irrelevant in the time of Hume. For instance, the question of what the targets and indicators of monetary policy should be would be meaningless (Meyer, 1980). Instead, it is better to look at some of the central economic ideas espoused by Hume, and ascertain whether they display monetarist aspects or indeed shaped future monetarist views.

Thus, in this essay I will first look at money neutrality and Hume's formulation of the theory. I will then analyse how Hume believed the interest rate was determined and whether this was similar to Monetarist theory. Following this I will look at Hume's view on private sector stability, free markets, and free trade, all of which are generally supported by monetarists. Finally, I will look at Hume's views on inflation, an occurrence that is not favoured by monetarists.

Money Neutrality

The neutrality of money in the long run is the central proposition to monetarist thinking. (Meyer, 1978). Hume was the forerunner of this idea – while he may not have been the first theorist to describe it, he was, according to Shrumpeter, the first to show that on an abstract level there is no specific level of nominal money that a country needs. (Mayer, 1980) Hume writes:

> "If we consider any one kingdom by itself, it is evident, that the greater or less plenty of money is of no consequence; since the prices of commodities are always proportioned to the plenty of money… It is a maxim almost self-evident, that the prices of everything depend on the proportion between commodities and money, and that any considerable alteration on either has the same effect, either of heightening or lowering the price… All augmentation (of gold and silver) has no other effect than to heighten the price of commodities and labour. Money is not, properly speaking, one of the subjects of

---

3 I.e. money neutrality.

commerce, but only the instrument which men have agreed upon to facilitiate the exchange of commodity for another."

<div align="right">(Hume, 1752, p.41)</div>

Thus, money for Hume was merely a unit of account, one that could loosen the wheels of trade but not drive it itself. Essentially, he believed money does not play a strong role in determining economic activity. In other words, Hume was subscribing to the classical theory on the neutrality of money

 Thus, Hume firmly understood the distinction between the real and nominal quantity of money which is so stressed by monetarists. (Mayer, 1980)

Interest Rate Determination

Related to this was Hume's theory of real interest rates which is very similar to the view subscribed to by monetarists (Mayer, 1980). He believed that a high stock of money does not necessarily imply low nominal interest rates, instead it would just increase prices:

> "The lowness of interest is generally ascribed to plenty of money. But money, however plentiful, has no other effect, if fixed, than to raise the price of labour… Interest in Batavia and Jamaica is at 10 percent, in Portugal at 6; through these places, as we may learn from the price of everything, abound more in gold and silver than either London or Amsterdam."

<div align="right">(Hume, 1752, p.61)</div>

This theory is remarkably similar to Friedman's theorising in his 1968 essay 'On the Role of Monetary Policy', the only differences being that Friedman has the benefit of the Fisher effect and he describes money growth rate rather than money stock:

> "High and rising nominal interest rates have been associated with rapid growth in the quantity of money, as in Brazil or Chile or in the United States in recent years, and… low and falling interest rates have been associated with slow growth in the quantity of money, as in Switzerland now, or in the United States from 1929 to 1933."

<div align="right">(Friedman, 1968, pp6-7)</div>

Furthermore, Hume attributed the interest rate to the level of profits. A low interest rate was the consequence of the growth of trade rather than a cause of it (Murphy, 2009). According to Hume, it was merchants who were crucial

to interest rate determination. These merchants[4], by stimulating economic activity and by pooling their subsequent savings, would drive down the interest rate. Hume greatly admired these merchants and felt they were the 'most useful race of men' in society (Hume, 1752, p.68). This is not unlike many monetarists who also display great gratitude to modern day merchants such as businesspeople (Meyer, 1980). Also by reversing the causal effect whereby lower interests rate lead to increased economic activity, Hume was espousing views very similar to modern day real business cycle [RBC] theorists who believe interest rates are endogenous to economic activity rather than exogenous. As RBC theory is seen as a subset of monetarism, this bolsters the case for Hume as a forerunner to monetarist thought.

Private Sector Stability

Monetarists tend to believe that the private sector is naturally stable if it is left to its own devices and not interrupted by irregular monetary policies (Mayer, 1978, p.14). In this respect, Hume was firmly in agreement. At the time, mercantilists fought to preserve a country's supply of specie in order to stabilise the private sector. Hume however was firmly against this idea, believing that the supply of specie should be left to look after itself. What is more, Hume maintained that prices are highly flexible and that this fact would ensure the stability of the private sector:

> "Nor is it probable, that the diminution of circulating money was ever sensibly felt by the people, or ever did them any prejudice. The sinking of the prices of all commodities would immediately replace it, by giving England the advantage in its commerce with the neighbouring kingdoms."

<div align="right">(Hume, 1752, p.94)</div>

Furthermore, Meyer (1980) points out that Hume was a man of great social sympathy and if he had thought unemployment was a serious problem, he would have likely discussed it at length.

Therefore, by placing firm faith in private sector stability Hume is adhering to monetarist theory.

Free Markets

Hume is often credited with following the monetarist practise of preferring the free market to government intervention. One conspicuous example is Hume's criticism of a large public debt that was proposed by his contemporaries:

4   Or Entrepreneurs as Cantillon would refer to them.

"The practice therefore of contracting debt will almost infallibly be abused, in every government. It would scarcely be more imprudent to give a prodigal son a credit in every banker's shop in London, than to impower a statesman to draw bills, in this manner upon posterity."

(Hume, 1752)

Consequently, it seems Hume would not look kindly on government intervention within the market.

Although, Viner claims that laissez faire doctrines can only be found in Hume's writings by implication, if at all (as cited in Mayer, 1980). Furthermore, Hume's free market convictions are possibly tempered by his support of taxation. He believed that mild taxation could be very positive since it stimulated industry (Mayer, 1980). His idea was that the imposition of a tax would actually increase the supply of effort, ingenuity, and enterprise, so that income rose to compensate for the tax (O'Brien, 1975, p.242). However, this must be qualified by the context of the time; most of Hume's peers advocated much harsher taxes, and Hume's support of tax was much more constrained in comparison (Humphreys, 1932).

A much more damning finding is Hume's description of spontaneous motivation found in his ground breaking philosophical work 'A Treatise of Human Nature'. Hume (1739, p. 125) claims that 'reason alone can never be a motive to any action of the will; and… that it can never oppose passion in the direction of the will'. For Hume (1739, p.127), 'reason is and ought only to be a slave of the passions'. This suggests that realms such as the market could be subject to irrational human behaviour and unintended consequences. His theorising sounds very similar to and must have been an influence on Keynes' idea of Animal Spirits:

"Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits – a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities."

(Keynes, 1932, pp.161-162)

It is hard to reconcile Hume's ideas on human motivation with him being a free marketer. In this respect, Hume sounds far more like a Keynesian than a monetarist.

Free Trade

Monetarists have generally been in favour of free trade, praising the mutual benefits experienced by all participants. In this respect it seems David Hume was firmly monetarist. Interestingly, this viewpoint was not popular in Hume's time, with most of his contemporaries espousing mercantilist ideals of protectionism and specie accumulation (Murphy, 2009). In a similar vein to how monetarists reacted against Keynesianism, economic historians have argued that Hume was reacting against the dominant mercantilist views (Meyer, 1980).

Hume's writings on trade are contained in the essays 'Of The Balance Of Trade' and 'Of The Jealousy Of Trade', which was published 6 years later. First of all, in 'Of the Balance of Trade', Hume demonstrated the self-defeating nature of specie and bullion accumulation through his price specie flow mechanism. He described how an increased money supply would lead to higher prices due to the increase in expenditure. As a result, exports would become less competitive while imports would become 'so cheap in comparison' (Hume, 1752, p. 83). Consequently, a balance of payments deficit would emerge, which would in turn cause the money supply to return to its previous level. The whole exercise thus would be completely pointless. Accordingly, Hume has been recognised for first applying monetarism to the open economy (Murphy, 2009; Wennerlind & Schabas, 2008).

However, it was left to 'Of The Jealousy Of Trade' to extoll the virtues and advantages of fully fledged free trade. Thus, Hume set out to show that trade was not a zero sum game. Open trade between nations would lead to great wealth and prosperity to all those involved:

> "That where an open communication is preserved among nations, it is impossible but the domestic industry of every one must receive an increase from the improvement of others… the increases of riches and commerce in any one nation, instead of hurting commonly promotes the riches and commerce of all its neighbours."
>
> (Hume, 1955, p.78)

It seems like Friedman (1968) was echoing Hume when he wrote in his article

'The Case for Free Trade', that 'free trade is in the best interests of trading countries and of the world'.

However, O'Brien (1975, p.36) questions Hume's free trade credentials, calling him a 'mild protectionist'. For instance, Hume feels imports on certain goods would be favourable, noting that that 'a tax on German linen encourages home manufactures, and thereby multiplies our people and industry' (as cited in O'Brien, 1975). Hume was also concerned with increasing the strength and power of Britain relative to other states. Johnson (1937, p.177) writes that although Hume considered that the 'happiness of individuals should be the purpose of policy, the greatness of the state should always be regarded as an even more important goal'.

Regardless, Hume clearly believes in the benefits of free trade and is only willing to accept trade protection for certain budding industries (Wennerlind, 2005). Hume, therefore, was predominantly a monetarist when it came to free trade and was a huge influence on the monetary approach to the balance of payments.

Inflation

The most prominent problem with Hume's supposed monetarism is his view on inflation. Unlike most monetarists, Hume did not display any great concern about inflation. He failed to acknowledge any great losses with inflation and felt that it was an important ingredient in driving the economy (Meyer, 1980). In a famous passage he writes:

> "The good policy of the magistrate consists in only keeping it, if still possible, still encreasing; because, by that means, he keeps alive a spirit of industry in the nation, and the encreases the stock of labour… a nation whose money decreases, is actually at that time, weaker and more miserable than another nation, which possesses no more money, but is on an encreasing hand."
>
> (Hume, 1752, p.50)

While monetarists have at times argued for an expansion in the short run, Friedman for example refers to a money illusion whereby consumers feel wealthier than the reality and therefore increase expenditure, they do maintain that the benefits will be short-lived. Hume on the other hand, while he also refers to this illusion in monetarist fashion, seems to argue for a permanent expansion in the money stock. This apparent faith in the long term advantages of inflation seems to contradict his views on the quantity theory.

By claiming a monetary expansion can have positive effects on

employment and output, Hume, according to Pearlman, writes 'one of the most controversial passages' in his economic writings (as cited in Wennerlind, 2005). The consensus therefore is that Hume favoured the policy of maintaining a gradually increasing monetary policy – thus, an inflationist policy. An idea that would run in direct opposition to the monetarists of the 20th century.

However, Wennerlind (2005), an expert on David Hume's economic writings, does not believe Hume should be labelled as an inflationist. By recognizing that Hume made an analytical distinction between endogenous and exogenous money supply, Wennerlind believes this inconsistency can be solved. Thus, a rise in the money supply should occur only if it is preceded by an increase in industry. Therefore when Hume claims that the magistrate's good policy consists of, if possible, increasing the money supply, he is really advising the state to promote industry through the creation of favourable laws. Hence, if read differently the passage does not imply that Hume is an inflationist and thus maintains his monetarist credentials.

