



World Economic Review - Issue 2

Regional Monetary Cooperation: Lessons from the Euro Crisis for Developing Areas? Sebastian Dullien, Barbara Fritz and Laurissa Mühlich	1
Institutional Support and Technological Upgrading: Evidence from Dynamic Clusters in Latin America and Asia Rajah Rasiah and Jebamalai Vinanchiarachi	24
Transacting Without Pricing, Pricing Without Transacting: A case for disconnecting the valuation function from the exchange function of financial markets Jacques-Olivier Charron	48
Merit Regulation via the Suitability Rules Ronald J. Colombo	54
Basel III and the Strengthening of Capital Requirement: The obstinacy in mistake or why "it" will happen again Bernard Vallageas	67
Deconstructing the Theory of Comparative Advantage Reinhard Schumacher	83
The Emergence of Profit and Interest in the Monetary Circuit Egmont Kakarot-Handtke	106
Book Review: <i>Why Nations fail - the origins of power, prosperity and poverty</i> Deniz Kellecioglu	118

World Economic Review ISSN: 2049-3517

Published by the World Economics Association

Regional Monetary Cooperation: Lessons from the Euro Crisis for Developing Areas?

Sebastian Dullien

HTW Berlin, Germany

Barbara Fritz

Freie Universität Berlin, Germany

Laurissa Mühlich

Freie Universität Berlin, Germany

Abstract

The euro crisis has highlighted the problems of the undertaking of an ambitious regional monetary integration project with only limited economic policy cooperation backing it up. We thus ask what lessons regional monetary cooperation schemes in other world regions can draw from this experience? This paper identifies three aspects as being crucial for the Euro crisis: first, the need for fiscal cooperation, including the enforcement of sovereign state insolvency; second, the need for a mechanism to extend lender of last resort facilities to solvent yet illiquid sovereign member states; and, third, the need for prudent financial regulation at the supranational level. Against this background, the paper analyzes monetary cooperation schemes in Latin America and Southeast Asia, namely the Latin American Reserve Fund and the Chiang Mai Initiative Multilateralization, together with the Asian Bond Markets Initiative. While the euro zone faces the alternatives of either deepening or breaking up, the study reveals that the cooperation schemes in Latin America and Southeast Asia – while less ambitious in scope – show surprisingly stable institutional settings, despite little economic policy coordination underpinning them. However, the euro experience shows that the need for more extensive economic policy coordination increases as financial integration becomes more profound.

Key words: euro crisis, regional monetary integration, lender-of-last-resort, intra-regional asymmetries, fiscal policy, financial regulation

1. Introduction

Does the current European sovereign debt crisis put a question mark over the entire project aimed at establishing a common currency among the sovereign states of Europe? Further and more importantly, does it demonstrate a crisis of economic integration in general, with it being inherently vulnerable to failure as long as it is not supplemented by the political unification of its members? And, therefore, should developing countries – who are actively engaged in monetary cooperation schemes in all regions of the world – refrain from participation in such initiatives, in order to avoid themselves ending up in such a crisis?

A widespread consensus with regard to the euro crisis holds that the paucity of fiscal coordination that exists among the sovereign member states constitutes an unstable institutional arrangement for the common monetary area. The survival of the euro as a common currency would, therefore, seemingly depend on decisive steps being taken towards intricate policy cooperation, including a clear shift in fiscal sovereignty from the national to the supranational level. However, the de-nationalization and centralization of not only monetary but also fiscal policies would leave no significant instruments with which sovereign member states could pursue national policies for the enhancement of the economic and social well-being of their citizens.

Such questions remind us of the neo-functionalist approach of Ernst Haas in “*The Uniting of Europe*” (1958). Based on his work, many analysts have depicted European integration as a continuous

process – one evolving towards an ‘ever closer union’. Haas sought to develop a theory that explained why, once an initial commitment had been made, the forward momentum of integration was inevitable. Endogenous ‘natural spillovers’ were seen as creating incentives to move forward from one form of regional integration to another. From a political economy perspective, he focused on organized economic interests and the pressure they exerted on governments to manage economic interdependence. Successful trade integration, for instance, would create incentives for entrepreneurs engaged in intra-regional cross-border trade to push for monetary cooperation endeavours – such as exchange rate cooperation – as a way to eliminate the disturbances in trade stemming from volatile exchange rates.

Based on the neo-functionalist approach, some scholars went on to argue that if integration dynamics ceased the European project would collapse – colloquially known as the ‘bicycle theory’ (Moravcsik 2005): as the rider slows down or stops, the bike loses its equilibrium and will eventually fall down. In the context of the current euro crisis, by this logic this would mean that unless the member countries move towards further political integration the European integration project is bound to perish. The application of the neo-functionalist approach to regional cooperation projects in emerging areas is obvious: if a lasting equilibrium short of political union cannot even exist in the euro zone, with its creation of a common currency and the European Central Bank (ECB) as the lender of last resort (LOLR) (at least for the private financial sector), then regional monetary cooperation schemes (RMCs) in emerging areas will surely find it even harder to keep the bicycle upright without constant pedalling – that is, without significantly deepening their regional ties in terms of economic policy cooperation.

In this paper, we call into question the validity of the neo-functionalist approach, on the grounds that such a theory may be too simplistic both for the European case (Moravcsik 2005, p. 250) as well as for other RMCs.¹ Rather, to a certain extent we refer to the ‘bicycle theory’ as a framework that may supply key regarding the institutional arrangement and other requirements necessary for achieving stable and long-lasting equilibrium. We define a regional cooperation and integration scheme as stable when exogenous economic shocks that arise can be absorbed by the institutional mechanisms of the region, instead of resulting in the breakup, scaling back or necessity-driven deepening of the respective form of regional integration.

In the case of the euro crisis, the German government has long argued that it is mainly a fiscal crisis and has hence suggested measures such as the fiscal compact to limit national government’s borrowing beyond what has already been legislated for in the Stability and Growth Pact.² Against this official German view, we detect a core problem in the incompleteness of the regional LOLR function, posited along the line of de Grauwe’s (2011) arguments. Despite the existence of the ECB and the absence of public or private debt in currency denominations other than the euro, we argue that the lack of LOLR facilities for solvent yet illiquid sovereign member states is deeply intertwined with problems of inefficient fiscal surveillance, on the one hand, and the lack of region-wide oversight and regulation of the financial sector on the other.

What lessons, then, can emerging markets and developing countries engaged in RMCs draw from this crisis? Are there also lessons for the handling of the euro crisis to be learnt from financial crises that have occurred in emerging markets during the last two decades? Various RMCs in Southeast Asia and Latin America have as their aim an increase in the liquidity of member countries. This is especially true for foreign exchange pooling agreements and financial market integration efforts being undertaken at the regional level.

The aim of this paper is, therefore, to detect stable institutional settings for specific forms of regional monetary cooperation and integration short of political union, both in the case of European integration as well

¹ From a comparative perspective, the idea of a standardized path to increasing levels of regional integration, with predefined sequencing such as trade integration before financial integration, is, in fact, inadequate when looking at RMCs and integration both in Europe and in other areas of the world. Hence, we will not follow the rationale of the neo-functional approach as such, but rather take from it the perspective of there being moments of institutional equilibrium and stability.

² For a detailed discussion of the German approach to the crisis, see Dullien and Guérot (2012).

as of RMCs among developing countries.³ We first discuss in detail concepts of liquidity provision and LOLR functions – in the cases of both advanced as well as developing economies – and the specific challenges of liquidity provision in the regional context (Section II). Next, we offer a synthesized analysis of the euro crisis (Section III), and analyze the stability of the institutional settings of three regional mechanisms in Asia and Latin America: the Chiang Mai Initiative Multilateralization (CMIM), the Latin American Reserve Fund (FLAR, according to its Spanish acronym) and the Asian Bond Market initiative (ABMI) (Section IV).

2. Liquidity provision and LOLR functions in a regional context

Within monetary approaches to economics, there is wide-ranging consensus about the key relevance of a LOLR for ensuring monetary and financial stability. Since the seminal work of Walter Bagehot (1873) on the emergence of the Bank of England as the first central bank in the modern world, this has been defined as the provision of unlimited liquidity for the banking system at a penalty rate.

However, in the actual provision of unlimited liquidity it is important to distinguish between illiquid and insolvent entities, in order to avoid moral hazard problems and any subsequent distortionary effects on the economy. This has not only been the experience of the recent global crisis, wherein central banks and governments were often forced into providing huge bail-outs and liquidity for banks without being able to make this distinction clearly. The occurrence of enormous financial crises in emerging markets during the 1990s already had revealed the necessity to distinguish between illiquid and insolvent entities. The latter are defined as being unable to serve their obligations in the medium- and long term, even if provided with additional short-term liquidity. The necessity to distinguish between these two concepts is an important lesson from crisis experiences in emerging markets. The case of Argentina shows that falsely treating a case of sovereign insolvency as a temporary liquidity problem actually deepens the problem, by prolonging the possible time to default and by increasing the level of debt. After a series of adjustment programmes which were implemented by the International Monetary Fund (IMF) as pre-condition for international liquidity provision from 1998 on, Argentina defaulted on its external debt at the end of 2001. Having learnt from such experiences, the IMF came out with a programmatic paper that puts forward a clear definition of solvency and liquidity:

An entity is solvent if the present discounted value (PDV) of its current and future primary expenditure is no greater than the PDV of its current and future path of income, net of any initial indebtedness. [...] An entity is illiquid if, regardless of whether it satisfies the solvency condition, its liquid assets and available financing are insufficient to meet or roll over its maturing liabilities. (IMF 2002, p. 5)

The distinction between a liquidity and a solvency crisis is vague in many cases because a liquidity crisis will, if not solved immediately, usually lead to rising financing costs and thus to an increase in the present value of debt. Any change in key parameters – especially growth rates, the interest rate and, in the case of foreign exchange-denominated finance, the exchange rate – is immediately reflected in changes in the liquidity and solvency status of the debtor.

From the literature on financial crisis and currency crisis (Obstfeld 1996), we learn that self-fulfilling debt or fiscal crisis is a possible outcome when there are multiple equilibria (Cole & Kehoe 1996). The logic here is simple: for an entity with a moderate, yet not extremely high, level of debt whether it is able to service its debt or not depends on the expectations of market participants. If investors believe that a country in debt is able to service their claims, they accept lower interest rates and the debt may therefore be sustainable. If in the case of a sovereign debtor they believe that the state in question might not be able to service the debt, they demand higher interest rates and the debt becomes unbearable – leading to a default. The catch in these models is that if a third party can guarantee continued access to loans at sensible interest rates,

³ In contrast to traditional Optimal Currency Area (OCA) theory (Mundell 1961, 1963), our aim is not to identify optimality criteria for efficient regional monetary cooperation and integration. While traditional OCA theory weighs the costs and benefits of forming a currency union against the alternative of a flexible exchange rate regime, we aim instead at the identification of both the stabilizing and destabilizing elements of regional monetary cooperation and integration in the context of different forms of cooperation.

expectations will permanently stabilize in the 'good' equilibrium and a self-fulfilling fiscal crisis is thus no longer possible. The action of this third party would help in the avoidance of huge costs for the economy, in case of successful crisis prevention.

Another lesson from the experiences of crisis of emerging markets is the one of Southeast Asia: in the Asian crisis of 1997 external liquidity provision by the IMF was clearly insufficient, both in terms of timing and of the conditionality involved – which together caused the deepening of recession and economic depression in the affected economies (Stiglitz 2002). In South Korea in 1997, for example, the IMF commenced activities only one month after the crisis had diffused. In the face of drastically depreciating exchange rates – dropping by more than 80 per cent in one year, as was the case in South Korea – the IMF insisted on continuously low inflation rates that required up to a doubling of short-term interest rates – thus choking off economic growth. The conditions placed upon drawing liquidity had a strong focus on fiscal balances, with a disregard for the negative effects thereof on domestic markets and economic growth. In light of this exposure to sudden stops and strong regional financial contagion, the Association of Southeast Asian nations (ASEAN), together with China, Japan and South Korea, initiated in 2001 regional financial and monetary cooperation in the form of regional swap arrangements intended to provide short-term regional liquidity provision – this was termed the Chiang Mai Initiative (CMI) and later CMIM, which will be discussed in greater detail in due course (see Section IV).

In contrast to the shortcomings in its reaction to the crisis in Southeast Asia, the IMF intervened rather successfully in Brazil in 2002. In this third example of emerging market experience with external liquidity provision, the IMF correctly classified the situation not as a case of clear insolvency but rather as one of temporary illiquidity – with an associated risk of it quickly turning into one of insolvency, due to mistrust of the market by investors. In 2002, such mistrust led to a speculative attack that provoked a currency crisis and a subsequent devaluation of the Brazilian real by more than 50 per cent. The IMF stepped in with a standby credit of US\$ 30 billion to be made available over a period of 15 months, which eventually stopped the speculative attack and stabilized the Brazilian currency.

The provision of the LOLR function faces two different sets of challenges at both the domestic and regional levels: First, liquidity provision for the financial system; and, second, for sovereign states. With regard to the first aspect, the question is whether insolvency can be imposed on individual institutions without generating any systemic risks. This is discussed as the 'too-big-to-fail' syndrome of large financial institutions that should be adequately addressed by prudential financial regulation. With regard to regional schemes, especially those with a high level of integration, an additional problem is faced in the form of financial institutions with intensive cross-border activities – therein potentially prompting spillovers to other financial systems, while regulation is usually restricted to the national level.

It is equally challenging to maintain an adequate level of liquidity within the public sector (in order to prevent a spillover into a solvency crisis), again an issue relevant at both the domestic and regional levels. The critical importance of public sector liquidity for the stability of the financial sector has come to the forefront of awareness in the current global financial crisis, as well as in the euro zone. Yet, agreement is a long way from being reached over central banks acting as LOLR, especially for governments. It is true that the guiding principles of central banking are clearly shifting away from a narrow orientation towards price stability (through inflation targeting) and towards achieving the overall goal of financial stability (Blanchard et al. 2010; Eichengreen et al. 2011). However, if this includes clearly defined efforts to maintain the liquidity of national governments and involves it being done irrespective of market expectations then it is an issue of great debate (Eichengreen et al. 2011, p. 24). This is not only due to the fear of fuelling inflation by monetizing public expenditures. Given that there is no legal orderly procedure for enforcing bankruptcy rules for insolvent sovereign debtors, both in the regional and international contexts, it might also encourage free-rider behaviour – especially among smaller member states, who are likely both to overspend and to expect a bail-out at the regional level.