Then again, while Wennerlind makes a strong case, his supposition is unconvincing as it does not take into account the gap between Hume's understanding of the economy and the underlying financial reality of the world they lived in (Murphy, 2009). For instance, in 'Of Public Credit', Hume was very inaccurate in his understanding of the financial innovations and the financial services sector which would later become an essential part of the British economy, and also made a foolish prediction that either Britain should 'destroy paper credit or public credit will destroy the nation' (Hume, 1752, p.135). Also, when Wennerlind (2005) makes his argument he presupposes that the law of one price holds. But as Murphy (2009), and Wennerlind and Schabas (2008) show by identifying the correspondence between Oswald and Hume, Hume does not recognise this law.

I would conclude therefore that Hume did favour an inflationist policy in a manner unlike Monetarists.

Conclusion

David Hume succeeded in many realms of intellectual thought. He published respected books in philosophy, political science, history, and as well of course economics. His macroeconomic work had a profound influence for years to come, even prompting Friedman to comment in 1975 that economists had learnt little since Hume (as cited in Wennerlind & Schabas, 2008).

This essay sought to establish as to whether Hume can be truly credited as being a forerunner to modern monetarist thought. Certainly, his formulation of money neutrality exerted a huge influence on monetarism,

becoming its central proposition. Hume's idea of interest rate determination was also very influential, as was his price specie flow mechanism, his belief in the stability of the private sector, and his preference for free trade. However, as I have demonstrated in this essay, his less than convincing faith in the free market and his fondness for inflation challenge his monetary pretensions. As discussed in the introduction, Hume's philosophical method also undermined his ideology and thus made the 'sceptic' hard to place within any school. Regardless, even with these caveats, Hume's enormous influence within monetarism should not be ignored. While not always monetarist in detail, Hume undoubtedly was a forerunner to many monetarist ideas.

References

Friedman, M. & Friedman, R. (1997, Oct. 30). The Case for Free Trade. *Hoover Digest*. Retrieved from: http://www.hoover.org/publications/hoover-digest/article/7125.

Friedman, M., 1968. The Role of Monetary Policy. *The American Economic Review.* Vol. 58 (Mar).

Hume, D., 1739. *A Treatise of Human Nature*. Edinburgh.

Hume, D., 1752. *Political Discourses*. Edinburgh.

Hume, D., 1955. *David Hume Writing on Economics*. Ed. Rotwein, E. London: Nelson.

Humphrey, D., 1932. *David Hume: Economist*. University of California, Berkeley.

Johnson, E. A. J., 1937. *Predecessors of Adam Smith*. London: P.S. King and Son

Keynes, J. M., 1936. T*he General Theory of Employment, Interest and Money*. London: Macmillan.

Mayer, T., 1978. *The Structure of Monetarism*. New York: Norton & Company.

Mayer, T., 1980. David Hume and Monetarism. *Quarterly Journal of Economics*. Vol. 95 (Aug.).

Murphy, A. E., 2009. *The Genesis of Macroeconomics*. Oxford: Oxford University Press.

O'Brien, D.P., 1975. *The Classical Economists*. Oxford: Clarendon Press.

Perlman, M., 1987. Of a Controversial Passage in Hume. *Journal of Political Economy*. Vol. 95 (April).

Viner, J. (1937). *Studies in the Theory of International Trade*. New York:

Harper and Row.

Wennerlind, C. & Schabas, M. (Eds.). (2008) *David Hume's Political Economy*. New York: Routledge.

Wennerlind, C. (2005). David Hume's Monetary Theory Revisited: Was he Really a Quantity Theorist and an Inflationist? *Journal of Political Economy*, Vol. 113/1.

From Thornton, to Bagehot, to the European Central Bank:

An Analysis of the Lender of Last Resort

Iain Snoddy
Senior Sophister

With the escalation of the euro area's sovereign debt crisis, calls for the European Central Bank to take a more active role have increased. Iain Snoddy looks at the history of economic thought with regard to central banks and applies this to the ECB of today, arguing that the euro area would benefit greatly should it begin to act as a lender of last resort.

Introduction

At present the future of the world economy looks bleak. The future of Europe looks particularly troubling,; with no end in sight for the sovereign-debt crisis that has engulfed the region since late 2009. Commentators have largely criticized European leaders for their failure to act decisively to stem the crisis. Much criticism has also been levelled at the European Central Bank (ECB) with many economists condemning its rigid adherence to the sole mandate of price stability and its refusal to act as a lender of last resort (LLR) to ailing sovereigns. This paper aims to address whether there is a role for the ECB in this regard. In the first section I shall analyse the initial conceptualisations of the LLR function, as put forward by Henry Thornton (1760-1815) and Walter Bagehot (1826-1877). Subsequently, I shall assess the adherence of the ECB to the classical view when acting as a LLR to the European banking sector. Finally, I shall argue that the ECB should act as a LLR to sovereigns and determine whether the advice of Thornton and Bagehot is useful in this regard.

Thornton & Bagehot: The Classical View

The first known conceptualisation of the LLR can be found in Henry Thorn-

ton's An Inquiry into the Nature and Effects of the Paper Credit of Great Britain, first published in 1802. In Paper Credit, Thornton addresses issues pertaining to monetary economics and does much to counter the orthodox thought of earlier writers. Notably he makes continual reference to the writings of Adam Smith. Murphy notes that Thornton had three main problems with Smith's monetary economics: his limited definition of the money supply, his real bill of exchange doctrine and his neglect of the role of the velocity of circulation (2008). However, despite Thornton's clear break with the classical tradition, it is impossible to accurately characterise his views into one school of thought. This is explained by the contrasting arguments presented in Paper Credit. Murphy highlights the apparent transformation from the 'anti-deflationist' arguments of Thornton to those of the 'hard currency man' (2003). Despite this seeming inconsistency, Paper Credit is a landmark work and contributes much towards monetary thought.

Thornton's most important contribution is his extensive analysis of the role of the Central Bank. He distinguishes between an 'external drain' of specie to foreign nations, and an 'internal drain' whereby panic-stricken domestic residents would increase their gold holdings through the conversion of paper currency into specie (Humphrey, 1975). In the latter case, Thornton warns against the effects of deflationary policies through a reduction of the note issue, as recommended by Adam Smith:

> '...however just may be the principle of Dr. Smith when properly limited and explained, the reduction of the quantity of Bank of England paper is by no means a measure which ought to be resorted to on the occasion of every demand upon the bank for guineas arising from the high price of bullion, and that such reduction may even aggravate that sort of rise which is caused by an alarm in the country.'

> (1802: 104)

Furthermore, he criticises the Bank of England for adopting the recommendations of Smith:

> 'If there has been any fault in the conduct of the Bank of England, the fault, as I conceive, has rather been, [...] on the side of too much restricting its notes in the late seasons of alarm, than on that of too much enlarging them. In doing this, it has happened to act (though in part) according to what seems likely to have been the

advice of DR. A. Smith ...'

(1802: 127)

In arguing against a deflationary restriction of paper credit, Thornton develops the concepts of price stickiness (Humphrey, 1989), liquidity preference (Hayek, 1939) and the marginal efficiency of capital (Murphy, 2008), preceding Keynes by over a century. Subsequently, he recommends an expansion of the note issue to protect the domestic economy from a drastic fall in the money supply (Humphrey, 1975).

Walter Bagehot addressed the topic of the LLR in his now seminal work, Lombard Street. To say that Bagehot further developed thinking on the subject would be untrue; throughout Lombard Street, much of Thornton's contributions are simply 'revised and restated' (Humphrey, 1989: 12). The main arguments made by Bagehot and Thornton are summarised as the 'classical view', which proposes that the LLR should lend freely to solvent institutions with good collateral, or in Bagehot's words, 'lend to all that bring good securities, quickly, freely, and readily' (1873: 173). The classical view further states that the Central Bank should make clear its willingness to lend (Bordo, 1989) and that '... these loans should only be made at a very high rate of interest' (Bagehot, 1873: 197). Some authors, such as Humphrey (1975; 1989), would also point out that Bagehot implied that lending should occur at penalty rates. However, Goodhart successfully debunks this notion, stating that while Bagehot implies raising lending rates during a crisis, he does not suggest raising these rates above new post-crisis market interest rates (1999).

Despite remarkable coherence between the authors' understanding of the role and position of the Bank of England, their views on the desirability of such a position diverge markedly. Laidler notes that Bagehot did not envisage the position held by the Bank of England as an optimum solution but rather as one generated by an 'historical accident' (2002). This stands in opposition to Thornton, who believed in the inherent desirability of the pivotal role of the Bank of England (Ibid, 2002). It is also important to highlight that Bagehot's analysis was driven by his concern for protecting gold convertibility (Ibid, 2002). In stark contrast, Thornton emphasised the importance of preserving the quantity, and hence the purchasing power, of the money stock (Humphrey, 1989).

Developments

Despite acting as a LLR to the financial system in 1847, 1857 (O'Brien, 2003) and again during the Overend Gurney Crisis in 1866 (Bordo, 1989), widespread acceptance of the bank's LLR function did not emerge until the mid

1870s (Fetter, 1965). Having accepted its role as a LLR, the Bank of England was able to prevent financial criseis developing into deep recessions in 1878, 1890 and 1914 (Bordo,1989). Most other nations had also developed LLR capacities at the end of the nineteenth century (Ibid, 1989), although development of an effective LLR in the United States did not occur until much later. The US Federal Reserve was not formed until 1914 and proved inexperienced and unsure of its role when faced with its first banking crisis in the wake of the 1929 stock market crash. The Federal Reserve's failure to support the domestic banking sector during crises in 1930, 1931 and 1933 led to the failure of 'more than one-fifth of the commercial banks in the United States holding nearly one-tenth of the volume of deposits at the beginning of the contraction...'(Friedman and Schwartz, 1966, p.299). The creation of federal deposit insurance in 1934 restored calm to the banking sector and assured the public of the safety of bank deposits. However, according to Schwartz, deposit insurance is 'not essential to prevent panics, given a responsible lender of last resort' (1987); deposit insurance therefore was required only to account for the Federal Reserve's inaptitude. However, the Fed has gradually accepted its LLR role and since 1970 has erred on the side of excessive action (Bordo, 1989).

The ECB and the Banking Sector

In 2008, following the collapse of Lehman Brothers in the United States, the world economy was faced with a deep global financial crisis, prompting a global recession. Central Banks, including the ECB, acted quickly, providing liquidity support and engaging in open market operations (OMO). The ECB slashed the main refinancing rate by 325 basis points between October 2008 and May 2009 (Trichet, 2010). Further measures taken are well outlined by Trichet (2010), notable among these is the unlimited liquidity lending to euro area banks, although these provisions have been phased out since December 2009 (2010). In December 2011 the ECB provided further support to the European banking sector through the issue of just over 500 billion Euros in three year loans.