In the case of developing economies, LOLR provision is faced with additional problems – as the LOLR functionality of the domestic central bank is beset by severe limitations. These countries are unable, to varying degrees, to borrow abroad in their own currency (see Eichengreen & Hausmann 2005; Hausmann & Panizza 2010). In most developing countries, the net foreign currency debt and the lack of long-term

financial instruments lead to their exposure to balance-sheet effects that increase the risk of currency, debt and to financial crises. The associated risks are exacerbated by the fragility of the countries' financial markets – caused by their insufficient size and a lack of diversification, capitalization and liquidity in them, as compared with the more advanced financial systems of industrialized countries (Aghion et al. 2009). Financial fragility may only partially be compensated for by the accumulation of foreign exchange reserves. Weaknesses in the financial market undermine economic stability and macroeconomic development, as financial crises are associated with short-term disruptions in economic growth and the long-term loss of economic output (see, for example, Bordo et al. 2009). Typically, financial crises also cause the severe regressive redistribution of wealth and income at both the global and domestic levels, leading to an increase in poverty (Halac & Schmukler 2004). Smaller developing countries in particular lack sufficient scales to enable the development of the mature, diversified and liquid financial markets that would allow them to achieve longer lending maturity and reduce foreign currency borrowing (Borensztein et al. 2008).

The issue of liquidity provision for these countries is thus twofold: on the one hand it is a question of foreign exchange liquidity, and on the other it is one of the restricted scope of the domestic central bank in acting as a LOLR – due to the limited volume of financial contracts that are denominated in the domestic currency. Both aspects are at the heart of the regional cooperation efforts coordinated between developing and emerging market countries. Regional schemes are seen as being supplementary to global and domestic mechanisms of liquidity provision, which appear to be either insufficient or inefficient (Bird 2010). One possibility is to pool foreign exchange reserves among neighbouring countries, to be used as insurance during periods of sudden stops and capital outflows. Even this – in most cases – does not substitute for a 'full' LOLR. Undoubtedly, a LOLR in a key international currency would be an effective way to increase liquidity.⁴ Another option at the regional level for augmenting the domestic LOLR function would be to increase the geographical size and reach of the financial systems of the countries in a region, achieved by financial market cooperation and integration and done so as to enhance domestic currency-denominated financial contracts in the region (Panizza 2006). *Ceteris paribus*, this would lead to an enlargement of the domestic LOLR function, thus reducing the risks attached to currency mismatches.

In the following sections, we analyze the euro crisis as a case of inadequate institutional arrangements – one which is thus confronted with the problem of an incomplete LOLR functionality in the regional context. These inadequate institutional settings are found to be especially challenging with regard to the prevention of the insolvency of financial institutions and states, as well to the lack of enforceability of insolvency for both of the aforementioned entities. Based on this analysis, we study the respective cooperation requirements for RMCs coordinated among developing economies with the aim of increasing international or domestic liquidity.

Based on the theoretical analysis of liquidity provision in the European regional context, and considering empirical evidence from the experience of both the euro zone as well of developing economies⁵, we will focus on:

- a) the requirements for fiscal cooperation, including for the ability to enforce sovereign state insolvency;
- b) the extension of the LOLR function to solvent yet illiquid sovereign member states; and,
- c) the requirements for financial regulation at the regional level.

3. The euro crisis

While the euro crisis is often portrayed as being simply a case of sovereign debt crisis, only a small part of the crisis has actually been a solvency issue for each particular national government involved. The spread and depth of the crisis can only be explained by the shortcomings regarding regional financial market

⁴ On the problems of the global monetary 'non-system', as well as its asymmetries with respect to developing economies, see Williamson (1976) and Mateos y Lago et al. (2009).

⁵ For a systematic overview on the variety of regional monetary cooperation schemes among developing countries and emerging markets, see Fritz/Mühlich (2012).

regulation, as well as by the inadequate provision of LOLR liquidity. Thus, in essence, the combination of an increasingly deep process of financial integration with an insufficient institutional framework has been at the root of the crisis.⁶

To understand this argument, one needs to be aware of the specific institutional features of the European integration project and of the European monetary union. Three points are especially important in this regard:

- a) First, the EU treaty has a specific 'no-bail-out' clause (Art. 125 TFEU). This provision prohibits either other countries or the EU as a whole from taking over the liabilities of other member countries. Such a clause was incorporated at the behest of certain countries who feared that the EU would morph into a transfer union. Moreover, it was believed that this clause would provide incentives for national governments to run prudent fiscal policies, as markets would punish high deficits with higher interest rates. This clause was seen to be credible, as it was believed that the fallout from a default in one country of the euro area could be contained.⁷
- b) Second, the EU treaty prohibits the ECB from directly financing government budgets (Art. 123 TFEU). While the article specifically prohibits the ECB from buying debt instruments 'directly' from member governments, there has been a debate about whether this provision also prohibits any large-scale purchases of sovereign debt in the secondary market. In particular, the German government has maintained a position that such purchases would be problematic. Up until the beginning of 2012, the ECB had been careful to emphasize that its limited purchases, under the 'securities market programme', of government bonds of the euro zone periphery were done in order to 'address the malfunctioning of securities markets and restore an appropriate monetary policy transmission mechanism' (ECB 2010, p. 1) – instead of to simply bring down yields on euro zone periphery debt.
- c) Third, financial market oversight and financial sector supervision has been exercised at the national level by implementing the stipulated EU guidelines and regulations for financial services. There has been only a loose coordination of national supervisors and no EU rule formulated for the liquidation of insolvent national or cross-border financial institutions. This set-up has been the outcome of secondary EU legislation. As will be described in more detail in due course, it has led to the outcome which one could expect if the arguments brought forward in the theory of fiscal federalism are applied: as financial sector regulation and oversight in a monetary union has a public goods character, the provision of these at a lower level of the federal structure has consequently led to a significant under-provision of effective regulation and supervision.

These three elements have interacted so as to first cause the inception of the euro crisis in Greece, and also to then spread it to other countries located on the eurozone periphery, such as Ireland, Portugal and Spain. From there, it has impacted on core countries such as Italy and France.

Greece has been the simplest case – namely one of clear and simple insolvency. Over a long period of time it has run an irresponsible and unsustainable fiscal policy, even if this reality was not openly visible at first. This is in contrast to other euro zone crisis countries, some of which (like Ireland and Spain) had been running budget surpluses prior to the onset of the global economic and financial crisis of 2008–9. Interestingly, some countries like Spain are being put under a fair amount of EU pressure at the present

⁶ A number of authors have emphasized the fact that the euro crisis also has elements of a balance-of-payment crisis, and also that underlying intra-regional divergences in competitiveness have additionally played a crucial role (cf. Giavazzi & Spaventa 2010). Given that regional trade imbalances are without doubt a crucial problem in the euro crisis (Dullien 2009), it is important that other RMCs learn the lessons of the European experience. Having this in mind, we believe that one can analytically focus on the issues specified above without covering the question of intra-regional current account imbalances. Specifically, there are empirical indications that the problems of a liquidity squeeze, the lack of a LOLR for governments as well as the problems related to the insufficient coordination of financial oversight and regulation are serious problems in themselves even if there are no regional imbalances. Ireland, for example, ran a rather modest current account deficit in the late 2000s and had already corrected that deficit by 2010. Yet, the country still has to live with a IMF–EU programme. In 2012, bond spreads for Austria and the Netherlands were increasing, while there was a talk of those governments having problems to finance their deficits – even as they were running current account surpluses. While not all problems of the euro zone would disappear if the question of fiscal surveillance, liquidity provision for governments and financial oversight were tackled, it would certainly remove important elements of its vulnerability.

⁷ In fact, even in early 2010 the view that contagion from a Greek default would be limited was widely shared by German economists. See, for example, Dieter (2010) or Enderlein (2010).

World Economic Review

juncture – even though their debt-to-GDP ratio has not been particularly high as compared to other OECD countries like Japan, the UK or the USA.

Table 1: Budget balance and debt to GDP ratios in 2007 and 2011

	Budget Balance as % of GDP		Gross Government Debt as % of GDP	
	2007	2011	2007	2011
Austria	-1.0	-3.4	63.4	79.9
Belgium	-0.3	-3.5	88.0	100.3
Estonia	2.4	0.1	7.3	12.3
Finland	5.3	-2.0	41.4	61.2
France	-2.7	-5.7	73.0	98.6
Germany	0.2	-1.2	65.6	86.9
Greece	-6.8	-9.0	115.0	165.1
Ireland	0.1	-10.3	28.7	112.6
Italy	-1.6	-3.6	112.1	127.7
Japan	-2.4	-8.9	167.0	211.7
Luxembourg	3.7	-1.2	11.3	28.2
Netherlands	0.2	-4.2	51.5	72.5
Portugal	-3.2	-5.9	75.4	111.9
Slovak Republic	-1.8	-5.9	32.9	49.8
Slovenia	0.0	-5.3	30.7	53.7
Spain	1.9	-6.2	42.3	74.1
United Kingdom	-2.8	-9.4	47.2	90.0
United States	-2.9	-10.0	62.1	97.6
Euro Area	-0.7	-4.0	71.8	95.6
Source: OECD Economic Outlook, November 2011				

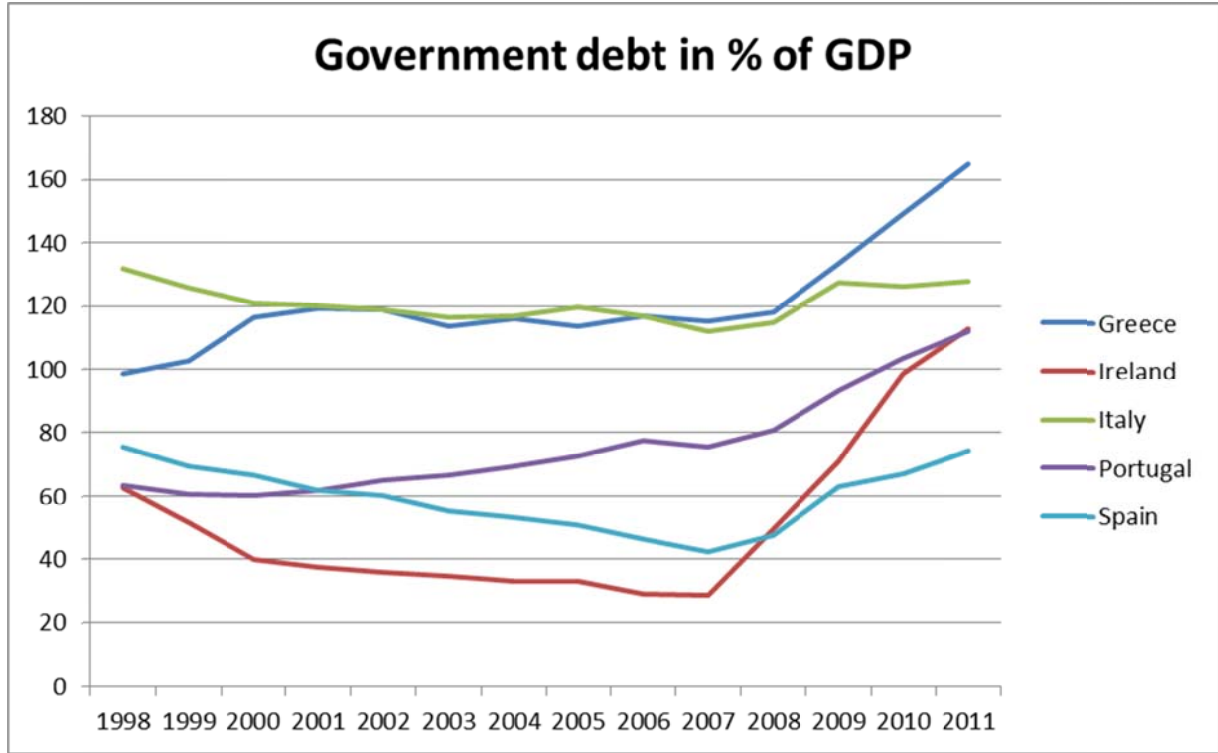
While Greece had actually seen a decrease in its debt-to-GDP ratio from 119 per cent in 2001 to 113 per cent in 2005, the underlying public finances were unsustainable. Greece had been violating the Stability and Growth Pact in any given year since its accession to the euro area (from 2001 onwards), with deficits always being clearly above 3 per cent of GDP. This problem was not detected, however, due to a number of incidents of misreporting by the Greek statistical authorities. Moreover, Greece experienced a benign debt-to-GDP trend in the 2000s as the nominal growth rate was artificially pushed up by high domestic inflation, which resulted in an artificially inflated nominal GDP (and hence mathematically in a lower debt-to-GDP ratio). As this high inflation also led to huge external imbalances these circumstances were simply unsustainable. When the global financial and economic crisis hit in 2008, Greece was in the position that the subsequent recession would push its public finances clearly and noticeably into unsustainable territory.⁸

While the no-bail-out rule was clear on paper, the euro area member states were nevertheless not willing to execute it in practice. As Greek government bonds were largely held abroad (and a significant share of them by banks of the euro area outside Greece, especially French ones), they quickly came to realize that a Greek default would have significant externalities for the other member countries even though

⁸ For an early assessment of the unsustainability of Greek debt, see Dullien and Schwarzer (2010).