Generally, the ECB's actions vis-a-vis the European banking sector are consistent with the classical view of the LLR; the bank has engaged both in OMO and has provided loans directly to the banking sector with the requirement of good collateral to separate the illiquid institutions from the insolvent. However, the one percent interest rate charged on loans in December 2011 hardly reflects 'at a penalty'. Despite this action, governments still advanced large bailout packages to ailing banks. It is likely that insolvency prompted these institutions to seek domestic government support,

rather than risk refusal by the ECB. These large bailouts served to worsen the fiscal position of European nations, most notably Ireland, while other nations such as Greece proved fiscally irresponsible prior to the financial crisis. In late 2009, concerns over euro area sovereign debt reached new heights as market yields on government bonds issued by the PIGS (Portugal, Ireland, Greece, Spain) jumped substantially. The crisis reached a new phase following market speculation over the state of the Italian economy.

The ECB and the Sovereign Debt Crisis

Throughout the sovereign debt crisis the ECB has consistently stated its reluctance to provide support through intervention in government bond markets. While the ECB did engage in bond purchases in late 2011, these actions were incredibly limited and were made alongside the self-defeating declaration that such stabilization measures would have only a temporary effect (Delong, 2011). The resignation of Jürgen Stark in protest also sent a signal to the market of the fierce resistance to such moves, as did the bank's 'sterilization' of bond purchases, reflecting a reluctance to forgo low inflation. The ECB did encourage the purchase of government bonds using the funds advanced to banks in December 2011, achieving some success in lowering bond yields. However, this indirect action signalled to markets the extent of the ECB's reluctance to get directly involved. Resulting from the ECB's inaction the EU was forced to create the European Financial Stability Facility (EFSF) and the European Stability Mechanism (ESM) to provide LLR support to struggling sovereigns. Along with the IMF, these institutions provided bailouts to Greece, Ireland and Portugal with the ECB acting in an oversight role through the troika arrangements.

De Grauwe (2011) notes that a sovereign debt crisis in a monetary union has many similarities with a domestic banking crisis; both lead to fears of contagion and speculation on illiquid but otherwise solvent institutions. This can lead to the self-fulfilling collapse of banks through runs, or the bankruptcy of nations through large increases in borrowing costs. The role of a LLR in the sovereign debt market is much the same as that of a LLR in the banking sector and as such, the central bank who already acts in this capacity is the main candidate for support to sovereigns. Consequentially an ECB bond purchasing program would lower the market interest rate on bonds, lowering borrowing costs for sovereigns and stemming the risk of a self-fulfilling crisis. The depth and speed of austerity measures could be reduced as borrowing costs are lowered and nations no longer need to convince the market of their solvency. Furthermore, through the prevention of sovereign default the ECB would serve to protect the European Banking sec-

tor, which is greatly exposed to euro area sovereign debt. Bond purchasing would also prevent the need for the 'spectacle' of state bailouts which harm market expectations. It is also worth noting that by simply announcing its willingness to act as a LLR the ECB could do much to sooth market fears.

The classical doctrine would suggest that the ECB should lend freely, at a high rate and to only solvent institutions with good collateral. Thornton and Bagehot both suggest that a LLR is not required to prevent crisis but should act quickly to prevent contagion. Unfortunately the ECB has acted slowly and as such the depth of action required at present is much greater than would have been required in late 2009. The ECB also faces a number of problems in enacting these principles. Firstly, lending freely to sovereigns could generate inflation within the euro area although it could be argued that the bleak growth prospects of Europe have assured low inflation for years to come.

Defining good collateral also presents a problem. Growth prospects could be used to measure future ability to repay, although estimates are prone to error and bias. Distinguishing between solvent and insolvent institutions presents a further issue, and one in which there is little agreement between economists. However, few would argue that the Greek position is tenable. The introduction of high or penalty rates to determine between those solvent and insolvent as suggested by De Grauwe (2011) is also largely impractical as the raison d'être of ECB intervention is to lower the interest charged on bonds and hence borrowing costs. This also prompts a further question; should the ECB refuse to lend to insolvent institutions? In answering this question it is important to note that refusing to lend would not only allow the sovereign to default, but might in fact bring about default through a complete collapse of market confidence.

Conclusion

There are of course a number of valid reasons for the ECB's refusal to act as a LLR to sovereign nations. The bank may be concerned about moral hazard through providing a safety net for imprudent sovereigns. The ECB may also fear a loss of credibility through subordinating its inflation goal. The legality of such action is also under question. However, the consequence of inaction is to further risk the stability of the European financial system and the European Monetary System itself. The remarkable similarity, and indeed interconnectedness between banking and sovereign debt crises in a monetary union means that the position of the ECB in acting as a LLR to both banks and governments is a 'natural' one. In performing these roles the ECB would do well to remember the wisdom of Thornton and Bagehot, in particular the ECB should

take swift decisive action to calm markets. Furthermore, the ECB should not underestimate the benefits to be gained from announcing their support; the action required of them might be much less than they would imagine.

References

Bagehot, W., 1873. Lombard Street: *A Description of the Money Market*. New York: Sribner, Armstrong. Edited with an Introduction by P.L. Bernstein, 1999. New York; Chichester: John Wiley and Sons Inc.

Bordo, M.D., 1989. The Lender of Last Resort: Some Historical Insights. NBER Working Paper No.3011.

De Grauwe, P., 2011. The European Central Bank: Lender of Last Resort in the Government Bond Markets? CESifo Working Paper No 3569.

Delong, J.B., 2011. The ECB's Battle Against Central Banking. *Project Syndicate,* [online] Available at
< http://www.project-syndicate.org/commentary/delong119/English> [Accessed 31 December 2011]

Fetter ,F.W., 1965. *Development of British Monetary Orthodoxy*. Massachusetts: Harvard University Press.

Friedman, M., and Schwartz,A.J., 1966. *A Monetary History of the United States 1867-1960*. Princeton: Princeton University Press.

Goodhart, C.A.E., 1999. Myths about the Lender of Last Resort. *International Finance,* 2:3:339-360.

Hayek, F.A. von., 1939. Introduction to Thornton (1802) in F.A. von Hayek (ed.) *Henry Thornton: An Enquiry into the Nature and Effects of the Paper Credit of Great Britain.*

Humphrey, T.M., 1975. The Classical Concept of the Lender of Last Resort. Federal Reserve Bank of *Richmond Economic Review*, 61:2-9.

Humphrey, T.M., 1989. Lender of Last Resort: The Concept in History. Federal Reserve Bank of *Richmond Economic Review,* 75:8-16.

Laidler, D., 2002. Two Views of the Lender of Last Resort: Thornton and Bagehot. In: *le Forum, le LED-EPEH and le Phare, Conference on the "Lender of Last Resort"*. Paris: France 23-24 September 2002.

Murphy, A.E., 2008. *The Genesis of Macroeconomics*. Oxford: Oxford University Press.

Murphy, A.E., 2003. Paper credit and the multi-personae Mr. Henry Thornton. *European Journal of the History of Economic Thought,* 10:3:429-453.

O'Brien, D., 2003. The Lender-of-Last-Resort Concept in Britain. *History of Political Economy,* 35:1:1-19.

Schwartz, A.J., 1987. The Lender of Last Resort and the Federal Safety Net. *Journal of Financial Services Research*, 1:1:1-17

Thornton, H., 1802. *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain London*: Hatchard. Edited with an Introduction by F.A. von Hayek, 1939. London: Allen and Unwin.

Trichet, J., 2010. State of the Union: The Financial Crisis and the ECB's Response between 2007 and 2009. *Journal of Common Market Studies*, 48:7-19.

THE MODIFIED GENERAL EQUILIBRIUM APPROACH TO KEYNESIAN ECONOMICS

GRAEME O'MEARA
Senior Sophister

In this especially timely and relevant essay, Graeme O'Meara analyses Keynesian economics through its departures from classical general equilibrium theory. The paper discusses the notion that Keynes' ideas exposed three major failings in the classical theory: the Walrasian auctioneer approach, a refusal to account for informational deficiencies, and the insistence on exclusively equilibrium-based theory. He concludes that perhaps the search for more rigorous, mathematical and micro-founded theory than that of Keynes has done macroeconomics more harm than good.

Introduction

John Maynard Keynes' 'General Theory of Employment, Interest and Money' (1936) is considered by many to be the one of the most profound contributions to economic thought. Notwithstanding the early contributions of such maestros as Petty, Law, Cantillon and Hume on what would now be classified as macroeconomic issues, the birth of modern macroeconomics as a coherent and systematic approach to aggregate economic phenomena can be traced to the publication of the *General Theory* (Snowdon and Vane, 2005). For decades following its publication, it ignited great debate and controversy amongst economists and policymakers. Keynes' objective upon retreating 'into his ivory tower at King's, was to embark on a supreme intellectual effort to save Western civilisation from the engulfing tide of barbarism which economic collapse was bringing about.' (Skidelsky: 1992: xxvii) This entailed a revamping of the orthodox economic theory, the postulates of which were applicable 'to a special case only and not to the general case' (1936: 3). In the 'long struggle for escape' Keynes set out to liberate economists from the intellectual confines that left them unable to deal with the Great Depression; confines created for the most part by what Keynes dubbed 'classical economics.'

However, the lack of precision and use of mathematical appliances to clarify his ideas would leave Keynes' tour de force susceptible to multiple interpretations. The most prominent interpretation was that initiated by Sir John Hicks (1937), which led to the renowned 'Income-Expenditure' model. This was later modified and extended by Modiglianni (1944), Samuelson (1948), Klein (1947) and Hansen (1953) which led to Keynesian economics being associated with such building blocks as the multiplier, the consumption function, liquidity preference theory and the marginal efficiency of capital. The ongoing debate between Keynes and the Classics was finally reconciled in the Neoclassical Synthesis, a truce which placed the *General Theory* as a special case of classical theory with restrictive assumptions on the latter, while maintaining that this Keynesian special case was nonetheless important as it was more relevant to the real world than general equilibrium theory (Leijonhufvud, 1967). This fusion of ideas proved unsatisfactory for subsequent scholars in what emerged as a counter attack on the Keynesian revolution, which strived to re-establish Keynes as an academic professional and theoretical innovator in his own right. In what Coddington (1983) brands as 'reconstituted reductionists,' Robert W. Clower (1965) and Axel Leijonhufvud (1968) reaffirm Keynes as a theorist engaged in the dynamics of disequilibrium processes. In this paper, I review and reappraise the work of the aforementioned scholars by highlighting what they perceived to be the true essence of the *General Theory*.

Reconstituted Reductionism

Classical theory analyses markets on the basis of choices made by individual traders, with the resultant theory operating at the level of individual choice and market phenomena; Coddington (1983) refers to this as 'reductionism,' the act of reducing market phenomena to (stylised) individual choices. Reductionist theorising confines its focus to situations of market equilibrium[1], a situation which makes choice theory relatively straightforward: there may be a gap between market demand and market supply, but the choice theory from which these respective schedules have been derived presumes all choices are realisable (Coddington, 1983). The work of Clower (1965) and Leijonhufvud (1968) established Keynesianism is a type of 'reconstituted reductionism' because it addresses not the problem of equilibrium, but the issue of attain-

1 Fundamentalist Keynesian Joan Robinson (1953-4) suggests that if the idea of equilibrium is followed relentlessly, then as the concept becomes all embracing, it becomes paralysed by its own logic: equilibrium becomes a state of affairs strictly unapproachable – unless it already exists, there is no way of altering it. (Coddington, 1983)

ing it. It poses the dilemma of how a decentralised market economy can, in the absence of the Walrasian auctioneer, generate an equilibrium vector of prices that clears all markets. To make the transition from Walras' world to Keynes world, the tâtonnement mechanism must be dispensed with and the auctioneer made redundant. Transactors have to become price makers and the control mechanism that prohibited sub optimal trades is no longer in force, since non-equilibrium prices may become established. Choice logic remains the same however – maximisation of utility and profit and belief in the un-trammelled workings of the price mechanism as a coordinating device. 'To be a Keynesian, one need only realise the difficulties of finding the market clearing vector.' (Leijonhufvud: 1967: 405)

The Keynesian Counter Revolution

Clower (1965) reignited the old 'Keynes and the Classics' debate by exposing theoretical flaws within the Keynesian revolution, and more generally in economic theory. Keynes challenged the Classics inter alia on two counts: firstly, its failure to recognise involuntary unemployment: 'if the classical theory is only applicable to the case of full employment, it is fallacious to apply it to the problems of involuntary unemployment,' (1936: 16)[2] and secondly, its unequivocal reliance on Walras' Law and the fictional auctioneer as the coordinator of economic activity. Clower proposed an ultimatum by inferring that 'either Walras' Law is incompatible with Keynesian economics, or Keynes had nothing fundamentally new to add to orthodox economic theory.' (1965: 278) If Walras' Law is compatible with Keynes, then 'literature on monetary theory makes it perfectly evident that Keynes may be subsumed as a special case of the Hicks-Lange-Patinkin theory of tâtonnement economics' (1965:279) which confirms that Keynes added nothing new to the existing orthodoxy. Alternatively, if Walras' Law is discordant with Keynesian economics, then the established theory of household behaviour was incompatible with Keynes; this would mean that in the general case, there exist market excess demand functions that include prices and quantities that would not satisfy Walras' Law, except in cases of full employment. (Clower, 1965)

Clower argued the latter case, claiming that there has been a 'fundamental misunderstanding of the formal basis of the Keynesian revolution' (1965:280) by pointing out that traditional price theory assumes market excess demands are independent of current market transactions, which would imply that income magnitudes are not independent variables in the demand or supply functions of a general equilibrium model; they are choice variables

---

2 With humorous reference to Euclidean geometers in a non Euclidean world.

for maximisation. Incomes are denoted in quantities and prices, and quantity variables never appear explicitly in market excess demand functions of traditional theory. This would then imply the Keynesian consumption function (which depends on disposable current income) and other market relations involving income as an independent variable cannot be derived explicitly from any existing theory of general equilibrium.

Notional v. Effective Demand

Clower (1965) was the first to draw the distinction between realised (effective) and planned (notional) magnitudes. Notional supply/demand refers to transactions that transactors would exercise at equilibrium prices, conditional on being unconstrained by an inability to buy or sell at the prevailing prices in any other market. Effective magnitudes refer to actual supply/demand backed up by an ability to pay. Notional and effective magnitudes are equal when actual transactions are unconstrained by actual transactions in another market. For Clower, his notional demand for champagne is his effective demand once he can sell as much of his economic consulting services as he desires. However, if he is constrained in selling his consulting services in the labour market, his subsequent fall in income will result in a curtailed effective demand for champagne, below that of his notional demand: 'the other side of involuntary unemployment would seem to be involuntary under-consumption.' (Leijonhufvud: 1968: 69) In such a world as this, the decision to sell is not automatically transformed into a decision to buy, since the sale has first to be realised before a purchase is made; this is the essence of Clower's 'dual decision hypothesis': planned (notional) purchases will not be made unless planned sales have been realised (made effective) (Snowdon et. al, 1994). Clower's income appears as an argument in his effective champagne demand function, contrary to the notional demand (and supply) functions of the Walrasian system which posits that incomes do not appear along with prices as independent variables. In Walras' world, transactors are constrained by money wages and prices, and given these constraints, they choose how much they desire to work and consume; in equilibrium, they are able to perform their desired level of labour and receive chosen their income. Clower maintains that Keynes added an additional constraint whereby transactors may be constrained in the level of labour they are able to sell, causing actual income to diverge from desired income.

When transactors are constrained in one market, this is likely to generate an 'income constrained process,' whereby as false prices constrain sales and production of certain commodities, their producers suffer income

cuts and find their effective demands reduced, hurting other sectors, and so invoking a ripple effect throughout the economy (Leijonhufvud, 1968). A contagion breaks out where a drop in production in the sector first thrown out of equilibrium spawns a drop in real buying power and in the real demand for outputs of other sectors and so on (Rabin, 2004). The bursting of the Irish property bubble in 2007 presents a fine example of this ripple effect: transactors involved in this sector 'felt the pinch' when demand for property related goods and services (construction, architecture, finance, landowners, raw materials manufacturers etc) dried up, which cumulatively pushed the Irish economy into recession.

Thus, from a disturbance to the economy, there ensues a downward spiral of income, far below the level to which it was brought by the initial shock; 'this deviation amplifying feedback loop is characteristic of Keynesian quantity adjustment models.' (Leijonhufvud: 1968: 53) Leijonhufvud argues that once deviation-amplifying feedbacks take hold, the deviation-counteracting prices adjustments will be less effective; because of the downward spiral of income, prices will need to fall below the new equilibrium price to jolt the system back toward full employment. Only in instances of perfectly flexible prices would income constrained processes be nonexistent.

Clower maintains that the demand functions of orthodox theory do not provide relevant market signals, because they do not distinguish between effective and notional demands by incorporating into the theory the constraint effective current receipts may place upon notional current consumption. Put simply, Keynes brings current transactions and the effect of constraints on realised transactions into price theory, where traditional theory leaves it out. And while general equilibrium theory is a useful instrument for thinking about abstract economic problems, the danger, in having 'schooled ourselves so thoroughly in the virtues of elegant simplicity (is that) we may refuse to recognise the crucial relevance of complications that do not fit our theoretical preconceptions.' (Clower: 1965: 295) This serves as a reminder of the recent controversy in macroeconomics of shaping circumstances around models:

> "Economists can become seduced by their models, fooling themselves that what the model leaves out does not matter. It is, for example, often convenient to assume that markets are "complete"—that a price exists today, for every good, at every date, in every contingency. In this world, you can always borrow as much as you want at the going rate, and you can always sell as much as you want at the going rate. "

> (The Economist, 2009)

Unemployment Disequilibrium

The reaction of real output and employment to fluctuations in the money rate of aggregate demand is a problem of short run dynamics; in the literature much of the analysis starts at a full employment equilibrium and then scrutinises the adjustment after the disturbance. Leijonhufvud (1968) examines the nature of the disequilibrium situation at some point within the time interval before returning to equilibrium. Despite using comparative static period analysis (which reflected Marshallian overtones), Leijonhufvud believes that 'Keynes' model was static, but his theory dynamic.' (1968: 62) Fundamental Keynesian[3] G.L.S. Shackle expresses this proposition with characteristic elegance:

> "At each curtain rise, the *General Theory* shows us, not the dramatic moment of inevitable action, but a tableau of posed figures. It is only after the curtain has descended again that we hear the clatter of violent scene shifting."
>
> (1967: 182)

Classical economic theory asserts that adjustments to changes in money expenditure (aggregate demand) will be via price adjustment, i.e. the benevolent brewer will likely adjust his prices in response to fluctuations in the demand for beer. This fall in prices would then eradicate the reduction in aggregate demand via the Pigou effect. However, in the Keynesian macrosystem, the Marshallian ranking of price and quantity adjustment velocities is reversed; in the shortest period, flow quantities are freely variable, but one or more prices are given, which limits the range of variation for the rest of the prices (Leijonhufvud, 1968). Thus, quantity adjustment is a synonym for a change in real income. In Keynes' world, trade continues at disequilibrium prices, so that changes in the quantity exchanged at these disequilibrium prices are the significant sources of adjustment. Prices are not rigid because of the assumption of complete price rigidity, but from the assumption that trade continues at disequilibrium prices. Transactors establish reservation prices for their labour (and commodities) and in the face of an information deficiency they search for the best price they can get based on past experience. For Leijonhufvud, this reversal of adjustment velocities was the revolutionary innovation in the *General Theory* and is comprehendible only in light of how deeply entrenched Marshallian dynamics were in the thinking of Keynes'

---

3   Fundamentalists regard the influence of unstable expectations due to uncertainty as a key element of Keynes' work, pointing to chapters 12 and 17 of the *General Theory* and to an article Keynes wrote in 1937 to reinforce this approach.

contemporaries.

Much of the preceding analysis can be illustrated graphically (albeit with the use of hydraulic Keynesian apparatus) in Figure 1: suppose there is a market-clearing vector of prices such that we are at a point of general equilibrium. A disturbance, for example in the form of a pessimistic revision of businessmen's expectations, would reduce investment: the IS curve shifts to IS_1. The archetypal producer-employer has less than perfect information in that he is uncertain as to whether this endogenous shift in demand is specific to his product or economy wide, temporary or permanent. In trying to determine these factors, it may be rational to leave price unaltered in the meantime; until information to the contrary presents itself, he may contend that the prevailing price for his bottle of champagne is the right one. The producer's initial response is to leave price unchanged and adjust quantities: this reduction in output will result in a fall in labour demand via the production function. However, this does not guarantee a shift of the labour demand curve; instead actual (effective) demand lies off the curve, and as no change in prices or wage rates have occurred, the real wage rate remains at original level and so the labour market is in disequilibrium at A. This illustrates Clower's distinction between notional and effective magnitudes: at output Y/P_1 effective demand for labour is only N_1, the discrepancy being the difference between notional N_f and effective demand. For an economy not in general equilibrium, the price mechanism does not convey the information to restore full employment. In the absence of perfect knowledge, the prescription of infinite price velocity disappears. For Keynes, trade continues at disequilibrium prices, so that changes in quantity exchanged at these disequilibrium prices are the significant sources of adjustment.
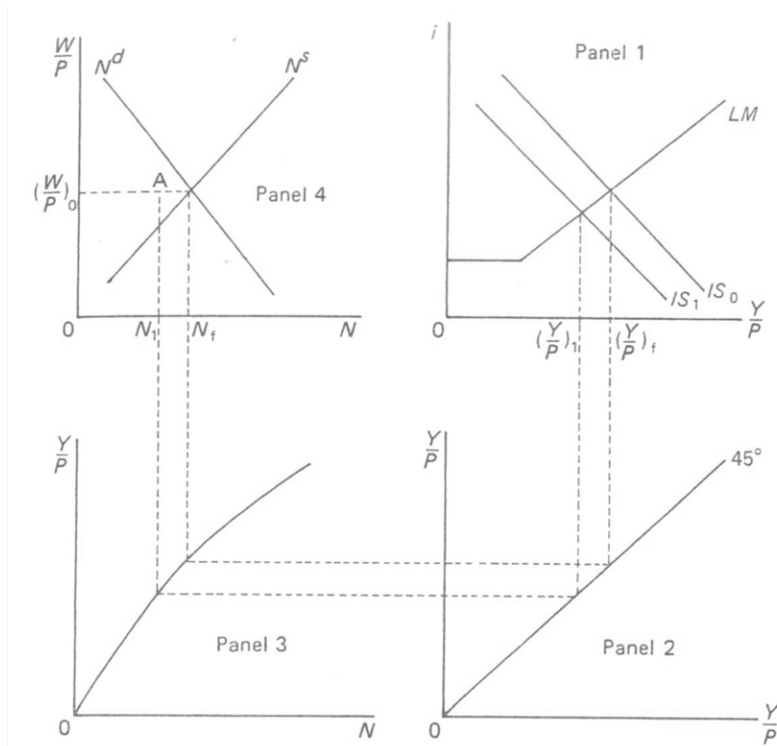
Figure 1: Unemployment from deficiency of effective demand (extracted from Pierce and Tysome, 1985)