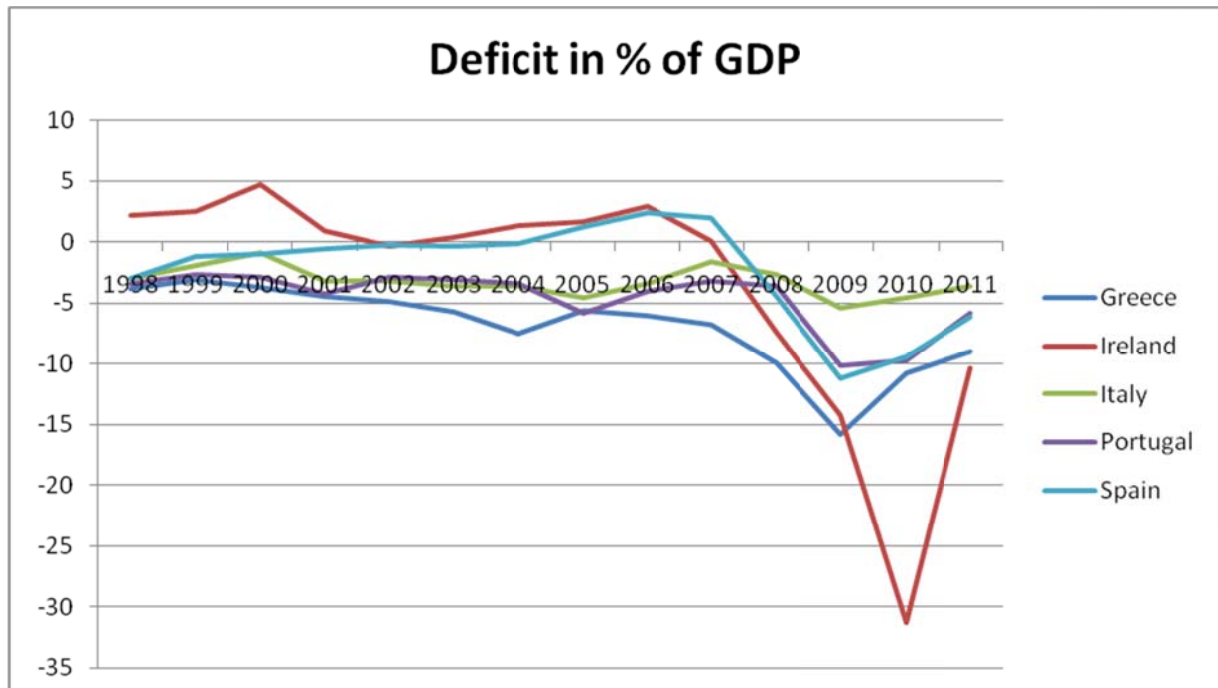
the share of the Greek economy in EMU GDP was rather tiny and trade linkages were rather limited. Hence, they decided to put together a rescue package for Greece in early 2010. The lesson from that country was that a fiscal framework that defers deficits and debts mainly to the national level in a regional integration agreement cannot be stable if the financial sectors are integrated, and if there is large cross-border holding of sovereign debt and/or large cross-border inter-bank lending.

Figure 1: Euro zone government debt as per cent of GDP



Source: OECD Economic Outlook, November 2011

Figure 2: Fiscal deficit in the euro zone as per cent of GDP 1998-2011



Source: OECD Economic Outlook, November 2011

The Irish case is more complicated and is linked to the lack of efficient regional financial regulation: Ireland had solid public finances prior to the crisis, with a surplus of 0.1 per cent of GDP in 2007 and a debt-to-GDP ratio of only 28.7 per cent. As the global economic and financial crisis hit Ireland, its domestic real estate bubble burst – alongside the same event happening in the United States – even though the country had ample fiscal space for standard counter-cyclical fiscal policies. However, the problem in Ireland was that the domestic banking system had expanded its balance sheet to several hundred per cent of GDP in the years prior to the crisis, with lending to the non-financial private sector alone having reached more than 200 per cent of GDP (Connor et al. 2010).⁹ Underpinning domestic borrowing was, to a large extent, the real estate and construction boom – which was in Ireland several times larger than it was in the US. In contrast to the US, however, securitization did not play a significant role in the Irish banking crisis. In 2008, the banks' net foreign liabilities amounted to more than 60 per cent of the Irish GDP (Honohan 2009), with a large share of them coming from other banks in the euro area (Connor et al. 2010). It is safe to say that the opportunities for Irish banks to borrow abroad to such an extent had only been made possible by the financial market integration that followed European integration and the establishment of the European monetary union.

When real estate prices in Ireland started to drop and problems in the banking sector began to emerge a bank run was inevitable. At the same time, the Irish banks' financing of the euro area inter-bank market dried up. In late 2008, when the Anglo Irish Bank was unable to roll over its foreign debt and ran out of collateral eligible for ECB refinancing, the government issued a blanket guarantee to all banks – virtually covering all liabilities. This guarantee has been estimated to have amounted to €370 billion, or 240 per cent of the Irish GDP (McGowan 2011). In the following months the government tried to resolve insolvent banks, but did so only slowly – and it turned out that the Irish banking system had significant solvency problems. The initial guarantee in the end came with genuine costs to the Irish government.

While the initial problem was a national one and the guarantee issued was aimed at calming a national bank run, it was soon accepted and actively supported by European partners who feared losses and potential bank failures in their own national banking systems should the Irish government renege on its guarantee. However, the banking guarantee ruined Ireland's fiscal position. With direct fiscal costs being estimated at more than 40 per cent of GDP (hence making the Irish banking crisis one of the most expensive to have ever hit a developed country; see McGowan 2011), the country has seen its debt-to-GDP ratio be catapulted up to more than 100 per cent of GDP. The sustainability of the Irish debt itself is now in question (Wheelan 2011).

In hindsight, it is now undisputed that Irish supervisors had not only been lax in their regulation of the offshore financial sector but also of the national banking sector (McGowan 2011), to such an extent that *The New York Times* even called Ireland the 'Wild West of European Finance' (Lavery & O'Brian 2005). Trying to attract financial business from the rest of Europe, Ireland developed a large offshore banking sector. From a rational choice perspective, such a policy makes sense for a small country as it can benefit from the additional business while at least part of the potential costs of insufficient regulation will effectively be borne by the rest of the monetary union. The Irish case can thus be seen as a sovereign debt crisis evolving out of the specific incentives offered to under-regulate a national financial system within a monetary integration agreement that has integrated financial markets.

The Irish case – with a banking system that had grown disproportionate to the country's economic size – is only the tip of the iceberg with regard to the problems of financial market regulation in the EU. As became evident during the crisis, in some countries individual financial institutions had not only grown too big to fail (since they had achieved a systemically important role in the financial system governing the euro area as a whole), but they had reached a size where it had become increasingly difficult for them to be saved by their national governments alone. One example is the financial group Dexia, which resulted from the merger in the late 1990s of French and Belgian financial institutions listed both in Brussels and Paris. With total

⁹ In addition to the domestically active banks, prior to the global crisis Ireland had developed a large offshore banking sector and as a country had foreign assets and liabilities amounting to more than 3000 per cent of GDP (Lane 2011).

assets of approximately €560 billion it can be said to be rather large. When Dexia got into trouble during the global crisis, it was not initially clear which of the two countries was responsible for the resolution of the problem and for capital injection. Finally, a deal was brokered between the governments of Belgium, France and Luxembourg to share the burden of providing financial assistance to this group.

Compared to Belgium's national GDP of roughly only €350 billion, Dexia's balance sheet is huge. Moreover, Belgium already in 2011 was heavily indebted, with a debt-to-GDP ratio of 100 per cent – so that any potential rescue packages for Dexia carried the risk of taking Belgium's public finances outside of the realm of sustainability. Compared to Dexia's huge balance sheet, the Belgian government could actually inject a relatively small amount of capital in 2009, to the tune of €6 billion. The public guarantees issued at the same time – to the tune of €150 billion, of which Belgium provided slightly more than 60 per cent – have the potential to make Belgium a second Ireland.

Hence, the lack of a common resolution framework and a fiscal authority that would be able to pay in times of debt crisis explains some aspects of the euro crisis. However, only in conjunction with the second element mentioned above – namely the lack of a LOLR for governments in the euro area – can the further spread of the euro crisis to countries such as Italy or Spain be explained. While the ECB, in principle, acts as a LOLR for illiquid yet solvent banks, and has demonstrated its willingness to do so with the large volume of long-term financing operations undertaken in late 2011 and early 2012, there is still no LOLR for sovereign debtors.

Yet governments can, similar to banks, be subject to a liquidity crisis that then leads to solvency problems, as explained in multiple equilibria models (see Part II). In countries such as Japan, the UK or the US, despite having much higher debt-to-GDP-ratios than countries in the euro area, the yields on government bonds have in general remained low. These countries all possess a congruity between their monetary authorities, their legislative bodies and their fiscal authorities. In addition, the governments in these countries are all predominantly indebted in their own currencies. In their case, it is generally expected that the central bank would lend to the national government in times of liquidity problems (possibly after a change in the central bank's legal framework). Hence, investors do not have to fear a liquidity squeeze that could turn into a solvency problem. Just by the mere existence of the option of liquidity provision, expectations are stabilized in the good equilibrium and a self-fulfilling fiscal crisis becomes unlikely or even impossible.

This is not so in the euro area: the ECB has explicitly declined to take on the role of LOLR for national governments, as a result of which a self-fulfilling fiscal crisis becomes a genuine possibility. In fact, the mechanism explained above has most likely been at play in the spread of the euro crisis to Italy and Spain. While the latter has also experienced a real estate bubble burst, its banks have remained much healthier than the Irish ones have. Moreover, Spain has had a very moderate debt-to-GDP ratio throughout the crisis years, yet spreads on its government bonds have risen strongly – fuelling doubts about the sustainability of Spain's debt.

The case of Italy has been similar: the country previously had a high level of government debt. Yet its deficit has remained moderate, even during the financial crisis. It is easy to envision an adjustment path that would quickly stabilize and reduce the debt-to-GDP ratio. Market mistrust of Italy erupted only in July 2010 in the wake of a relatively minor dispute between then Prime Minister Silvio Berlusconi and his Finance Minister. At some point, however, spreads reached alarmingly high levels, and it was only then that sustainability was called into question.

In the euro area, however, the pre-crisis integration approach has not only led to the current euro crisis, but has also caused institutional change. Relatively early on, and already during the financial crisis of 2008–9, it became apparent that the framework of European financial sector regulation and supervision was not sufficient. With the increasing integration of financial markets, the earlier regulatory framework in the EU now seemed outdated. As a consequence, the EU revamped its regulatory and supervisory structure with the intention of creating a pan-European system of supervisory agencies for financial markets, banks and insurances. Moreover, a systemic risk board for the European Union has been created. In principle, this new system is supposed to also be responsible for coordinating the oversight and resolution of cross-border financial institutions. An additional element that was decided upon at the Euro Area Summit in June 2012 – when European leaders committed to form a 'banking union', under the auspices of which the ECB will be

given the right to supervise at least systemically important institutions and the possibility of bank recapitalization from a pan-European level. While the new system is far from being perfect (and important details remain open at the time of writing), it shows at least some strong institutional dynamics in the direction of the centralization of financial market regulation and oversight.

The sovereign debt crisis has led to further institutional changes: with the new Fiscal Compact Treaty (or formally Treaty on Stability, Coordination and Governance in the Economic and Monetary Union) negotiated in 2012, which place much stricter limits on national budget deficits than the old Stability and Growth Pact did, and the so-called 'six-pack regulation' (a set of secondary pieces of EU legislation tightening the Stability and Growth Pact and providing specific rules for fiscal and macro-economic policy coordination), which aims at the much closer coordination and monitoring of fiscal policies among euro member states, the EU has reacted to the challenges posed by the virtual absence of fiscal policy coordination. Of course, this solution might in future prove to be dysfunctional as it forces the euro area to synchronize austerity – which will most likely dampen economic growth and worsen the debt problem. However, the passage into law of these new rules demonstrates the endogenous dynamics driving the EU towards further integration, and hence underlines the instability of the current institutional set-up.

Finally, Europe has started to tackle the problem of a missing LOLR. While euro member states have not altered the legal foundations of the ECB, they have created nonetheless the European Financial Stability Facility (EFSF) and the European Stability Mechanism (ESM). Both mechanisms are designed to work with cash deposits from members and guaranteed loans from financial markets, and will be able to give loans to countries with liquidity problems. Together with the IMF, these two mechanisms now provide a framework for liquidity provision. Within the typical adjustment programmes, such as those implemented in Greece, Ireland and Portugal, the IMF as well as the European institutions have lent funds, while the former has directed the technical implementation of the programme as the EU commission was seen to lack the necessary technical expertise to do so.

Of course, the EFSF and the ESM cannot fully substitute for a LOLR as their resources are limited to the cash injections and guarantees provided by member states. However, they do offer a variation on some key LOLR services and can be expected to at least prevent the emergence of self-fulfilling fiscal crises for smaller euro member states. Whether they are large enough to solve the simultaneous crises of confidence of several larger EMU member states – such as the one being experienced at the moment – remains to be seen however.

Should the EFSF and the ESM fail to provide the necessary means, an introduction of euro bonds still seems to be a feasible possibility and is one currently under discussion. In this way, euro countries could issue bonds with joint and several liabilities.¹⁰ If these bonds were adequately constructed, one could expect the interest rates on them to be below the current average for euro bonds. While self-fulfilling crises could, of course, still be possible, they would be less likely given that the overall fiscal position of the euro area is more stable than that of the individual crisis countries. Moreover, the ECB's President Mario Draghi's announcement in the summer of 2012 to intervene in government bond markets for countries with an ESM programme if necessary also vastly increases the possible impact of liquidity provision by the latter.

In sum, the experience of the euro area over the past years has clearly shown, therefore, that the current level of integration – a full monetary union with integrated financial markets, but without a centralized financial oversight structure, without stronger fiscal policy coordination or centralization and without a LOLR for sovereign states – is not a suitable proposition for long-term stability.

4. Liquidity pooling and regional financial market initiatives in Asia and Latin America

4.1 Regional liquidity pooling

Regional self-insurance mechanisms, such as swap arrangements or regional liquidity pooling, hold a strong appeal as efficient ways of self-insuring against short-term liquidity shortages (Ocampo 2006) during periods

¹⁰ For an overview of the different proposals, see EU Commission (2011). The debate is made more complicated, however, by the limits that the German constitutional court has spelt out for such euro bonds, meaning that it might be necessary to amend the German constitution in order to be able to introduce them.

of massive private capital outflows. In this context, mechanisms such as the Chiang Mai Initiative Multilateralization (CMIM) (Sachs et al. 2010) and the Fondo Latinoamericano de Reservas (FLAR) (Imbs & Mauro 2007) have increasingly gained attention, particularly during the recent global financial crisis.