Textual Evidence

In chapter three of the *General Theory,* Keynes explicates his principle of effective demand, that is, the point at which the aggregate demand function intersects the aggregate supply function, where entrepreneurs expectations of profits and so employment is maximised. There is however, no evidence of a principle of notional demand. And although Clower finds little textual evidence of a dual decision hypothesis in Keynes' writings, he claims that indirect evidence is plentiful: Keynes' treatment of the orthodox theory governing household behaviour, discussions of Say's Law, the consumption function concept, his account of interest theory and discussions of wage and price determination. Clower advocates also that unless the orthodox theory of household behaviour is modified to recognise the dual decision hypothesis, the aggregate consumption function does not make sense, the distinction

between transactions and speculative balances is meaningless, the liquidity preference theory of interest is indistinguishable from the classical theory of loanable funds and excess supply in the labour market will not diminish effective excess demand elsewhere in the economy. 'In short, either Keynes had a dual decision hypothesis at the back of his mind, or most of the *General Theory* is theoretical nonsense.' (Clower: 1965: 290) This type of reasoning led Coddington to comment:

> "The picture here seems to be one of Keynes with a mind full of ideas *some* of which he got onto the pages of the *General Theory,* the task being to work out what the remainder must have been. This is a problem of reading not so much between the lines as off the edge of the page."
>
> (1983: 106)

It is interesting to note also that Walras is mentioned only once the in *General Theory* (Ch. 14: 'The Classical Theory of Interest' pp.176) and the Walrasian system was a post-Keynes product. Leijonhufvud (1976) later contends that Keynes' struggle to escape was from the Cambridge of Marshall and Pigou, whilst his and Clower's aim was to debunk elements of Walras to re-establish Keynes' message. In this respect, the arguments of Clower and Leijonhufvud are only really addressed to the neoclassical synthesis debate; they tell us very little about the actual Keynes episode, since they fail to separate the economics of the classics from the classical economics fictions created after Keynes (Snowdon et. al, 1994). However, in a comment on a paper by Leland B. Yeager concerning monetary disequilibrium, Leijonhufvud (1986) retorted Yeager's criticism of his and Clower's interpretation of Keynes by stating that:

> "Now, although "what Keynes really meant" is not at all as good and useful a question as, for instance, "could macroeconomics have evolved along a more fruitful path from the General Theory," it so happens that on these particular points we now do know precisely what he meant. Volume 29 of Keynes's Collected Papers, which appeared only in 1979, contains outlines and drafts of introductory chapters (pp. 63—102) that Keynes eventually discarded in favour of his brief and cryptic chapter 2. This material leaves absolutely no doubt whatsoever that the conceptual experiment of Keynes's analysis was exactly that which Clower and I have attributed to him."
>
> (1986: 415)

Here, Leijonhufvud is referring to the discovery of some unpublished manuscripts of Keynes' writings found in the famous 'Tilton laundry basket' when Lady Keynes was vacating the Tilton estate during the winter of 1975-76. In these draft pages, Keynes contrasts a cooperative economy (where labour is exchanged for goods, in which case a deficiency of effective demand is impossible) and an entrepreneur economy (where goods are exchanged for money, which would permit effective demand deficiencies and involuntary unemployment) (Yeager, 1988). While Leijonhufvud purports this to be clear evidence, Yeager maintains that if Keynes had a theory of incoordination 'at the back of his head he had it only in a fuzzy, rudimentary fashion' (1988:207) and attaches little importance to these pages since 'the fact that he wrote certain ideas down, considered them, and then *discarded* them would suggest that they were *not* what he meant.' (1988:207) Thus, it is not clear, nor will it ever be, whether the reinterpretation put forward by Clower and Leijonhufvud aligns with what Keynes really meant or pointed towards.

Conclusion

The work of Clower and Leijonhufvud, in attempting to expose Keynes as a true theoretician, was a response to the state of macroeconomics in the 1960s, which saw a rift appearing between micro and macro, with the latter lacking clear micro foundations upon which to base its conjectures. Clower first proposed the 'dual decision' process, which dissolved the simplistic market clearing ideology of the Walrasian auctioneer, and the process of tâtonnement. Leijonhufvud, in following this line of thought, emphasised quantity adjustment to demand driven disturbances and perceived the *General Theory* as a dynamic theory of disequilibrium. Leijonhufvud attributes involuntary unemployment to the information deficiency transactors encounter in the face of a demand disturbance, triggering income constrained processes and a multiplier effect.

While textual evidence for this interpretation is ambivalent, it inspired a new generation of economists with no stake in the orthodox IS-LM approach to follow a disequilibrium route to macroeconomics. Researchers such as the widely cited Barro and Grossman (1971) and Malinvaud (1977) reconsidered the theory of unemployment in terms of rationing models within a framework of the macroeconomics of non-clearing markets. This path of analysis was then intercepted by the rational expectations revolution and the emergence of New Classical economics perpetrated by Robert E. Lucas. In light of the recent turmoil in the global economy, it is rather unfortunate that disequilibrium theorising lost its fervour, as perhaps this may have pre-

vented the rigorous mathematization of economics and the ascendance of models proclaiming rationality into policy circles. As capitalism implodes once again, perhaps it is time to rejuvenate Keynesian ideas and shake off, once again, the classical doctrines that continue to pervade our thoughts.

References

Barro, R.J. and H.I. Grossman, 1971. A General Disequilibrium Model of Income and Employment. *American Ecnomic Review.* 61: 1 pp.82-93.

Clower, R.W., 1965. The Keynesian Counter Revolution: a theoretical appraisal. in R.W. Clower (ed.) *Monetary Theory.* Middlesex: Penguin.

Coddington, A., 1983. *Keynesian Economics: The Search for First Principles* .London: Allen and Unwin.

Hansen, A.H., 1953. *A Guide to Keynes.* New York: McGraw-Hill.

Hicks, J.R., 1937. Mr. Keynes and the "Classics": A Suggested Interpretation. *Econometrica* 5: 2 pp. 147-159.

Keynes, J.M, 1951 [1936]. *The General Theory of Employment, Interest and Money.* London: Macmillan.

Klein, L.R., 1947. *The Keynesian Revolution.* New York: Macmillan.

Leijonhufvud, A., 1967. Keynes and the Keynesians: a suggested interpretation. *American Economic Review,* 57: 2, pp.401-410.

Leijonhufvud, A., 1968. *On Keynesian Economics and the Economics of Keynes: A Study in Monetary Theory.* London: Oxford University Press.

Leijonhufvud, A., 1976. Schools, "Revolutions," and Research Programmes in Economic Theory in Snowdon, B., Vane, H.R. and P. Wynarczyk (1994). *A Modern Guide to Macroeconomics: An Introduction to Competing Schools of Thought.* Hampshire: Edward Elgar.

Leijonhufvud, A., 1986. Real and Monetary Factors in Business Fluctuations. *Cato Journal.* 6:2 pp. 402-407.

Malinvaud, E., 1977. *The Theory of Unemployment Reconsidered.* Oxford: Basil Blackwell.

Modigliani, F., 1944. Liquidity Preference and the Theory of Interest and

Money. *Econometrica*, 12: 1 pp. 45-88.

Pierce, D.G. and Tysome P.J., 1985. *Monetary Economics: Theories, Evidence and Policy.* London: Butterworths.

Rabin, A.A., 2004. *Monetary Theory.* Cheltenham: Edward Elgar.

Robinson, J. The production function and the theory of capital. *Review of Economic Studies* 21: 2 pp. 81-106 in A. Coddington (1983) *Keynesian Economics: The Search for First Principles.* London: Allen and Unwin.

Samuelson, P.A., 1948. *Economics.* New York: McGraw-Hill.

Shackle, G.L.S, 1967. *The Years of High Theory.* Cambridge: Cambridge University Press in A. Coddington (1983) *Keynesian Economics: The Search for First Principles.* London: Allen and Unwin.

Skidelsky, R., 1992. *John Maynard Keynes, Vol. 2: The Economist as Saviour 1920-1937.* London: Macmillan in B. Snowdon and H.R. Vane (2005) *Modern Macroeconomics – Its Origins, Development and Current State.* Cheltenham: Edward Elgar.

Snowdon, B. and H.R. Vane, 2005. *Modern Macroeconomics – Its Origins, Development and Current State.* Cheltenham: Edward Elgar.

Snowdon, B., Vane, H.R. and P. Wynarczyk, 1994. *A Modern Guide to Macroeconomics: An Introduction to Competing Schools of Thought.* Hampshire: Edward Elgar.

The Economist, 2009. 'The State of Economics' 16-07-2009 Accessible at: http://www.economist.com/node/14030288.

Yeager, L.B., 1986. The Significance of Monetary Disequilibrium. *Cato Journal* 6:2 pp. 369-408.

Yeager, L.B., 1988. On Interpreting Keynes: Reply to Leijonhufvud. *Cato Journal* 8:1 pp. 205-208.

Capitalism as religion

Arthur Thoby
Erasmus Student

In this unapologetically provocative essay, Arthur Thoby explores the fundamental nature of capitalism. He skilfully argues that Capitalism meets the relevant criteria for classification as a religion, with economics providing its theological foundations and money occupying the role of divinities. He then discusses the harmful consequences of 'Religious Capitalism' for society and its limiting influence on individual thinking.

"The charm of history and its enigmatic lesson consist in the fact that, from age to age, nothing changes and yet everything is completely different. "

Huxley, Aldous.

Introduction

Voluntarily enigmatic, this assertion of Huxley could, in certain respects, constitute the basis for the definition of postivism, especially regarding the sense of history. This paper, consistent with a certain school of thought that regards history as highly symbolic, will posit that modern Western societies do not fiercely differ from pre-Enlightenment societies in the sense that one still witnesses the realisation of a theology. As provocative as this idea may seem, thinkers from various fields of studies including anthropology, sociology, philosophy and even economists such as Fogel, laureate of the Nobel Prize in Economics in 1993, acknowledge that capitalism and economics convey religious or theological dimensions. In this context, it is argued that the rise of capitalism and Western ideologies had the progressive decline of the Christian Church as a corollary. Accordingly, this papers postulates that capitalism (and economics) is the new recipient of a fervent religious expression. Doing so, this article (i) briefly reviews the notions of capitalism and progress, (ii) establishes the grounding for considering capitalism as religion and finally (iii) analyses the implications of the second point.

Capitalism and Progress: What is Capitalism?

Rarely has a concept been more controversial than capitalism. Of particu-

lar fashion in the 19th and the very beginning of the 20th century, its use progressively declined during the 20th century to finally vigorously reappear in the writings of Hayek and Friedman. Yet, the very concept of capitalism makes economists uncomfortable. Such uneasiness is to be sought in the history of thought. Indeed, during the 20th century, capitalism was traditionally opposed to other ideologies such as communism or totalitarianism. This opposition gave capitalism the dimension of a doxa that does not fit with the scientific representation the field of economics has of itself. For that particular reason, most economists now prefer to capitalism more fashionable words such as 'liberalism', 'market capitalism', 'rational capitalism' and so on. Thus, defining capitalism is particularly tricky and depends on the standpoint one is willing to adopt. This paper being, by nature, inter-disciplinary, one will settle for a broad definition that entails the two following definitions.

Xing and Hersh (2004:100) define 'market capitalism' as a 'process of societal development involving historical, cultural and religious causes. Their approach is all the more original that they regard capitalism as a political realisation rather than 'the logic of history' or a 'natural outcome'.

In his last lectures, Weber outlined a powerful theory of rationalised capitalism (Collins, 1980). He depicts it as a system aiming to provide human needs thanks to 'the entrepreneurial organisation of capital, rationalized technology, free labour, and unrestrained markets' (Collins, 1980:925). Here, the rise of the bureaucratic is central because the latter is responsible for the implementation and enforcement of a calculable system of laws along with the notion of citizenship.

Intellectual Revolution and the Notion of Progress

As many historians of thought, economists, or sociologists underline, understanding capitalism requires a broad inquiry into history and the history of thought. That said, there is a large consensus placing capitalism at the heart of progressism or modernism.

As pointed out by Lash (2004), this revolution corresponded to the shift from an accidentally chaotic conception of the atom to a state in which nature became a place of exchanges. In light of the 'Enlightenment', the chaos of collisions and exchanges began to form patterns and laws. This is properly illustrated by the Kantian critique of metaphysics, whereby metaphysics is limited to the condition of understanding. In short, the 19th century witnessed a new scientific and metaphysical order celebrating the notion of reason or, as Xing and Hirsh (2004) put it, a shift from the 'Age of Faith' to the 'Age of Reason'.

Simultaneously, great changes occurred in the political realm. As

Weber notes this period was characterised by the struggle of bourgeoisie to takeover political institutions and subsequently expand its economic activities. However, consistent with the Enlightenment principles, the new institutions were designed to convey the ideas of democracy, citizenship and freedom, thus encouraging people's participation in the system.

In sum, post-Enlightenment societies were characterised by an unfettered faith in science and reason, freeing humanity from belief and celebrating democracy. For the first time in its history, humanity was thought to be capable of taking on its own destiny, giving birth to the notion of progress. Thus, progress is a forward-looking concept whose pillars, science and democracy, ultimately aim to ensure humanity's well-being through freedom and the control of nature.

Capitalism as Religion: Defining Religion

Traditionally, religion or theology consists of the study of the transcendent or metaphysical.

In The Elementary Forms of the Religious Life, Durkheim (1964) defines religion as 'a more or less complex system of myths, dogmas, rites and ceremonies or, put more simply, a system of beliefs and rites. In this sense, beliefs consist of opinions and representation while rites determine the mode of actions. The French sociologist adds that religious phenomena are generally associated with the observation of (i) supernatural, (ii) divinity and (iii) the belief in spiritual beings.

Bourdieu defines religious power as 'the authority to modify, in a deep and lasting fashion, the practice and world-view of lay people through the absolutization of the relative and legitimation of the arbitrary' (Verter, 2003:153). In other words, he posits that religion structures the perception and thinking of the world and especially the social world. Religion does so by the imposition of a system of practices and representations (social capital) whose structure presents itself as the natural-supernatural structure of the cosmos. In this framework, religious need is an inner compulsion seeking understanding of the world or, to put it in a Heideggerian fashion, the expression of the need for causation.

Capitalism: a Religion?

In a short piece, Walter Benjamin declares that capitalism should be examined as a 'purely cultic religion', i.e. which does not appeal to a dogma or theology (Löwy, 2009). Social actions (rites and ceremonies), he argues, are the only forms of the cult, implying that the latter is solely practical, hence non-transcendental. Finally, capitalism's main origi-

nality, asserts the German philosopher, lies in the fact that it is not expiatory but guilt-producing. Though fascinating the approach is, the latter suffers nevertheless from two important limitations. In short, it fails to recognise the theological and expiatory nature of capitalism.

Fritz Mauthner, further develops the conceptualisation of money already present in Walter's thesis. Basing its assumption on an etymological ground, he posits that God, as Idols, is essentially a human creation that gains life, becomes a significant phenomenon, and then ultimately subdues humanity (Löwy, 2009). Viewed from this optic, banknotes would constitute the new divinities of the cultic phenomenon named capitalism. On this particular subject, Lash proposes an insightful interpretation. The latter postulates that capitalism is the encounter of metaphysics and physics, where metaphysics is in-itself and physics for-itself (Lash, 2004). In this sense, he stresses that money, being abstract and tangible, is both metaphysical and physical. Thus, as exchange values are abstract concepts, money solely constitutes the physical face of capitalism.

Nelson (2001, 2004), argues that capitalism is greatly transcendent. As a matter of fact, it offers to achieve the Kingdom of Heaven (myth) in this world rather than in the hereafter (Nelson, 2004). In substance, economic progress (through the application of progressive methods) is expected to alleviate poverty while providing humanity with the goods and services necessary to its worldly well-being, its salvation. Thus, Fogel notes that the progressive era was a period of intense religious expression (Nelson, 2004). Ironically, historians usually describe this movement as the 'Gospel of efficiency'. Moreover, the abusive association of capitalism with democracy, whose pregnancy appears clearly in the writings of Hayek and Friedman, symbolically reinforces the idea of salvation, capitalism helping humanity in its quest for freedom.

Market mechanisms in themselves carry theological dimensions. Thus, as Cox (1999) points out, free markets and God are both ascribed the same characteristics, namely omnipotence, omniscience and omnipresence. In a more detailed fashion, market laws are elevated to the rank of natural laws. This is the very idea of the Invisible Hand, whose universal and absolute nature justifies and determines social interactions (Xing and Hersh, 2004).

In the light of Bourdieu's definition, economics undoubtedly constitutes a religion. First, economics is a doxa that presents itself as an organised scientific system of explanations aiming to describe the real world (Xing and Hersh, 2004; Nelson 2001, 2004; Oslington, 2000). Here, economists are in charge of the production and reproduction of the theology (Bourdieu, 1998). In effect, scientific aspects seem to have partly taken over the field of economics. Yet, economics failed as a science (notably with reagrd to prediction)

while it thrived as a theology (Nelson, 2001). In addition, beyond the purely scientific aspect, economic laws and their corollary capitalism are regarded as rational, inevitable and natural (Fukuyama, 1992). Second, economics as a theology produces the basis for moral judgment and values. As a matter of fact, it produces new values corresponding to the old Christian good and evil (Nelson, 2001, 2004). In this sense, actions encouraging the well-functioning of free markets are seen as being good (for the public interest) whereas limiting measures are ascribed an evil nature. From this standpoint, economics enables people to resolve an ethical dilemma regarding the sole pursuit of self-interest. In addition, this moral basis constitutes the point of departure of the notion of expiation. Consistent with the idea of salvation, economics requires short term sacrifices (transition costs) that should be all the more significant if heaven is reachable in this world. Third, economic rhetoric is partly metaphysical (Nelson, 2004, 2001; Oslington, 2000; McCloskey, 1983). Thus, McCloskey (1983) notes that economists do not follow their methodologies (modernism), even arguing that if they had done so, great theories such as the one of Keynes would have never seen the light of the day. McCloskey adds that econometrics is too narrow and illustrates, taking the example of purchasing power parity, that at the end of the day, the interpretation of mathematical results (not to mention the construction of models) is a matter of faith. This, he argues, is perfectly exemplified by the existence of the semantic of evidence. Thus, economists' abuse of sentences and words like: 'significant', 'in some absolute sense', 'evident', 'obvious', 'it is natural to assume' and so on. Moreover, it is ironic to notice that economics' animosity towards metaphysics is itself metaphysical (Oslington, 2000; McCloskey, 1983). More importantly, one should note that in the absence of such a metaphysical rhetoric, several key hypotheses of economic theory could not be sustained.

In sum, one may consider capitalism as a religion in the sense that it is a system of beliefs and rites in which economics plays a theological role. It is a hegemonic system that combines norms, values, divinities and laws. The content of its faith is directly derived from the progressive movement that celebrates the salvation of humanity through science and democracy.

Capitalism as Religion: Implications

Probably the most important implication lies in what Bourdieu calls the habitus. Religion, as a symbolic system, incorporates the individual level in the form of the habitus. Here, the habitus is a structured structuring structure which could be assimilated to a lasting disposition of the individual to act in conformity with a systematic view of the world (Verter,

2003). In this sense, the habitus is responsible for the construction of beliefs and representations at the individual level. Consequently, it is anterior to conscious thought and structures the actions of the individual.

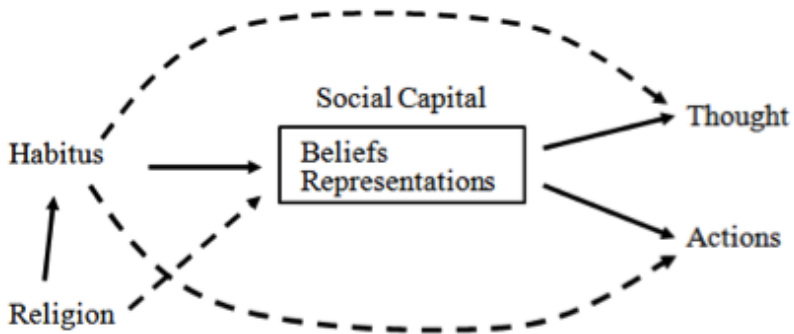Figure 1 illustrates this concept:

Fig.1 Bourdieu's conception of *Habitus*.

Why does this matter? Essentially, it means that the great principles of capitalism and free markets are no longer processed through the close scrutiny of reason, but rather, the former being the only way to envisage the understanding of the real world, they are given for granted. In the end, conscious thought may be exercised solely within the boundaries set by the theology.

With respect to the habitus, one has to underline the importance of education. It is indeed education that ensures the expansion of social capital and relays the Divine Parole. The educative process reduces the message to its essence, i.e. the key representations that need to be spread in order for the system to ensure its functioning. But in doing so, it both denatures and radicalises it. Moreover, the message is essentially a negative discourse, which constructs itself by rejecting and pointing out the flaws in other systems. A prime example of this assertion is the association of capitalism and democracy. As one remarked earlier, the very notion of progress entailed both science and democracy. But in the course of the 20th century, as communism and totalitarianism rose, capitalism was more and more associated with democracy, becoming, in the end, a condition to the establishment and development of democracy. Note that the process of education may be hazardous for the system in itself as it promotes representations that are not consistent with its

philosophy or may lead to actions inconsistent with the latter.