A regional liquidity pool is usually coordinated through a common regional fund to which member countries subscribe and dedicate a previously agreed amount of their reserve holdings. Once agreement on the volume, maturity, fees, interest rate payments and conditionality of the financing has been reached, member countries gain access to immediate, short-term or medium-term financing, depending on the volume and structure of the fund. Alternatively, liquidity pooling can be realized by regional bilateral swap arrangements concluded between the participating central banks of a region.

Regional and global liquidity provision mechanisms supplement each other (Henning 2011). In this context, regional reserve funds may on the one hand constitute a more flexible tool for reserve provision – one that can be more easily and rapidly accessed than can international mechanisms of assistance. On the other hand, regional reserve funds are relatively small compared to those of global mechanisms like the IMF. In addition to its comparatively small size, for example, the CMIM partly failed to prove itself as a strong institution capable of providing an insurance mechanism during the recent crisis (ADB 2010; Aizenman et al. 2011; Sussangkarn 2011).

In regional liquidity pooling mechanisms, the coordination requirements in terms of fiscal policies and financial market regulation are rather low – as compared not only to the case of monetary integration projects such as the euro, but also to other forms of regional monetary cooperation. This leads to a more fundamental question about the kind of coordination required with regard to fiscal policies, liquidity provision and financial regulation in order to create a liquidity pool at the regional level.

First, in a regional reserve pooling mechanism monetary policy is independently formulated by each of the member countries, and hence regional fiscal policy coordination does not seem necessary. At the same time, a regional reserve pooling mechanism requires an incentive structure and transparent conditionality criteria that ensure sustainability. Along with Ocampo and Titelman (2009), who point out that regional ownership of the common liquidity pool facilitates the enforcement of conditionality criteria, one may consider such an ownership structure a *de facto* substitute for fiscal coordination.

Second, when evaluating the stability of regional liquidity pooling schemes, the liquidity provision in foreign exchange (typically the US dollar) to solvent yet illiquid sovereign states, through the pooling of reserves, represents the core feature of this form of regional monetary cooperation. In contrast to a regional central bank in a common monetary area such as the euro, these schemes can be described as a kind of 'quasi-LOLR'. In the context of our analysis of the fallacies in the design of the ECB and revealed by the current euro crisis, whether such a mechanism is able to prevent a self-fulfilling crisis from unfolding strongly depends on three factors: (i) the absolute size of the regional pooling; (ii) the relative size of liquidity available for single member countries; and, (iii) the symmetry between the reactions of member countries to the crisis.

The question regarding absolute size (i) is easy to address: the higher the amount of shared liquidity, the higher its potential stabilization effect. Yet, especially with regards to the relative size of liquidity available to individual member countries (ii), the effectiveness of regional liquidity sharing for each member country varies according to the size of the member: smaller member countries usually benefit more. At the same time, while a regional reserve pool requires the existence of diversified and financially strong members of a large size who can substantially contribute to the volume of the pool, larger countries benefit less since the amount of funds available are likely to be too small relative to their liquidity needs. Regarding the symmetry of the member countries' reaction to shocks (iii), regional self-insurance mechanisms only work if the pooled resources are not drawn on by all member countries at the same time (Eichengreen 2006; Imbs & Mauro 2007). On the one side, asymmetric business cycles and crisis reactions provide for an effective use of the regional pooling mechanism, while on the other they may endanger further regional monetary cooperation if the member countries develop too disparately – as has been correctly identified by OCA approaches.

Finally, as regional liquidity pooling of foreign exchange reserves does not per se create an incentive for regional financial integration, this mechanism does not necessarily require any efforts to be

made in terms of coordinating financial regulation beyond the multilaterally agreed standards.

Fondo Latinoamericano de Reservas (FLAR)

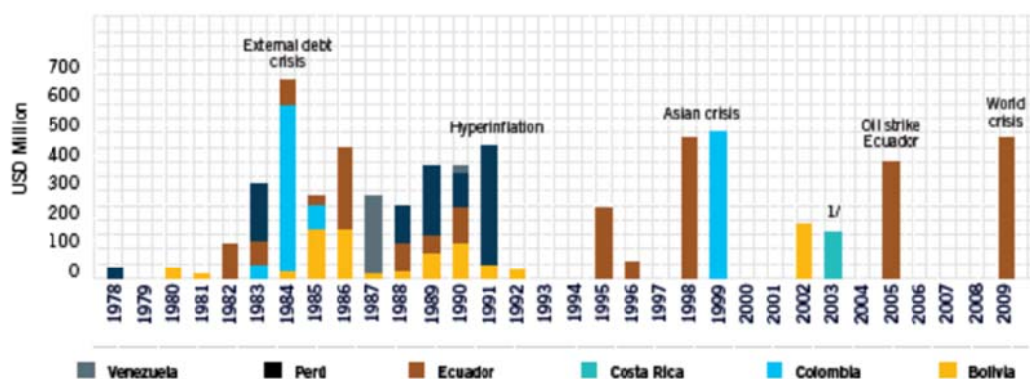
The FLAR was founded in 1978, and currently has six member countries originating from the Andean region – Bolivia, Colombia, Ecuador, Peru, Uruguay and Venezuela plus Costa Rica,. It is of a rather small size, with a subscribed capital of US\$ 2.3 billion and paid-in capital that did not reach US\$ 2 billion in 2010 (FLAR 2010). The FLAR operates as a trust where central banks from member countries may borrow in proportion to their capital contribution in five differently designed credit facilities. The paid-in capital of the member countries defines their borrowing capacity from the fund. The FLAR is a quickly disbursing institution, having less strict conditionality associated with borrowing than the IMF does (Ocampo 2006).

With regard to balance of payment support and liquidity provision, the FLAR disbursed credits of US\$ 5.3 billion between 1978 and 2007. The largest share was disbursed in the 1980s, as well as during the crises of 1998–1999; in these periods, FLAR financing was larger than that of the IMF (Ocampo & Titelman 2009). Bolivia and Ecuador draw most funds as recipient countries (above 50 per cent of all disbursements), while Venezuela – the largest economy among the member countries – is a non-borrowing member. During the recent financial crisis the aforementioned one-sided benefit for smaller countries was true in the case of the FLAR too, since only Ecuador relied on FLAR borrowing to counter potential crisis contagion (Moody's 2009). In terms of the aforementioned considerations for fiscal cooperation, the FLAR's excellent performance rates are based on strong enforcement conditions that appear to be a de facto substitute for fiscal rules. The FLAR has a zero default rate that Ocampo (2006) links to the strong sense of ownership that exists among members of the regional fund (FLAR 2011).

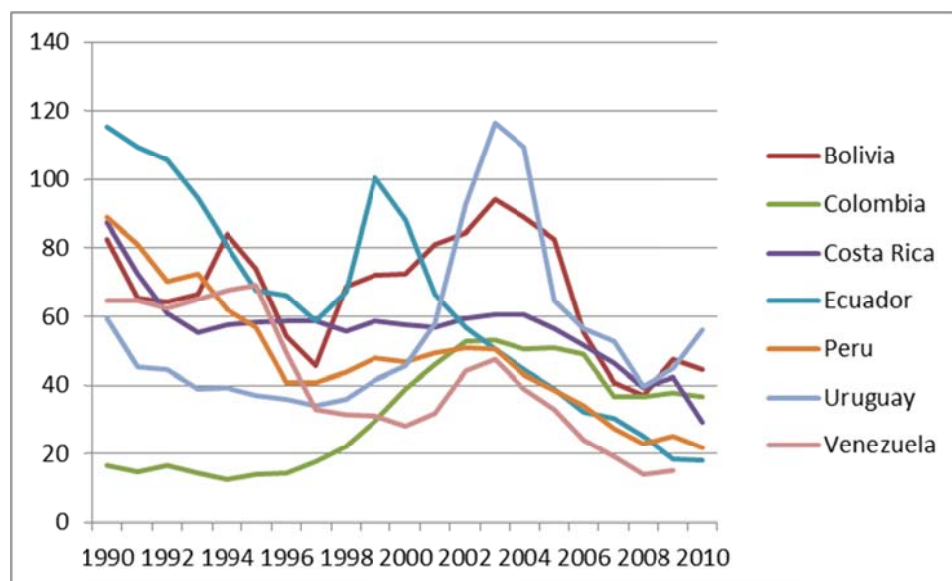
We will now look for the case of the FLAR at the three aforementioned criteria for the stability of regional liquidity pools and their ability to prevent a liquidity crisis from turning into a deep solvency crisis. Even though it is able to counter-cyclically provide liquidity to member countries, overall borrowing has so far been limited in terms of the total volume taken (see Figure 3). One reason for this is that the fund's absolute size (i) is very small.

The small size of the FLAR may eventually be compensated for by its high speed of liquidity provision, a feat which has been repeatedly praised by its member countries (Ocampo 2006). Due to the absence of conditionality, the FLAR was able to react quickly – as compared to most kind of IMF disbursements – to the liquidity demands of member countries. However, given the moderate debt-to-GDP ratios of the member countries the mechanism seems to provide a stable mechanism of temporary liquidity provision – at least for the smaller ones (see Figure 4).

Figure 3: The FLAR – Approved Credit



Source: FLAR (2011)

Figure 4: FLAR Members: government debt as a per cent of GDP

Source: Reinhart & Rogoff (2009). Note: government debt as per cent of GDP defined as total (domestic plus external) gross central government debt/GDP (where gross central government debt is not available, general government debt was used).

With regard to its relative weight for single member countries (ii), the comparison of FLAR credit disbursements to those of the IMF for Bolivia and Ecuador shows the relative importance and use of the FLAR for the small member countries (see Figure 4). This would be different if, for instance, Venezuela – as the largest member country – would draw on the fund and hence quickly exceed the latter's volume.

Table 2: Credit provided by the FLAR and the IMF to FLAR member countries (in US\$ million)

Time Period	FLAR	IMF	FLAR/IMF
1978–1981	90	626	0.1
1982–1985	1382	1005	1.2
1986–1989	1372	1494	0.9
1990–1993	860	3228	0.3
1994–1997	267	1012	0.3
1998–2001	994	305	3.3
2002–2005	737	342	2.2
2006–2010	480	0	N/A

Source: Machinea & Titelman (2007) (translated into English by ourselves)

Regarding the symmetry between the reactions of member countries to crisis (iii), the business cycles are sufficiently uncorrelated, providing for certain stability in this regional cooperation arrangement (Machinea & Titelman 2007; Ocampo & Titelman 2009).

Chiang Mai Initiative Multilateralization (CMIM)

In 2000 the original ASEAN-5 countries – Indonesia, Malaysia, Philippines, Singapore and Thailand – expanded their original bilateral swap arrangements (ASEAN Swap Arrangement, ASA) to include bilateral swaps between each of these five countries and China, Japan and South Korea (the 'plus-three' countries),

known collectively as the Chiang Mai Initiative (CMI). At this point, the ASEAN member countries – Brunei, Cambodia, Laos, Myanmar and Vietnam – who had joined more recently were not involved. The funding underpinning the CMI was later expanded and increased in size to US\$ 90 billion. The CMI was multilateralized in 2010 (known as CMI Multilateralization, CMIM).¹¹ Meanwhile, reserve accumulation in Southeast Asia boomed – with about half of total world reserve holdings and a total amount of more than US\$ 3 trillion being held in China alone by the end of 2011. The perceived need for a regional liquidity pool has hence diminished in light of these developments (ADBI 2011).

First, with regard to fiscal cooperation, one needs to take into account the link between the CMIM and the IMF. A core conditionality of the CMIM, similar to the CMI set-up, is that only 20 per cent of the swap amount can be used without it being done under an IMF-supervised programme. This IMF-linked conditionality can be considered as a *de facto* substitute for regional fiscal cooperation. On the one hand, this may significantly increase the stability of the mechanism. On the other, however, following Ocampo and Titelman (2009), IMF involvement has a detrimental effect in that it reduces regional ownership and thus weakens the enforceability of the regional cooperation arrangement. In addition, the IMF's involvement in the lending facilities for disbursements of more than 20 per cent of the country's drawing rights involves an inevitable time lag that has to be seen as a major hindrance to speedy short-term liquidity provision, and may thus need to be discussed again regarding future use by the CMIM. During the global financial crisis, CMI swaps were not yet useable since a surveillance unit was still missing, and hence countries were unwilling to participate in currency swaps. While the process of setting up the CMIM is ongoing, the AMRO is supposed to fill this gap (Rana 2011).

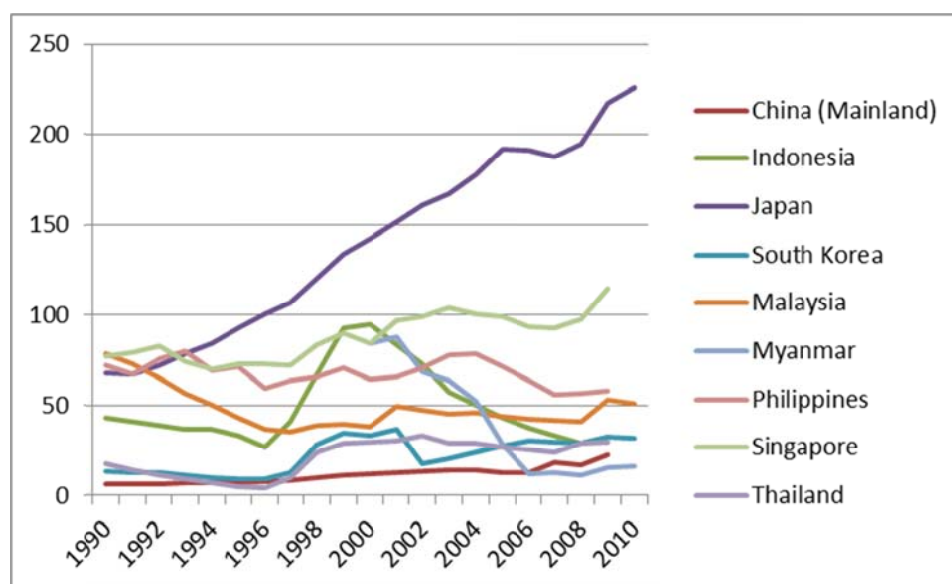
Second, regarding the efficiency of liquidity provision to illiquid yet solvent sovereign states, the absolute size (i) of the CMI has been growing substantially over time. Yet, there is an intense debate currently taking place about the adequate size of regional liquidity pooling in Asia (see ADBI 2011a).

During the global financial crisis, most emerging economies were in the position to draw on their own accumulated foreign reserves to counterbalance the consequential reduction in foreign financing. In addition, however, rather than relying on existing regional mechanisms such as the CMIM or international mechanisms of liquidity provision channelled through the IMF, some countries were offered the possibility to draw on bilateral swap arrangements with the Federal Reserve Bank of the US (Brazil, Mexico, South Korea and Singapore). It was South Korea that, in addition to drawing on its own foreign exchange reserves, made use of this swap arrangement with an amount of US\$ 30 billion. Yet, it needs to be taken into account that such bilateral swap arrangements are provided in the economic and political interests of the offering country, in this case the US – and hence two important aspects should be borne in mind. First, bilateral swap arrangements are a viable option of liquidity provision only for well-developed emerging markets, while, second, *ad hoc* bilateral swap arrangements cannot be considered a satisfactory medium- to long-term strategy in response to liquidity crises (see also Aizenman et al. 2011). Hence, reconsidering the core design and set-up of regional mechanisms such as the CMIM remains important (see also Volz 2012).

With regard to the relative size of the CMIM for its member countries (ii), it is the role of China and Japan in particular that gives the mechanism a very high level of stability. Both countries participate with very high shares in liquidity provision, yet *de facto* are not expected to draw on these funds – due to their domestically accumulated foreign reserves which far exceed the volume of the CMIM. On the contrary, and especially for the more recent ASEAN members such as Brunei and Vietnam, the current level of CMIM liquidity would be more than sufficient. As the debt-to-GDP ratios of ASEAN+3 countries illustrate, smaller member countries in particular do not endanger the mechanism as a whole and therefore benefit the most from it as borrowers (see Figure 5). If China and Japan were to draw on the CMIM, however, the mechanism certainly could not be considered a stable arrangement.

¹¹ The CMIM's financial size is currently US\$ 120 billion, with ASEAN countries contributing 20 per cent while China, Japan and South Korea provide the remaining 80 per cent – wherein China and Japan each share 40 per cent of the burden. The Executive Level Decision Making Body (ELDMB) – the deputy level of the member countries' finance ministers – is entitled to decide upon lending, renewal and default by a two-thirds majority. The ASEAN+3 Macroeconomic Research Office (AMRO) is responsible for managing the CMIM.

Figure 5: ASEAN+5 members: government debt as a per cent of GDP*



Source: Reinhart & Rogoff (2009).

*Note: Government Debt as per cent of GDP defined as total (domestic plus external) gross central government debt/GDP (where gross central government debt is not available, general government debt was used).

At the same time, high business cycle synchronization (iii) due to far-reaching trade and financial integration, outlined in more detail below, may de facto limit the effectiveness and stability of this regional mechanism – as the member countries tend to demand international liquidity in a rather synchronized manner that may exceed the fund's actual volume.

4.1 Regional financial market initiatives

One possible way that developing countries could avoid dependence on costly foreign finance – and the knock-on effect of financial instability – is by working to foster regional financial market development, especially with respect to bond markets. Such a strategy has received a degree of attention within RMCs, for example through the Asian Bond Markets Initiative (ABMI) of the ASEAN and its three partner countries China, Japan and South Korea (ASEAN 2008). Creating regional financial markets through the regional expansion of issuance and the demand for local currency bonds (LCBs) represents one promising strategy for enhancing financial development and providing financial stability (Borensztein et al. 2008).

Regional financial market development initiatives aim to provide market infrastructure and financial means for commencing and/or intensifying regional and international local currency lending and borrowing for member countries. Such initiatives hence focus on both the supply and the demand side of the bond market, and can be established through a regional multilateral financial institution – be it a multilateral development bank or a regional fund. A regional multilateral financial institution is likely to be in a better position than individual countries are to attract international investments – and thus to facilitate not only the issuance of LCBs but also the increased demand for them, thereby bridging the gap between international and regional financial markets (Birdsall & Rojas-Suarez 2004). Additionally, regional financial markets can be developed through the collaborative creation of a regional market for LCBs – achieved through the provision of the necessary infrastructure and funding at the regional level.

Such regional financial market development initiatives can be tailored to the participating countries' requirements regarding financial stability and market sophistication (UNCTAD 2007). Furthermore, regional markets provide an opportunity to introduce additional innovative financial instruments that involve less risk of balance-sheet mismatches and financing costs for the issuing countries (Lee & Park 2010). For the

majority of smaller developing economies in particular, the creation of financial markets at the regional level is more likely to result in success than if each country were to individually try to establish a market for local currency, debt instruments or other financing mechanisms (Eichengreen et al. 2006).

The coordination mechanisms required for regional financial market development initiatives to be stable in the aforementioned sense may be best understood as a process that increasingly gains importance with stronger cross-border financial links between the member countries' currencies.

First, fiscal coordination is increasingly required to the extent that rising regional cross-border holdings of government bonds increase the danger of regional spillovers of liquidity and solvency crises among neighbouring sovereign states, as the case of the euro area demonstrates. Hence, the destabilizing effects of national sovereign debt crises at the regional level increase with more successful initiatives. In case of a successfully developed financial market initiative, fiscal coordination may not only be a supporting factor for developing regional markets but may also become a crucial ingredient in the medium- to long-term survival of the initiative.

Second, the likelihood of a liquidity crisis turning into a solvency crisis depends on the economic size and strength of its member countries and the success of the initiative. Financial market initiatives aim to increase liquidity provision in local currencies, thus increasing the regional central banks' capacity to act as a LOLR. Again, intra-regional asymmetries seem to play a crucial role in the degree of stability that regional financial market initiatives experience. Stability may increase substantially if at least one financially strong member country exists that is in a position to act as a benchmark for the development of market standards and infrastructure – and as a market maker in the establishment of LCB markets. Such mechanisms gain additional stability if smaller member countries indebted themselves in the currency of such an 'anchor' country, and are thus able to service their debt through intra-regional trade surpluses against it. Furthermore, the anchor country provides for the financial endowment of either a regional multilateral institution or a regional fund that can help to develop the demand for LCBs.

Third, the design of the initiatives' approach to regional financial market regulation is important in order to prevent the stagnation or break-up of the arrangement upon increasingly successful cross-border financial integration. Regional LCB issues are intended to decrease risks linked to unhedged foreign currency debt/foreign exchange exposure. Depending on the degree of success and sophistication of the arrangement, as mentioned above cross-border LCB issues may increase the risks associated with domestic financial crises if the national and/or regional financial markets are under-regulated. With a highly integrated regional financial market, such dangers of contagion – in terms of a lack of surveillance and of prudential regulation – exist especially if individual financial institutions develop a regional systemically important size whose bail-out may affect the financial stability of other member countries. Hence, with the advent of increasing success, regulatory oversight is needed to prevent the arrangement from collapsing in the event of an external shock.

Asian bond market initiatives I+II

Apart from regional liquidity pooling, the second major field of regional cooperation in the ASEAN/ASEAN+3 is the development of financial market initiatives (Ma & Remolona 2006). Three major initiatives are currently in place to enhance regional financial market development. The first of these is the Asian Bond Market Initiative (ABMI) that was launched in 2002, with the aim of developing liquid primary and secondary bond markets and of recycling external surpluses into financing investment within the region (ASEAN+3 2008).

Within the framework of the ABMI, the Asian Development Bank (ADB) plays a crucial role: first, as a market maker for regional financial markets; second, as a facilitator of regional policy dialogue; and, third, as a promoter of the dissemination of information. The ADB's market-making role in financial market development primarily reveals itself in the provision of credit and political risk guarantees and bond issues denominated in regional currencies. Also, the ADB provides technical assistance to strengthen market infrastructure for regional bond markets. The bank's triple-A credit rating and diversified risk-structure of its portfolio play a crucial role as they enable the bank to raise funds in international financial markets at more favourable conditions than most of the member countries themselves. They also allow the ADB to attract extra-regional market participants to regional financial markets.

Second and third, in addition to the ABMI, two complementary initiatives were set up in 2003 and 2004 respectively: Asian Bond Funds (ABF) I and II (Ma & Remolona 2005). ABF I has a capitalization of US\$ 1 billion and is managed by the Bank for International Settlements (BIS). ABF I invests in US dollar-denominated bonds issued by governments or quasi-government institutions of eight of the Executives' Meeting of East Asia Pacific Central Banks (EMEAP) countries – namely China, Hong Kong, South Korea, Indonesia, Malaysia, the Philippines and Thailand. Ma and Remolona (2005) note that ABF initiatives were the first in which a regional financial institution contributed resources to the setup of a regional bond fund, based on regional reserve pooling. ABF II was established about a year later, so as to directly address the problem of currency mismatches: ABF II is capitalized with US\$ 2 billion and takes on investments in local currency-denominated bonds issued by the same eight East Asian governments or quasi-government institutions (Henning 2005).

Their still small size to date does not allow us to draw final conclusions at the regional level about ABF I+II's requirements for fiscal cooperation and financial regulation, nor about the extension of the LOLR function to illiquid sovereign member states. Overall, investment opportunities in the region's local currency bonds have increased. From the point of view of the provision of liquidity in local currency, and hence the minimization of exposure to foreign exchange volatility, ABF I+II have developed into relatively stable mechanisms. Yet, more requirements in terms of coordination efforts will be needed with the further development of both these initiatives. As such, ABF I+II cannot be evaluated in light of temporary liquidity provision but rather by assessing their contribution to preventing negative spillovers from external crises.

During the global financial crisis, Southeast Asia generally experienced little negative spillover effects – yet this was more due to strong economic growth in the region and its financial market conditions, as well as to its minimal exposure to US subprime mortgages. Both the financial market development initiatives and the CMIM were at that time not (yet) readily available mechanisms for weathering the storms of international volatility. The region hence needs to further develop their own mechanism(s) for regional temporary liquidity provision and regional financial market development and regulation. In this vein, the global financial crisis considerably contributed to the enhancement of efforts regarding regional and domestic market development in Southeast Asia, which have been escalated so as to further reduce dependence on extra-regional capital inflows (Shimizu 2010).

5. Conclusion

This paper has asked what lessons regional monetary cooperation schemes around the world can learn from the current euro crisis. Specifically, we asked whether the neo-functionalist notion of an 'ever closer union' holds true – in other words, whether there are inherent dynamics that promote ever-increasing integration towards final political union once the integration process has started, or whether there might be a stable equilibrium of financial and monetary integration before political integration. We have defined such a stable equilibrium as the point at which a regional integration arrangement can absorb economic shocks without causing the breakup, scaling back or necessity driven deepening of the respective form of regional integration.

In the case of the euro zone, we are of the view that this arrangement runs the risk of breaking up if its institutional cooperation mechanisms are not deepened. Yet, we argue that tighter and more strict control of national budgetary policies – and especially government deficits and debt – which is the understanding of the German government when it refers to political union – and which has been embodied in the fiscal compact treaty agreed upon in late 2011 – is not the key challenge that needs to be addressed in order to solve this crisis. While some degree of closer fiscal cooperation in terms of surveillance and sanctioning mechanisms may be sufficient, we find that the full provision of LOLR facilities for solvent but illiquid sovereign debtors is the key to preventing further spillover effects diffusing to other member countries.

We found relevant lessons from developing countries to be taken into account for the euro crisis resolution. This includes the realization that liquidity and solvency problems need to be clearly distinguished from each other. In the case of liquidity issues ample funds need to be made available quickly and without unrelated conditionality. Liquidity provision by a third party, such as a regional central bank, can be highly

efficient in preventing negative expectations from becoming self-fulfilling according to models with multiple equilibria. In the case of insolvency, only debt restructuring that aims at a sufficient reduction of the debt level is capable of preventing further crisis aggravation and contagion effects on other countries in the region. Had the euro area followed these lessons – and had the member states promised funds more quickly to countries with liquidity issues such as Spain, while addressing to a sufficient extent early on the solvency issues in Greece – the recession that has occurred in the euro area might have been less harsh and the risks of a euro area breakup might have been mitigated earlier.

What lessons can the ongoing euro crisis offer to RMCs in other regions of the world? In this paper we have analyzed such two arrangements, one in Southeast Asia – the Chiang Mai Initiative Multilateralization, CMIM – and one in Latin America – Fondo Latinoamericano de Reservas, FLAR – that have as their intention the provision of international liquidity. Furthermore, we have also addressed the Asian Bond Markets Initiative (ABMI), which aims to increase the share of domestically denominated financial contracts in the region, in order to increase the capacity of the various domestic central banks to act as a LOLR.

The overall finding has been that, according to our definition, the institutional frameworks of these mechanisms showed a high degree of stability in terms of crisis resilience, even if the member countries of these arrangements were adversely and asymmetrically hit by shocks from the global financial crisis. The FLAR in particular plays a relevant role for its smaller member countries, even in the absence of any formal fiscal coordination and surveillance mechanisms.

Yet, at the same time, the experience of the euro crisis reveals how the underlying financial volume of a regional arrangement is relevant to the ability to provide liquidity with adequate timing and in sufficient quantity. In this light, both the FLAR and the CMIM appear rather small, especially for the bigger economies in the respective regions. One might thus ask if the ‘sense of ownership’ among the group of mostly small Andean countries might be sufficient as a coordination mechanism if the FLAR were enlarged. Related to the question of ownership, the lesson is different for the case of the Asian CMIM, which was founded in the aftermath of the Asian crisis and which has had access to an increasing volume of foreign exchange reserves over the course of the past decade. It seems that the current institutional setting, which externalizes fiscal surveillance for drawings from the CMIM fund facilities to IMF conditionality, has not destabilized the mechanism – yet at the same time it has caused a stigmatization effect, which makes the use of the common liquidity pool less attractive for its members. Thus, developing a more adequate institutional arrangement – especially for fiscal coordination – seems to be relevant for this nascent institution.

A second key lesson from the euro zone is that far-reaching financial integration makes a key difference in cooperation requirements, especially in the fields of fiscal and financial regulation, as the interdependence of financial institutions beyond national borders creates significant spillover effects. This is relevant especially for the ABMI, which explicitly aims at augmenting the financial integration of the member countries. The volumes that are involved in creating, among other endeavours, regional bond markets are still rather small, so the deepening of coordination seems to be an issue that will need to be faced in the near future.