A Blind Faith

In many respects, one may say that the last two decades have witnessed an unprecedented period of blind faith in the latest version of capitalism, namely neoliberalism. The fall of the Soviet Union marked the triumph of capitalism and its corollary, democracy (Fukuyama, 1992). What the economists call now the Great Moderation could be depicted, in Fagel's semantic, as the fifth awakening, i.e. a blind faith in the 'natural' laws of markets. Two to three decades of transmission of Samuelson's principles and the relative failure of socialist capitalism led eventually to an enthusiastic if not zealous application of neoliberalism whose ambition was to release the providential forces of the market.

First, it decided the implementation, by policy makers that only partly understood the implications of their actions, of an unbridled globalisation. Again, this is not to say that globalisation is a negative phenomenon (I am quite convinced it is positive), but it is quite clear now that this rather was the resultant of dogmatic forces (if not theological) than the product of cold reason. This resulted, in many countries, in severe non-addressed competitiveness issues.

Second, it triggered the development of financial markets in a hazardously deregulated environment. The example of derivatives perfectly illustrates the latter assertion. In the mid-2000s, when it came to regulate derivatives in the US, the American Congress that did obviously not understand the nature of the matter, gave, in a sense, carte blanche to American banks to expand their activities recklessly, accounting for the ability of the market to regulate itself.

Thus, poor understanding of both the competitiveness of issues arising from globalisation and the financial innovation is, in a way, the cause of the current crisis, whose point of departure was a blind faith in the market and capitalism as whole.

From a Society With Markets to a Market Society

Regarding capitalism as a religion allows making certain comparisons. The comparison of the history of Catholicism to that of capitalism, at least of their structures, is worth highlighting. Interestingly, looking at the expansion of the catholic faith, the tipping point is certainly its association to the Crone of France, at a time when Clovis sought efficient administrative institutions to rule his kingdom. Note that faith did not play a great role in this association, the goal of Clovis being essentially instrumental. Then, in the course of the following decades, religion expanded its influence and became the holy

authority, being the only source of empowerment of government structures. In short, one may say that the religious entity progressively reversed the balance of power, ultimately subordinating the political realm to its normative supremacy. How does that fit with the history of capitalism? Capitalism being the new religion, it is arguable that modern societies are still tripartite. Figure 2 presents the structure of the tripartite model:
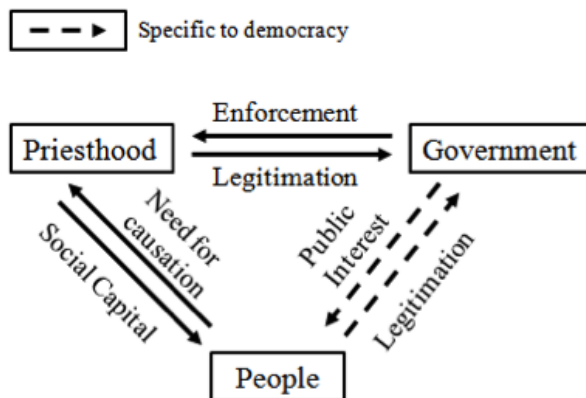
Fig. 2 Tripartite Structure.

In such a framework, in equilibrium (in the sense of sustainability), religious and political power counterbalance each other (here priesthood and government). Should the priesthood overcome government's authority, it would put the entire system at risk. In many ways, the history of capitalism resembles the one of Catholicism. Thus, in its early conception, capitalism aimed to serve humanity's well-being (instrumental). Progressively and consistently with its diffusion, the new secular religion gained momentum and counterbalanced government authority. Ultimately, with the relative failure of socialist capitalism (interventionism), religious authority subdued governments, the condition to their legitimacy being related to their accomplishment regarding the hegemonic theology. In sum, capitalism became an end in itself (Friedman, 1982). Symbolically, one may view the current crisis as the natural outcome of a system in which: (i) the private sphere took advantage of the public sphere; (ii) the priesthood's doxa became an end in itself, subordinating governments to its will.

Conclusion

Voluntarily provocative, the ambition of this paper was to highlight the partly theological nature of capitalism and show that it can be considered as an essentially religious phenomenon.

Firstly, capitalism was described as a process of social development finding its essence in the notion of progress seeking humanity's well-being through the advances of the modern sciences and the establishment of democracy. Following Weber's definition, this system is characterised by the release of the market forces in which the pursuing of self-interest (entrepreneurial spirit) is the key element.

Then, considering religion as (in short) an organised system of rites and beliefs seeking to provide an absolute understanding of the real world while determining social actions, one concludes that, in many respects, capitalism contains religious and theological dimensions. As a matter of fact, one may assume that social actions constitute rites while money, due to its physical metaphysical nature, plays the role of divinities. Besides, capitalism is highly transcendent. Indeed, it proposes no less that the salvation of humanity via the alleviation of worldly poverty and the implementation of democratic principles. In addition, market laws are ascribed the traditional characteristic of God. Furthermore, economics, though it presents itself as a modern science, presents metaphysical and theological elements in its practice (rhetoric, models, and hypotheses). Accordingly, one may envisage comparing capitalism, at least symbolically, to a religion.

Finally, this paper posits that religious capitalism bears significant implications. From a sociological standpoint, it implies that capitalism becomes directly embodied at the individual level in the form of the habitus, the latter determining conscious thought and actions. Doing so, it sets the boundaries in which thought may be exercised. This, added to the distorting nature of education that diffuses the great religious principles, is the point of departure of developments that have characterised the two last decades. These developments entail (i) a blind faith in capitalism and market laws, (ii) the progressive transmission of power from the political to the religious (economic) authority.

References

Bourdieu, P., 1998. L'Essence du Néolibéralisme. *Le Monde Diplomatique*, March 3rd.

Collins, R., 1980. Weber's Last Theory of Capitalism: A Systematization. *American Sociological Review* 45:6:925-942.

Cox, H., 1999. The Market as God. *The Atlantic Monthly* 283:3:18-23.

Durkheim, E., 1964. *The Elementary Forms of the Religious Life.* Translated from the French by Swain, J.W. . London: George Allen & Unwin LTD.

Friedman, M., 1982. *Capitalism and Freedom*. Chicago: The University of Chicago Press.

Fukuyama, F., 1992. *The End of History and the Last Man*. New York.

Lash, S., 2004. Capitalism and Metaphysics. Paper Presented at the Meeting "Capitalism and Metaphysics", organised by the Institute of Sociology of the Faculty of Humanities of the Porto University, 2004, November 3.

Löwy, M., 2009. Capitalism as Religion : Walter Benjamin and Max Weber. *Historical Materialism* 17 :60-73.

McCloskey, D.N., 1983. The Rethoric of Economics. *Journal of Economic Literature* 21:2:481-517.

Nelson, R.H., 2001. *Economics as Religion: From Samuelson to Chicago and Beyond.* University Park: The Pennsylvania State University.

Nelson, R.H., 2004. What is "Economic Theology"? *The Princeton Seminary Bulletin* 25:1:58-79.

Oslington, P., 2000. A Theological Economics. *International Journal of Social Economics* 27:1:34-44.

Verter, B., 2003. Spiritual Capital: Theorizing Religion with Bourdieu against Bourdieu. *Sociological Theory* 21:2:150-174.

Xing, L. and Hersh, J., 2004. The Genesis of Capitalism : The Nexus Between "Politics in Command" and Social Engineering. *American Review of Political Economy* 2:2:99-143.

The Solutions to Externalities: From Pigou to Coase

Andrew Winterbotham
Junior Sophister

In this essay, Andrew Winterbotham provides a well researched account of the development of the theory of externalities, through an assessment of the work of Arthur Pigou and Ronald Coase. The global nature of externalities in the twenty first century is also highlighted, and the author concludes that the Coasean approach is a better solution to the vexing problem of global externalities.

Introduction

"...the essence of the matter is that one person A, in the course of rendering some service, for which payment is made, to a second person B, incidentally also renders services or disservices to other persons C, D and E"

Arthur C. Pigou (1924: 161)

The concept of externalities is as relevant to the field of economics as when Arthur Pigou first postulated the idea in 'The Economics of Welfare' (1924).[1] Externalities have hugely significant effects on societal welfare. Take for example the positive externalities deriving from education, which not only lead to increased future output but have also been shown to reduce crime (Hillman, 2009). Moreover, it is essential that the negative externalities arising as a result of economic growth, the most fundamental way of improving living standards, are minimised. Paul Romer likens economic activity to cooking in The Concise Encyclopedia of Economics (2007):

"...most cooking in the economy produces undesirable side effects. If economic growth could be achieved only by doing more and more of the same kind of cooking, we would eventually run out of raw materials and suffer from unacceptable levels of pollution and nuisance".

---

1 Externalaties essentially arise as a result of one agents actions affecting the welfare of another. Put simply, every action taken by an individual or firm has consequences for others, either positive or negative, consequential or inconsequential.

The goal of this essay is to discuss and evaluate the solutions to externalities put forward by the two schools of thought; that of the public policy means of taxes and subsidies advocated by Pigou, and the private resolution approach, as advocated by Coase (hereafter referred to as the Pigovian and Coasean approaches respectively). Firstly, externalities shall be defined by way of some history, and then the modern formulation is outlined. The solutions to externalities are then explored. We then highlight the now global nature of externalities, and the suitability of the Coasean approach in dealing with it. At this point it is worth noting that the Coase theorem has drawn criticism from many, but this is beyond the scope of this paper. Nevertheless, as shall be demonstrated, Coase's propositions are still of some policy relevance.

The Marshall-Pigou Proposition

The Marshall-Pigou formulation essentially views externalities as an example of market failure. Market forces, they argued, will lead to insufficient output by industries enjoying external economies and will cause industries with rising supply curves to overexpand. Hence, the Marshall-Pigou prescription is to harmonize private production decisions with government intervention. They advocated taxing the latter set of industries and subsidizing the former (Bator, 1958).

The Marshall-Pigou proposition can be unraveled by making a distinction between 'pecuniary' and technological externalities. A 'pecuniary' externality is said to exist where the actions of one firm inadvertently bid up the factor prices or lowers the price of the product of another (Goldin, 1975; Merewitz and Sosnick, 1971). Technological externalities exist when actions taken by one firm affect the production function of another (Holcombe and Sobel, 2001; Scivotsky, 1954).

Pigou's thoughts on technological externalities are correct, while his proposition on pecuniary externalities was fundamentally wrong. If an industry enjoys external economies, a subsidy should not be required. The implied gains in efficiency are adequately signaled by the input price, and profit-maximizing output levels by the A-firms are socially efficient (Bator, 1958). So unlike technological externalities, pecuniary externalities do not result in inefficiency, and thus do not warrant corrective action. In fact, it is the ability of firms to inflict pecuniary losses that generates efficiency in competitive markets. Efficiency merely requires that people have clearly defined property rights over the ownership of property but not over the market value of that property (Holcombe and Sobel, 2001). We may even rule out pecuniary externalities, as they ultimately arise from technological external economies from

somewhere within the system (Bator, 1958).