We draw a third lesson from the problems that intra-regional asymmetries in terms of the economic size of member countries may cause for RMCs. For the case of the euro, due to the lack of regionally agreed and monitored financial regulation and oversight, smaller member countries such as Ireland previously had incentives for financial under-regulation. Thus they attracted large financial inflows and gains for their domestic financial sector that resulted in significant costs – in terms of the bailing out of financial institutions – that operated across borders in other eurozone countries. Thus for the case of Europe – and at the current level of financial integration – intra-regional asymmetries call for deeper cooperation.

On the contrary, in the case of emerging markets and developing countries, where financial integration up to now has not reached significant levels, the participation of bigger – and especially of financially stronger – economies seems to increase the stability and efficiency of liquidity-providing mechanisms, even in regions with low levels of economic coordination. In particular, the somewhat benevolent participation of regional leaders – such as Brazil in the case of the FLAR, or China and Japan in the case of Asian regional initiatives – may enable a significant increase in liquidity provision for the smaller

countries of these regions. Thus, a crucial task for future research is to identify the incentives that can be offered to entice engagement in these emerging markets, so that member states can be persuaded to actively participate in regional schemes based on their specific interests. Another related research question that will need to be tackled is the one of how intra-regional trade and financial imbalances can be adequately addressed in such arrangements.

In sum, the idea of an 'ever closer union', as first advocated by the neo-functionalist approach, in reality obscures the diversity of potentially stable institutional solutions that can precede the level of political integration – both in the case of the euro area and of the scrutinized arrangements in Southeast Asia and Latin America.

References

ADB 2010, *Institutions for Regional Integration. Towards an Asian Economic Community*. Available: <http://www.adb.org/documents/books/institutions-regional-integration/institutions-regional-integration.pdf>

ADB 2011, *Institutions for Economic and Financial Integration in Asia: Trends and Prospects*. Available: <http://www.adbi.org/files/2011.09.07.wp308.economic.financial.integration.asia.trends.prospects.pdf>

ADB 2011a, 'Lessons of the Eurozone Financial Crisis for Asia', *Presentation at the National Economic Outlook Conference 2012–13*, 22/11/2011. Available: <http://www.mier.org.my/presentations/archives/pdf/ADB1.pdf>

Aghion, P., Bacchetta, P., Ranciere, R. & Rogoff, K. 2009, 'Exchange Rate Volatility and Productivity Growth: The Role of Financial Development', *Journal of Monetary Economics*, vol. 56, pp. 494–513.

Aizenman, J., Jinjarak, Y. & Park, D. 2011, 'International Reserves and Swap Lines: Substitutes or Complements?', *International Review of Economics and Finance*, vol. 20, pp. 5–18.

ASEAN+3 2008, ASEAN+3 New ABMI Roadmap.

Bagehot, W. 1873, *Lombard Street: A Description of the Money Market*. Henry S. King and Co., London.

Bird G. 2010, 'The World's Liquidity Arrangements: The Easiest Item on a Holistic Agenda for International Monetary Reform?', In *Regional and Global Liquidity Arrangements*, eds U. Volz & A. Caliarì, German Development Institute, Bonn.

Birdsall, N. & Rojas-Suarez, L. (eds) 2004, *Financing Development – The Power of Regionalism*. Center for Global Development: Washington, DC.

Blanchard, O., Dell'Ariccia, G. & Mauro, P. 2010, *Rethinking Macroeconomic Policy*. Research Department, IMF, Washington, DC. Available: <http://www.imf.org/external/pubs/ft/spn/2010/spn1003.pdf>

Bordo, M.D., Meissner, C.M. & Stuckler, D. 2009, *Foreign Currency Debt, Financial Crises and Economic Growth: A Long Run View*. NBER Working Paper 15534, Cambridge, MA, Available: www.nber.org/papers/w15534

Borensztein, E., Cowan, K., Eichengreen, B. & Panizza, U. 2008, *Bond Markets in Latin America: On the Verge of a Big Bang?* MIT Press, Cambridge, MA.

Cole, H.L. & Kehoe, T.J. 1996, 'A Self-fulfilling Model of Mexico's 1994–1995 Debt Crisis', *Journal of International Economics*, vol. 41, pp. 309–30.

Connor, G., Flavin, T. & O'Kelly, B. 2010, *The U.S. and Irish Credit Crises: Their Distinctive Differences and Common Features*. Irish Economy Note No. 10. Available: <http://www.irisheconomy.ie/Notes/IrishEconomyNote10.pdf>

Dieter, H. 2010, Die Internationalen Finanzmärkte stellen die Eurozone auf die Probe. SWP Aktuell, Berlin.

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

Dullien, S. 2009, 'Divergences in EMU: Scope of the Problem and Policy Options', *Intervention: European Journal of Economics and Economic Policies*, vol. 6, pp. 24–32.

Dullien, S. & Schwarzer, D. 2010, *Umgang mit Staatsbankrotten in der Eurozone: Stabilisierungsfonds, Insolvenzrecht für Staaten und Eurobonds*. SWP-Studie S 19, Berlin.

Dullien, S. & Guérot, U. 2012, *The Long Shadow of Ordoliberalism: Germany's Approach to the Euro Crisis*. European Council on Foreign Relations Policy, Brief 49, London.

ECB 2010, *Decision of the European Central Bank of 14 May 2010 Establishing a Securities Markets Programme (ECB/2010/5)*, Frankfurt. Available:

http://www.ecb.int/ecb/legal/pdf/l_12420100520en00080009.pdf

Eichengreen, B. 2006, *Insurance Underwriter or Financial Development Fund: What Role for Reserve Pooling in Latin America?* Available: www.econ.berkeley.edu/~eichengr/research/insur_under_july06.pdf

Eichengreen, B., Borensztein, E. & Panizza, U. 2006, *A Tale of Two Markets: Bond Market Development in East Asia and Latin America*. Hong Kong Institute for Monetary Research Occasional Paper No. 3, Hong Kong.

Eichengreen, B. & Hausmann, R. (eds) 2005, *Other People's Money – Debt Denomination and Financial Instability in Emerging Market Economies*, University of Chicago Press, Chicago.

Eichengreen, B., Prasad, E. & Rajan, R. 2011, *Rethinking Central Banking*. Committee on International Economic Policy and Reform.

Enderlein, H., „Griechenland muss sich selbst helfen“, *Süddeutsche Zeitung*, 9 March 2010, p. 2.

EU Commission 2011, *Green Paper on the Feasibility of Introducing Stability Bonds*. EU Commission, Brussels.

FLAR 2011, *History and Performance in Perspective*. Available:

<https://www.flar.net/ingles/contenido/contenido.aspx?catID=215&conID=700>

Fritz, B. & Mühlich, L. 2012, *Varieties of Regional Monetary Cooperation in Emerging Areas: Systemizing Different Forms and their Potential for Reducing Volatility*. FU Berlin, Manuscript

Giavazzi, F. & Spaventa, L. 2010, *Why the Current Account may Matter in a Monetary Union: Lessons from the Financial Crisis in the Euro Area*. CEPR Discussion Papers 8008.

Haas, E. 2004, *The Uniting of Europe*, Stanford University Press, Stanford.

Halac, M. & Schmukler, S. 2004, 'Distributional Effects of Crises: The Financial Channel', *Economia*, vol. 5, pp. 1–68.

Hausmann, R. & Panizza, U. 2010, *Redemption or Abstinence? Original sin, Currency Mismatches and Counter-Cyclical Policies in the New Millennium*. Centre for International Development at Harvard University, Working Paper No. 194.

Henning, C.R. 2005, *Systemic Contextualism and Financial Regionalism: The Case of East Asia*. Available: http://www.cgu.edu/include/1118_henning1.doc.

Henning, C.R. 2011, *Coordinating Regional and Multilateral Financial Institutions*. WP 11-9. Available: <http://www.iie.com/publications/wp/wp11-9.pdf>

Honohan, P. 2009, 'Resolving Ireland's Banking Crisis', *The Economic and Social Review*, vol. 40, pp. 207–221.

Imbs, J. & Mauro, P. 2007, *Pooling Risk among Countries*. IMF Working Paper WP/07/132, Washington, DC.

IMF 2002, *Assessing Sustainability*. Prepared by the Policy Development and Review Department, in consultation with the Fiscal Affairs, International Capital Markets, Monetary and Exchange Rate Affairs, and Research Departments, Washington, DC.

Lane, P. 2011, *The Dynamics of Ireland's Net External Position*. Available: <http://www.philiplane.org/SSIS/2011LANE.pdf>

Lavery, B. & O'Brian, T.L., 'For Insurance Regulators, Trails Lead to Dublin' *The New York Times*, 1 April 2005. Available: <http://www.nytimes.com/2005/04/01/business/worldbusiness/01irish.html>

Lee, J.W. & Park, C.Y. 2010, *New Financial Reforms and Their Effects on Emerging Asia*. ADB Briefs No. 5/2010. Available: <http://www.adb.org/publications/new-financial-reforms-and-their-effects-emerging-asia>.

Ma, G. & Remolona, E. 2005, 'Opening Markets through a Regional Bond Fund: Lessons from ABFI', *BIS Quarterly Review*, pp. 81–92.

Ma, G. & Remolona, E. 2006, *Learning by Doing in Market Reform: Lessons from a Regional Bond Fund*. HWWI Research Paper No. 2-6, Hamburg Institute of International Economics, Hamburg.

Machinea, J.L. & Titelman, D. 2007, 'Less Volatile Growth? The Role of Regional Financial Institutions' *CEPAL Review*, vol. 91, pp. 7–28. Available: <http://www.eclac.cl/publicaciones/xml/7/29497/lcg2333iMachinea.pdf>

Mateos y Lago, I., Duttgupta, R. & Goyal, R. 2009, *The Debate on the International Monetary System*. IMF Staff Position Note 09/26.

McGowan, M.A. 2011, *Overcoming the Banking Crisis in Ireland*. OECD Economics Department Working Paper No. 907, Paris.

Moodys 2009, *Credit Analysis Fondo Latinoamericano de Reservas*. Available: https://www.flar.net/documentos/2580_Moodys.pdf

Moravcsik, A. 2005, 'The European Constitutional Compromise and the Neo-Functionalist Legacy', *Journal of European Public Policy*, vol. 12, no. 2, pp. 349–386.

Mundell, R. 1961, 'A Theory of Optimum Currency Areas' *American Economic Review*, vol. 51, pp. 509–517.

Mundell, R. 1963, 'Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates', *Canadian Journal of Economics and Political Science*, vol. 29, no. 4, pp. 475–485.

Obstfeld M. 1996, 'Models of Currency Crises with Self-fulfilling Features', *European Economic Review*, vol. 40, pp. 1037–1047.

Ocampo, J.A. 2006, 'Regional Financial Cooperation: Experiences and Challenges', in *Regional Financial Cooperation*, ed. J.A. Ocampo, Brookings Institution Press, Washington, DC.

Ocampo, J.A. & Titelman, D. 2009/10, 'Subregional Financial Cooperation: The South American Experience', *Journal of Post-Keynesian Economics*, vol. 32, no. 2, pp. 249–269.

Panizza, U. 2006, 'Original Sin and Monetary Cooperation', in *New Issues in Regional Monetary Coordination – Understanding North–South and South–South Arrangements*, eds B. Fritz & M. Metzger, Palgrave Macmillan, London.

Rana, P. 2011, *Asian Monetary Fund: Getting Nearer*. RSIS Commentaries No. 79/2011. Available: <http://www.rsis.edu.sg/publications/Perspective/RSIS0792011.pdf>

World Economic Review

Reinhart, C.M. & Rogoff, K.S. 2009, *This Time is Different: Eight Centuries of Financial Folly*. Available: <http://www.reinhartandrogoff.com/data/browse-by-topic/topics/9/>

Sachs, J.D., Kawai, M., Lee, J. & Woo, W.T. (eds.) 2010, *The Future Global Reserve System – An Asian Perspective*. Available: [http://aric.adb.org/grs/papers/Kawai 5.pdf](http://aric.adb.org/grs/papers/Kawai%205.pdf)

Shimizu, S. 2010, *The Development of Asian Bond Markets since the Global Financial Crisis – Significance and Challenges*. No. 38, RIM Pacific Businesses and Industries, vol. X. Available: <http://www.jri.co.jp/MediaLibrary/file/english/periodical/rim/2010/38.pdf>

Stiglitz, J. 2002, *Globalization and Its Discontents*, Penguin Books, London.

Sussangkarn, C. 2011, *Institution Building for Macroeconomic and Financial Cooperation in East Asia*. Available: <http://www.amro-asia.org/wp-content/uploads/2011/11/CS-CMIM-AMRO.pdf>

UNCTAD. 2007, *Trade and Development Report: Regional Cooperation for Development*, United Nations Publications, New York & Geneva.

Volz, U. 2012, *Lessons of the European Crisis for Regional Monetary and Financial Integration in East Asia*. ADBI Working Paper 347, Asian Development Bank Institute, Tokyo. Available: <http://www.adbi.org/working-paper/2012/02/21/5007.lessons.european.crisis.east.asia/>

Wheelan, K. 2011, *Ireland's Sovereign Debt Crisis*. UCD Economic Research Center Working Paper WP11/09, Dublin.

Williamson, J. 1976, 'The Benefits and Costs of the International Monetary Nonsystem', *Essays in International Finance*, No. 115, Princeton University, Princeton, NJ.

Institutional Support and Technological Upgrading: Evidence from Dynamic Clusters in Latin America and Asia

Rajah Rasiah¹

University of Malaya, Kuala Lumpur, Malaysia

Jebamalai Vinanchiarachi²

University of Malaya, Kuala Lumpur, Malaysia

Abstract

In light of wide differences in economic outcomes in the world, this paper uses an evolutionary set of lenses to examine the clusters of Buenos Aires' automotive, Los Lagos' salmon, Penang's electronics and Qiaotou's buttons with an elucidating view towards evaluating the significance of institutional support in driving technological upgrading in firms. The purpose is to demonstrate if industrial and location specificities and industrial policy instruments matter in upgrading outcomes. The results show that transnational corporations drove automotive and electronics clusters in Buenos Aires and Penang respectively, while domestic firms dominated the origin of salmon and button clusters in Los Lagos and Qiaotou. Domestic organizations have been the prime drivers of upgrading in Los Lagos and Qiaotou. Whereas the meso organizations in Los Lagos adapt knowledge from frontier clusters abroad, they are the basis of knowledge generation in Qiaotou. Whatever the differences, the role of government through institutional change has been critical in stimulating upgrading, but the extent and nature of intervention in the four clusters were industry and location specific.

Key words: institutions, technological upgrading, economic synergies, clusters

1. Introduction

Dynamic clusters have originated in several ways (Best, 2001). Some emerged from the relocation of labour-intensive and low value added segments of value chains (Helleiner, 1973; Rasiah, 1988)). Some were deliberately spawned through government policy (Amsden, 1985; Lall, 1996). Others arose as a result of lengthy processes of skills and craft development and socio-economic interactions (Piore and Sabel, 1984; Pyke and Sengenberger, 1992; Rasiah, 1994, Becattini et al., 2009, Guerrieri et al., 2001). Some arose from natural resource endowments (Narula and Dunning, 2000). Export processing zones have also acted as the spearhead for the development of dynamic industrial locations (Rasiah, 2007). The development of clusters is often underpinned by the geographical agglomeration of firms of a given sector or different sectors and industrial networks.

Clustering in developing countries typically starts as fragmented activities: either as an agglomeration of domestic firms or of transnational corporations (TNCs) relocating labour- or resource-intensive activities from abroad. However, whatever the circumstances of origin, institutional direction have acted as critical drivers of technological upgrading and integration in export markets (Markusen, 1996).

This paper extracts some aspect of the evolutionary argument that technological upgrading experiences are industry and location specific (Nelson, 2008) to answer the question of why technological upgrading experiences are uniquely different by evaluating the experiences of four locations from the

¹ Corresponding author; comments from Jorge Katz, Norbert Haring, Jayati Gosh, Richard Nelson and Carlo Pietrobelli are gratefully acknowledged. While the data used in this paper comes from the United Nations Industrial Development Organization (UNIDO) that was collected through the supervision of the authors, it does not reflect the views of UNIDO. The usual disclaimer applies. The authors declare no conflict of interest.

² Former principal advisor to UNIDO.

developing world where clustering has reached reasonable maturity. Two clusters each were selected from Latin America and Asia to identify the drivers of technological upgrading and economic synergies.³

The first cluster, automotive in Buenos Aires, evolved initially from the 1950s when giant transnational corporations (TNCs) relocated subsidiaries to assemble automobiles for the domestic market. The cluster faded in the 1970s and 1980s before re-emerging strongly following reconstruction policies from 1991. The second cluster, salmon in Los Lagos, emerged in the 1970s following government efforts to promote the industry in Southern Chile to alleviate poverty (Chile, 2003). It is from the 1980s that the cluster began to grow strongly following strong coordination between meso organizations, the industry association and government instruments as successful incubators from Fundacion Chile (FC) were actively sold to private firms. The third cluster, electronics in Penang, emerged from export processing zones from the early 1970s when TNCs relocated assembly operations to take advantage of cheap labour and tax holidays (Rasiah, 1988). From assembly operations production was transformed to knowledge-intensive activities from the 1980s. The final cluster, buttons in Qiaotou, evolved when Italian firms outsourced manufacturing to Chinese firms since 1982 to check spiralling production costs. Knowledge flows from meso organizations helped the upgrading of button production from 2006.

Hence, the aim of this paper is to show how industrial and location specificities and the nature of institutional support drove technological upgrading in the four clusters. The rest of the paper is organized as follows. Section two discusses the main theoretical arguments. Section three presents the methodology and data. Section four analyses technological upgrading, and growth in exports, employment, wages, and skills intensities in the four clusters. Section five evaluates the drivers of economic synergies. Section six presents the conclusions.

2. Theoretical considerations

Since the focus of the paper is on institutional support clusters are used as examples of dynamic locations characterized by a unique configuration of economic activities. It is important to note that elements of industrial policy exists so long as governments strategize by targeting any or more of the following economic instruments - e.g. tariffs, tax holidays, grants, subsidies, preferential capitalization, specialized support for particular educational fields on skills and foreign exchange exemptions to stimulate industrialization. Hence, in this paper we consider efforts of governments to spawn particular industries through any of these elements even if no preference is offered to national capital over foreign capital or if protective tariffs are no longer used as elements of industrial policy. Also, we prefer to assess the original contributions to understand the critical concepts and relationships than reviewing the latest publications.⁴

There is recognition that mature clusters of economic activities integrated in global markets demonstrate strong elements of technological upgrading and economic synergies (Best, 2001, Becattini et al., 2009). Clusters here are defined as regionally or locally networked sets of economic agents that connect firms, organizations and institutions, and are considered to produce the most synergies when the requisite macro institutions, meso organizations and micro-agents are in sync to drive learning, innovation and competitiveness through circular and cumulative causal processes. What Young (1928), Abramowitz (1956), Kaldor (1967) and Cripps and Tarling (1973) argued at an abstract and aggregate level can be presented dynamically through the concept of clusters.

The systemic nature of knowledge flows was demonstrated by Marshall (1890), and Nelson and Winter (1982: 63). Scitovsky (1954) and Rosenstein-Rodan (1943: 207-208) distinguished non-pecuniary from pecuniary external economies. Nelson (2008) went further to discuss the imperfect, non-linear, and coordinated and uncoordinated nature of knowledge flows. The specificity of industries, initial structural conditions and the timing of upgrading strategies have also attracted different institutional roles (Hobday, 1995; Malerba et al, 2008). Mature firms in open integrated clusters gain new ideas to support continuous

³ Although there are considerable publications on the automotive and salmon clusters in Argentina and Chile, and the electronics cluster in Malaysia, we use in the paper fresh data collected in 2008-09. The story of the button cluster in Qiaotou is fairly new.

⁴ Unfortunately several journals are more interested in promoting the most recent literature even if they added no new knowledge to the old ones.

organizational change as old employees are replaced to make way for fresh ones, while new firms benefit from the released entrepreneurial and technical human capital to start new firms (Best, 2001; Rasiah, 1995). Saxenian (2006) documented the development and movement of human capital, which has supported new firm creation and several fast growing regions in the world.

However, unlike its modeling application by neoclassical economists who often reduce it to an exogenous black box (Rosenberg, 1982), knowledge appropriation requires considerable effort as pointed out by Lall (1992). The historical sequence of the development of technological capabilities through industrial policy started in Britain when Henry the VII imposed taxes on exports of wool in 1485 (Reinert, 1994: 175). A series of follow up industrial policies helped the United States, Germany, Sweden, Japan, Korea and Taiwan achieve technological superiority in increasing returns industries. In examining institutional support, we assume North's (1990) definition of institutions as the 'rules of the game', and meso organizations and firms the players. However, following Nelson (2008) we use a wide role for institutions with markets being only one of them.

Freeman (1988) had demonstrated using the experience of Japan that international flows of stocks of knowledge from developed to developing economies take a sequential movement from imports to adaptation, assimilation and innovation. Katz and Stumpo (2001: 137) observed movement of motor vehicle firms from Argentina and Brazil that were integrated through TNCs as trade transformed from process adaptation to designing activities from the 1990s.

While the Marshallian systemic doctrine of knowledge flows remains critical in the generation and diffusion of technological spillovers institutions other than markets, such as laws and government directives, trust relationships supported by particular socio-cultural and intermediary organizations have been no less important in driving technological upgrading (Piore and Sabel, 1984; Becattini et al., 2009). Also, as Hirschman (1970) had argued, host-governments have the potential for translating potential to real spillovers. Hence, whereas the government can support steep catch ups by supporting firms to appropriate latecomer advantages (Veblen, 1915; Gershenkron, 1952; Abramovitz, 1956), markets alone only offer gentle structural change experiences (see Chakravarthy, 1993).⁵

The strong and responsible government of South Korea ensured that performance standards drove technological catch up by Samsung in electronics, Hyundai in shipbuilding and automobiles and Posco in steel in the 1970s (Amsden, 1989; Kim, 1997). The government in South Korea also insulated successful chaebols from the destabilization caused the oil crisis of 1973-75. The Industrial Technical Research Institutes created by the Taiwan government in 1974 drove technological catch up *inter alia* in machinery (Amsden, 1985; Fransman, 1986). The government financed the acquisition of Radio Company of American (RCA) in 1977-79 and the founding of the joint-venture company of Taiwan Semiconductor Manufacturing Corporation (TSMC) with Philips in 1987, which by the end of 2000 had become the world's leading contract manufacturer of fabricated wafers.

Governments can create or strengthen institutions to promote agglomeration effects, and also screen particular clusters to identify bottlenecks, gaps and weaknesses. Such problems can take the form of weaknesses in basic infrastructure, high tech infrastructure, network cohesion or integration in global markets (Rasiah, 2009). Given the problems of information asymmetries between government and firms, and the public good characteristics of knowledge, intermediary organizations such as chambers of commerce, training institutions and R&D labs often help solve collective action problems. Interdependent relationships that are driven by the discipline of the market, direction of government when public goods are involved and complementation through trust-loyalty to extract social commitment from the participating humans has been vital in the development of competitive clusters (Brusco, 1982). Stakeholder coordination (e.g. through industry, government, consumer and labour coordination councils) often help root and expand social capital.

A lack of firm-level drive, human capital, high tech institutions and external competition to stimulate innovation and competitiveness have often undermined the capacity of clusters to enjoy sustainable differentiation and division of labour, which also explains the stagnation that has characterized industrial estates in many developing economies.

⁵ Chakravarty (1993) advanced the notion that markets are good servants but bad masters.

Frontier clusters are characterized by innovation. The focal point of innovation in a dynamic cluster is essentially the interdependent and interactive flow of knowledge and information among people, enterprises and meso organizations, which must include coordination between the critical socioeconomic and technological agents across value chains who are needed in order to turn ideas into processes, products or services in the marketplace. In dynamic clusters, such as, the Silicon Valley, Route 128 and the Hsinchu Science Industrial Park, innovations evolve from a complex set of inter-relationships among actors located in a range of enterprises, universities and research institutes. The role of user-producer interactions in driving learning and innovation was articulated lucidly by Lundvall (1992). The effective execution and appropriation of innovations often requires the involvement of other actors in dynamic clusters, such as intermediary organizations (including R&D labs), suppliers, venture capitalists, property rights lawyers and marketing specialists. While it is possible for large firms to internalize the function of creating public goods in R&D labs, the embedding meso organizations must evolve to higher levels to support firms' participation in R&D activities. The United States' government funds strategic research in the military, universities and other laboratories in recognition of knowledge as a public good (Mazuzan, 1994). Also, clusters do not evolve in a linear way. They have often started at different levels of integration and follow different trajectories (Guerrieri et al., 2001). In addition, although some clusters have fragile origins in developing countries, firms may still appropriate certain advantages from agglomeration economies and meso organizations (see Rasiah, 2009, 2010).

Value chains of particular commodities in dynamic clusters are globally integrated in factor and final markets. Global markets provide the economies of scale and scope and the competitive pressure to innovate. Global value chains assist economic agents in clusters to orientate their strategies to the critical dynamics that determine upgrading and value addition (Gereffi, 2002; Gereffi, Humphrey and Sturgeon, 2005, Pietrobelli and Rabellotti, 2007). Examples of such changes include the introduction of cutting edge just-in-time and flexible specialization techniques in electronics, the proliferation of software technology in the use of cad-cam machines and the interface between firms' assembly activities and major markets. However, the depth of integration and potential for technological upgrading in particular host-sites in value chains depends very much on the embedding institutions and meso organizations.

Demand-supply influences from buyers and users in domestic and export markets and the embedding institutions and organizations play a critical role in driving technological upgrading in firms (Lundvall, 1992; Nelson, 2008). While it is important to examine the supporting strength of the embedding environment it is also important to evaluate its impact on the level of technological activity of firms located in particular locations.

In doing so we attempt to widen the scope and flexibility to strengthen the old arguments in support of industrial policy, such as 'getting relative prices wrong' (Amsden, 1989), 'governing the market' (Wade, 1990) or targeting 'institutional quality' (Rodrik, 1997) so as to increase the explanatory power of heterodox approaches. We hypothesize that the nature of industrial policy pursued in the four clusters will be somewhat different because of industry and locational differences. Not only will initial conditions matter, the institutions and meso organizations created and the responses of the micro agents will also be critical for technological upgrading to occur.

3. Methodology and data

The empirical investigation was started with the identification of 10 clusters enjoying at least 20 years of experience in which export growth of lead products from the locations exceeded the national growth rates over the period 2000-2006 (see UNIDO, 2009).⁶ The four most dynamic upgrading experiences among the 10 clusters researched are used for the purpose of this paper. The locations selected, the principal initiators and breakdown of firms chosen by ownership are shown in Table 1.

⁶ Wages were estimated after deflating for inflation using consumer price indices supplied by national consultants.

Table 1: Dynamic industrial clusters, Latin America and Asia, 2006

Country	Industrial cluster	Product	Average annual growth rate of exports 2000-06(%)		Number of surveyed firms	Foreign ownership in sample* (%)	Origin
			National	Location			
Argentina	Buenos Aires	Automotive	11.7	15.3	50	70.0	Started as a special economic zone
Chile	Los Lagos	Salmon	11.1	18.6	50	37.5	Started as a specially designated salmon farming region in the early 1980s
China	Qiaotou	Buttons	12.0	12.4	100	0.0	Originated through entrepreneurs investing in button manufacturing in the 1980s
Malaysia	Penang	Electronics	4.6	6.7	103	83.7	Started with Free Trade Zones from 1972

* Ownership was classified as foreign if foreign equity constituted at least 50% of total equity of firms and the rest as national.

Source: UNIDO, *Survey of Dynamic Industrial Locations, 2007-08*

Only when firms enjoy technological upgrading will they be able to sustain long run growth in exports, value added, skills and wages. The brainstorming sessions with the industry association in Buenos Aires, Los Lagos, Santiago, Penang, Qiaotou and Hubei helped the delineation of the typology by taxonomies and trajectories. We differentiated technological upgrading to include all vertical and horizontal integration that entail participation in higher technological capabilities with a commensurate increase in value added. We preferred this alternative to the separate classifications by Schmitz (2004) because functional integration does not always translate into the utilization of higher knowledge content. For example, TSMC integrated lower value added wafer bumping to higher value added wafer fabrication (Yap and Rasiah, 2013).