The Modern Formulation

The modern notion of externalities involves "direct interaction". This may take the form of producer-producer, consumer-producer or employee-employee interaction. Whatever form it takes, it consists of interdependencies which are external to the price system, hence unaccounted for by market valuations. Analytically, this implies the nonindependence of various preference and production functions and its effect is to cause digression between private and social costs (Bator, 1958). For example, if industrial activity degrades the environment, the true cost is not only the cost to the producer, but also the cost to the environment. It is precisely this divergence which leads to an inefficient market outcome.

We may prove this using a simplified variant of the famous production model of bees and apples suggested by Meade (1952), as is contained in Bator (1958). Let us assume a world of perfect competition where a single purchasable and inelastically supplied input, labour (L), is used to produce two homogenous (of degree zero) and divisible goods, honey (H) and apples (A), at nonincreasing returns to scale. The output of A is dependent only on L_A: $A = A(L_A)$. Honey production also depends on the level of apple output: $H = H(L_H, A(L_A))$. When we solve the usual constrained maximisation problem for the production possibilities curve, it can be shown that Paretian production efficiency implies:

$$(1) \qquad p_H \frac{\partial H}{\partial L_H} = w$$

$$(2) \qquad p_A \frac{dA}{dL_A} + p_H \frac{\partial H}{\partial A} \frac{dA}{dL_A} = w$$

where p_H, p_A and w represent the prices, respectively, of honey, apples and labour[2]

Equation (1) is consistent with profit maximizing and efficiency. Each competitive honey producer will hire labour until the value of its social as well as marginal product equals the wage rate. The apple producer will be inefficient, unless , that is the cross effect of apples on honey, is zero. Concretely, if apples have a positive external effect on honey output, the market-determined level

---

2 The maximization of $p_A A + p_H H$ subject to the production functions and the supply of labour, is equivalent to finding the critical value for a Lagrangian expression.

of apple output, , will be less than is socially desirable.[3]

Pigou touched on this topic a quarter of a century earlier, arguing that this divergence was sometimes necessary, for the maximization of societal welfare. He used the example of the postponement of the construction of an electric utility to protect a gas utility. If the technological improvement had been allowed proceed, the gain in consumer welfare would offset the loss in the capital value of the gas utility.

Solutions
1. The Coasean Approach

Coase's powerful article 'The Problem of Social Cost' (1960) revolutionized the way the economics profession viewed externalities. In this groundbreaking article, he argued that the Pigovian approach would "lead to results which are not necessarily, or even usually, desirable" (Coase, 1960: 2). The alternative that Coase proposed is perhaps best described using the real life example involving a doctor and a confectioner[4]. The doctor's business was impeded by machinery used by the confectioner and the court forced the confectioner to compensate him.

Coase argued that the same optimum result could have been achieved through private bargaining between the two parties. The ruling dictated that the doctor had the legal right to prevent the confectioner from using his machinery. Intuitively, the doctor would have been willing to receive a payment from the confectioner for him to continue using his machinery, so long as this was greater than the cost to the doctor of moving premises, or building a wall etc. This relates to the fact that Pigou failed to take into account the concept of opportunity cost. If the shoe was on the other foot i.e. the confectioner had the legal right to use his machinery, the outcome would be analogous. The doctor would simply have to pay the confectioner to stop using his machinery. If the doctor's income would have fallen more through continued use of this machinery than it added to the income of the confectioner, then there would clearly be room for bargaining.

Such rearrangement of legal rights through the market would take place whenever this led to an increase in the value of production. When we take transaction costs into account, this will only take place if the increase in production is greater than the cost involved in bringing it about (Coase, 1960). More precisely (despite Coase never explicitly mentioning it), the Coase theorem states the following:

3 For a more detailed analysis of the private and social costs of these externalities, see Bator (1958)

4 Sturges v. Bridgman 11 Ch. D. 852 (1879)

1. Externality problems are efficiently resolved by assignment of legal rights.
2. The efficient resolution of an externality is independent of who has legal rights.

<div align="right">(Hillman, 2009: 325)</div>

Coase argued, in some respects anticipating the current 'NIMBY' culture, that the Pigovian approach is inherently wrong in assuming that building construction is necessarily "anti-social". It may or may not be. It is necessary to weigh the harm against the good that will result. He claims "nothing could be more "anti-social" than to oppose any action which causes any harm to anyone" (Coase, 1960, 35). Put simply, this would erect barriers that would make any sort of construction work nigh on impossible. The reciprocal nature of externalities had also never been acknowledged beforehand. If the confectioner mentioned above had been using his machinery for years before the doctor set up his practice, then which party deserves compensation becomes more ambiguous.

Coase also alluded to the idea of government failure. He purports that externalities that a government tries to solve are actually caused by governmental action itself (Coase, 1960). Wolf (1987) expands on this argument, labeling them 'derived externalities'. On the other side of the debate, Grand (1991) argues that government organizations should not necessarily be singled out as being subject to the 'derived externality' problem. Any activity, whether undertaken in the public or private sector, has the potential to cause unintended side effects. The effects of certain actions cannot always be adequately predicted, a fact of life to which private organizations are as subject as public.

2. The Need for Government and the Means of Public Policy

The failure of the Coasean approach in certain circumstances (e.g. if a market for Coasean trading simply does not exist or is illegal) is the primary reason for the survival of the Pigovian approach. Civil societies, in a way, efficiently allocate the resources of government, by limiting the sources of externalities that can be declared within the domain of public policy. For example, personal freedoms can be taken away only in cases threatening public health or security i.e. for the containment of a contagious disease. People speaking with a certain accent say do not meaningfully impinge on the welfare of others and so corrective action here is neither desirable nor necessary.

The motivation for Pigovian taxes stems from the belief that a corrective

tax on producers forces them to internalize the external cost. This therefore causes them to produce at the point which cost-benefit analysis has deemed to be the level of output that corresponds to an efficient outcome for society i.e. where societal welfare is maximized. Changing the behavior of the firm through a Pigovian tax however may give rise to the substitution effect, creating a deadweight loss; the sum of the lost producer and consumer surplus. In its defense however, it has been argued that technological innovation may occur as a result of a Pigovian tax on producers, encouraging or even forcing them to innovate (Hillman, 2009). This will arguably even generate positive externalities. For example, it may lead to the discovery of breakthrough technologies that facilitate greater productivity, along with reduced pollution. Furthermore, recent research has proven that that a federal tax on alcohol in the US reduced deaths by 4.7%, or almost 7,000, in 1991 (Cook and Durrance, 2011).

3. Other Means

Approaches other than the aforementioned are of limited policy relevance and by no means offer a 'catch-all' solution. The confectioner described above, for example, may decide to be incredibly considerate and moderate his use of the machinery. Conversely, the doctor may be considerate and allow the confectioner free rein over the use of his machinery. If either occurred, the externality would then be non-existent. This would impose a cost to the considerate party though, and so would be difficult to envisage in reality. In addition, people may behave in a certain way, creating positive externalities, either to gain personal satisfaction or social approval. Either way, utility is gained by the individual who creates the positive externality. Finally, social norms play a role in deterring certain actions that create negative externalities. By and large, people conform to these social norms, causing some to behave differently than they would otherwise. This may go some way towards explaining the problem of crime in areas where violence and drug and alcohol abuse is widespread; disadvantaged youthes are particularly prone to replicating this sort of behaviour.

Global Externalities

Globalisation has dampened the effect of national and regional policy actions in counteracting externalities. International consensus has been incredibly difficult to achieve however. With the absence of a so called 'world government' it is likely that little progress will be made (O'Hagan and Newman, 2008). Yet achieving this has never been so imperative. An example of this

difficulty in gaining compliance to international agreements is the Montreal Protocol of 1987, which attempted to phase out emissions of ozone depleting substances. Compliance was uneven on account of the difficulties faced by poorer nations (Hillman, 2009).

This problem may be illustrated using a simple prisoner's dilemma payoff matrix, assuming democratic governments.[5] It can be shown that an inefficient Nash equilibrium will be reached, whereby neither country in the game co-operates. The prisoner's dilemma of international externalities is thus a case of the tragedy of the commons, where the "commons" refers to the global environment where each government only internalizes the externalities on its own population. The case of a common field being depleted by overgrazing is an analogous example. This can be relatively easily solved; privatization forces the farmer to internalize the costs of overgrazing.

The inefficient Nash equilibrium in this case however can only be escaped by the Coasean trading of emissions rights. Efficiency is achieved as the Coasean approach allows firms which can profitably take advantage of emissions rights to purchase rights from producers anywhere in the world who have lower demands for the rights. The Pigovian approach of taxes and subsidies is not feasible in this case because the global demand function expressing willingness to pay for emissions rights cannot be known with certainty (Hillman, 2009).

Conclusion

This essay has sought to examine and compare the various solutions to negative externalities, namely those put forward by the two schools of thought; the Pigovian and Coasean approaches. Firstly, we evaluated Pigou's propositions, and concluded that his failure to distinguish between 'pecuniary' and technological externalities somewhat limited his belief that all externalities require government intervention. We then formally proved the modern definition of externalities, i.e. that they essentially represent a divergence between private and social costs. We then examined the solutions to externalities by firstly exploring the Coasean approach and emphasizing its prevailing relevance. We conceded that the Pigovian approach is desirable in certain circumstances, however, particularly when a market for Coasean trading is simply non-existent. Finally, the pressing issue of global externalities was highlighted, where it was concluded that the Coasean approach is the only feasible means of dealing with the problem.

---

5 Dictatorships have a notoriously poor record of complying with international agreements and so their inclusion here would not be conducive to any sort of meaningful analysis.

References

Bator, F. M., 1958. The Anatomy of Market Failure. *The Quarterly Journal of Economics* 72(3), 351–379.

Coase, R. H., 1960. The Problem of Social Cost. *Journal of Law and Economics* 3, 1–44.

Cook, P. J. Durrance, C. P., 2011. The Virtuous Tax: Lifesaving and Crime-Prevention Effects of the 1991 Federal Alcohol-Tax Increase. Working Paper 17709, National Bureau of Economic Research.

Goldin, K. D., 1975. Price Externalities Influence Public Policy. *Public Choice* 23, 1–10.

Grand, J., 1991. The Theory of Government Failure. *British Journal of Political Science* 21(4), 423–442.

Henderson, D. R., 2007. *The Concise Encyclopedia of Economics*, 2nd edn, Liberty Fund.

Hillman, A. L., 2009. *Public Finance and Public Policy*, 2nd edn, Cambridge University Press.

Holcombe, R. G., Sobel, R. S., 2001. *Public Policy Toward Pecuniary Externalities.* 7 4(4), 195–210.

Meade, J. E., 1952. External Economies and Diseconomies in a Competitive Situation. *Economic Journal* 62(245), 54–67.

Merewitz, L., Sosnick, S.H., 1971. *The Budget's New Clothes*, 2nd edn, Markham.

O'Hagan, J., Newman, C., 2008. *The Economy of Ireland*, 10th edn, Gill and Macmillan.

Pigou, A. C., 1924. *The Economics of Welfare*, 2nd edn, Macmillan.

Scivotsky, T., 1954. Two Concepts of External Economies. *Journal of Political*

Economy 62(2), 143–151.

Wolf, C., 1987. Market and Non-Market Failures: Comparison and Assessment. *Journal of Public Policy* 7(1), 43–70.