The first round interviews with officials of the industry associations show that firm-level technological upgrading often rely extensively on knowledge accumulation stimulated by institutional change. The evidence we obtained shows that change not only requires the introduction and enforcement of the appropriate institutions but also the creation and strengthening of intermediary organizations to solve collective action problems. Drawing from Rasiah's (2007, 2009) work, four systemic pillars were identified and differentiated by taxonomies and trajectories to stimulate firms' movement from simple activities to the technology frontier (see Table 2). Despite the specificity of particular industries and locations, the evidence shows that a typology of taxonomy and trajectory can be drawn up to map these developments with sufficient amount of openness to allow for the capture of new information, which is consistent with Keynes' (1890) argument that inductive approaches cannot be assumed to evolve completely in a vacuum and are generally intertwined with deductive postulations. Table 3 shows a typology of taxonomies and trajectories of different levels of knowledge accumulation in firms, which is developed from the generic explication of taxonomies presented by Dosi (1982) and Pavitt (1984). In the formative years, incentives are often useful only in moving firms from level 1 to levels 2 and 3. Training and human capital organizations become

important from level 2 onwards, which increases in importance as clusters mature technologically. Grants become more important to attract firms' participation in levels 4 and 5 activities. R&D scientists and engineers, and R&D labs, and incubators become important at levels 4 and 5 activities.

The paper first locates the depth of technological capabilities achieved by firms by taxonomies and trajectories, as well as, their contribution to growth in exports, employment and wages, and skills intensity level achieved in 2006. The paper then assesses the individual experiences to identify the drivers of technological upgrading and economic synergies. Because each experience is assumed to be different the study used an inductive approach with a ledger of critical generic questions directed at industry associations in the locations. Consistent with Keynes (1890) notion of inductive theorizing, evidence was compiled along a generic typology by taxonomy and trajectory.

The data on the technological depth of operations, and exports, employment, wages and skills was collected by national consultants under our supervision. The firms were selected using a stratified sampling procedure on the basis of ownership and size of firms in the clusters.⁷ Location was not used as one of the criterion because the firms are from a particular location. The sample size of firms in Buenos Aires and Los Lagos was 50 each, and 100 in Qiaotou and 103 in Penang. The response rate was close to 100% in all the locations, which is largely because of the use of national consultants who were chosen because of their tacit links with the firms. The numbers in Latin America were smaller compared to those in Asia because of the smaller population of related firms in the former over the latter.

Whereas the evaluation of institutional support facing firms is based on the typology presented in Table 2, the technological capabilities of firms were rated using the typology shown in Table 3. The scoring in Table 2 was based on consensus reached from brainstorming sessions held with industry association officials in 2007-08, while individual firms' responses were tallied for Table 3. The sessions for the scoring reported in Table 2 were conducted by national consultants in Buenos Aires (automotive cluster), Santiago (salmon cluster), Penang (electronics cluster), and Qiaotou (button cluster) in 2007-08.

⁷ The sampling was undertaken by national consultants: Penang (Rajah Rasiah), Buenos Aires (Gonzalo Bernat), Los Lagos (Lydia Vidal) and Qiaotou (Xin Xin Kong).

Table 2: Systemic pillars and technological upgrading

	Basic infrastructure	High tech institutions	Network cohesion	Integration in international markets
Initial Conditions (1)	Political stability and efficient basic infrastructure	Presence of some economic agents seeking knowledge inputs	Social bonds driven by the spirit to compete and achieve	Connecting to the international economy
Learning Phase (2)	Strengthening of basic infrastructure with better customs and bureaucratic coordination	Import, learning by doing and duplicative imitation. Human capital development	Expansion of tacitly occurring social institutions to formal intermediary organizations to stimulate connections and coordination between economic agents	Access to foreign knowledge through machinery and equipment import and FDI Integration in global value chains
Catch Up Phase (3)	Basic infrastructure capable of providing strong essential services.	Import, creative duplication and innovation. Beginnings of Mark I system of learning	Smooth links with meso organizations connecting and coordinating to solve collective action problems	Access to foreign knowledge through licensing, acquisition and imitation, through imports and exports. Upgrading in global value chains, IPR regulation starts here.
Advanced Phase (4)	Advanced basic infrastructure instruments	Developmental research to support creative destruction (Schumpeterian Mark 1).	Participation of intermediary and government organizations in coordinating technology inflows, initiation of commercially viable R&D	Access to R&D human capital and collaboration with R&D institutions, high tech resources and markets abroad
Frontier Phase (5)	Novel Basic infrastructure to support new developments in basic infrastructure	Basic research to generate new knowledge paths (Schumpeterian Mark II system).	Participation of intermediary organizations in two-way flow of knowledge between producers and users	Global pursuit to connect with frontier nodes of knowledge

Source: Authors

Table 3: Firm-level technology by taxonomies and trajectories

Knowledge depth	HR	Process	Product
Simple activities (1)	On the job and in-house training	Dated machinery with simple inventory control techniques	Processing of raw materials, and assembly of components, CKDs and CBUs using foreign technology
Minor improvements (2)	In-house training and performance rewards	Advanced machinery, layouts and problem solving	Use of precision engineering and minor adaptations to products
Major improvements (3)	Extensive focus on training and retraining; staff with training responsibility	Cutting edge inventory control techniques, SPC, TQM, TPM	Cutting edge quality control systems (QCC and TQC) with original equipment manufacturing (OEM) capability
Engineering (4)	Hiring engineers for adaptation activities; Separate training department	Process adaptation: layouts, equipment and techniques	Large scale product adaptation (including improvement of fish species)
Early R&D (5)	Hiring engineers for product development activities; Separate specialized training activities	Process development: layouts, machinery and equipment, materials and processes	Product development capability. Some firms take on original design manufacturing (ODM) capability. Strong emphasis on ergonomics, sanitary conditions of fish and consumer safety features of products
Mature R&D (6)	Hiring specialized R&D scientists and engineers wholly engaged in new product research	Process R&D to devise new layouts, machinery and equipment prototypes, materials and processes	Novel product development capability, with some taking on original brand manufacturing (OBM) capability. Among lead firms putting up new automotive, electronics and button models, and improvements in fish standards

Source: Authors

4. Institutional upgrading

This section uses information gathered from the brainstorming sessions with industry associations and secondary documents to locate the level of institutional development of the four clusters. The assessment of the systemic pillars is based on the results shown in Table 4.

The basic infrastructure in all four clusters were equally good with good roads, flight connections, power supply, water supply, primary and secondary schools and internet support qualifying them all at stage 4 level. We did not find the globe's most novel planning (level 5) initiatives targeted at technological upgrading in firms or improvements to the environment in the four clusters. Nevertheless, government planning showed considerable efforts to support technological upgrading and the strengthening of supporting infrastructure.

Industry association officials reported that the high tech infrastructure facing automotive firms in Buenos Aires, salmon producers in Los Lagos and electronics firms in Penang have reached level 3 upgrading phase.⁸ Whilst these locations have training, standards and testing organizations, they do not have fully equipped R&D labs and designing centres to support firms apart from a handful of universities engaged in specific developmental activities. FC and CORFO only support adaptive development activities with their R&D labs not equipped to support new knowledge creation activities. Hence, when the Infectious Salmon Anemia (ISA) virus struck in 2007, FC and CORFO could not respond (Iizuka and Katz, 2011).

⁸ The Malaysian government approved grants and import of human capital to foreign firms in 2005 to attract wafer fabrication and designing activities.

Button firms in Qiaotou enjoy the most sophisticated support from the designing, training and testing centres and R&D labs and universities that assist with the development of new materials and button designs. However, most button firms in Qiaotou have only advanced to level 4 technology-intensive activities as the designing, testing and developmental aspects of button technology is undertaken in the meso organizations and universities.

The intensity of interactions between the firms, and between firms and meso organizations, and those governing the regulatory instruments in Buenos Aires, Los Lagos and Penang have reached level three. Yongjia County officials in Zhejiang Province, China, have worked hard to strengthen connections and coordination between the button firms, the meso organizations, and the incentive providers to reach level 4 technological activities.

All four clusters are highly integrated in global markets, especially when it comes to exports, and access to foreign knowledge. Button firms in Qiaotou originally learnt from their Italian masters to eventually move to the technology frontier by seeking support from designing centres, R&D labs and universities in China. The TNCs have introduced best practices – including designing activities – in Buenos Aires and Penang. CORFO and FC provide adaptation support utilizing foreign knowledge from Norway, Scotland, United States and Canada for salmon production in Los Lagos.

Because high technology meso-organizations have reached only level 3, the human capital and knowledge generated domestically in Penang is not adequate to support firms' extensive participation in designing and R&D activities. The inability of government programmes to attract on a large scale human capital inflows from abroad has restricted the capacity of Penang to support further upgrading, though firms such as Intel, Motorola, Advanced Micro Devices, Fairchild and Alterra have captured the limited supply of engineers to engage in designing activities.

Qiaotou has reached the technology frontier of button manufacturing as the masters in Italy from whom they learnt button production over the period 1976-1990 now attend exhibitions in Zhejiang Province to monitor the introduction of new designs and materials. The level 4 sophistication of R&D labs and universities has facilitated this development. However, we classified the cluster as enjoying level 4 institutional support rather than level 5 because the technologies realized in materials used in button manufacturing are a result of the coevolution of existing technology in metal, chemicals, plastics, and testing and manufacturing processes undertaken in Gansu and Guangdong states.⁹

All four clusters enjoy strong integration (level 4) in global markets. However, whereas the automotive and electronics clusters of Buenos Aires and Penang are extensions of TNC production chains, the salmon and button clusters of Los Lagos and Qiaotou enjoy strong national ownership.

Table 4: Embedding institutional pillars, selected locations, 2006

	Basic infrastructure	High tech organizations	Network cohesion	Integration in international markets
Catch Up Phase (3)		Buenos Aires Los Lagos Penang	Buenos Aires Los Lagos Penang	Buenos Aires Penang
Advanced Phase (4)	Buenos Aires Los Lagos Penang Qiaotou	Qiaotou	Qiaotou	Los Lagos Qiaotou

Source: UNIDO, *Survey of Dynamic Industrial Locations, 2007-08*

⁹ This decision was validated by industry officials in Qiaotou when we interviewed them on 24 December 2007. We are grateful to Kong Xin Xin for assisting us with the interview.

5. Technological upgrading and economic synergies

This section analyses the level of technological upgrading, and growth in exports, employment, wages, and skills intensity in the four clusters. Over 50 per cent of firms in the clusters were at least engaged in level 3 knowledge-intensive activities. The upgrading achieved by these clusters has helped to sustain exports, employment creation, wage increase and high intensity of skills.

Technological upgrading

All four clusters are technologically advanced. With the exception of one TNC in level 6 activities and significant numbers of TNCs in level 5 activities in Penang, the percentage shares of firms in the four clusters are mainly concentrated in level 4 technological capabilities (see Table 5).

Interviews show that the automotive cluster in Buenos Aires faced a dualistic evolution of supplier capabilities as TNC assemblers and their suppliers continued to enjoy strong capabilities, while independent suppliers targeting the second hand market and exports were characterised by inferior capabilities. Whereas suppliers to TNCs were engaged in tier 1 activities to provide completely knocked down modules, tier 2 firms supplied components to these firms, and the TNCs. Whereas TNC assemblers were reported to operate at level 5 capability and their suppliers were reported to function at level 4 capability, independent national suppliers were reported to operate at level 3 capability. Modularization and changes in the regulatory environment created by the Automotive Industry Decree and the *Mercado Común del Sur* (MERCOSUR) agreement that pressed for the introduction of new models drove the relocation of designing of small vehicles by TNCs, thereby stimulating the development of human resource and process technology capabilities in Buenos Aires (Bernat, 2008). Foreign TNCs are the only car assemblers in the automotive cluster in Buenos Aires. Despite Brazil providing a much larger market, TNCs have retained Buenos Aires and Sao Paulo as major regional production locations.

As in Norway, salmon farming has become horizontally integrated since the 1990s as price competition intensified (see Aslesen, 2007). However, whereas vertical coordination through a blend of competition and cooperation has intensified in Norway, vertical and horizontal integration has increased in Chile, especially among a firms seeking to internalize development work to raise quality and yield. Millions of Atlantic salmon eggs are imported every year from Norway, Ireland and Scotland, though the figure fell as national producers of salmon eggs worked with universities, such as, the University of Chile to reduce dependence on foreign suppliers (Katz, 2006: 203). The hatcheries handle hatching and infant growth of the salmon smolts in vats placed in fresh water lakes. Highly skilled employees ensure that the water and temperatures remain ideal through the infantile period of growth of the salmon. Once the salmon grows to adult size it is then transported to the sea by trucks, boats or helicopters where it is farmed using rectangular net cages until it is harvested. Apart from diseases, sea lions often poach a number of the salmon before they are harvested. Skilful divers are often required to stitch up nets damaged by sea lions. Salmon are fed in the dark. Once harvested, the salmon is then sent to factories for processing, which is undertaken through automated conveyor belts where skilled workers handle the different aspects of cleaning, slicing, finishing and packing. While most of the processed fish are frozen some are smoked for export in the same factories. Throughout these activities firms have absorbed best practices from the United States, Canada, Scotland and Norway. Some of the larger firms have also installed automated machinery and equipment to carry out research to improve water and salinity control techniques, feeding behaviour of salmon and processing techniques to reduce throughput time.

Although Qiaotou was utilizing the newest materials, machinery and equipment, chemicals and metal mixes, we did not classify them in level 6 because its inputs were supplied by R&D labs located outside the clusters studied. Whereas the firms studied were directly involved in the manufacture of buttons significant knowledge inputs attracting new designs and materials comes from labs located in the University of Lanzhou and Huanan University of Technology. Hence, the firms studied were not highly knowledge-intensive as their activities were confined to manufacturing only. Nevertheless, improvements in layouts,

automation and the implementation of best practices were becoming important in these firms.

Motorola was the only firm reporting that their plant in Penang was the most sophisticated in their microcosm and were engaged in new product development. Whereas Intel, Alterra, AMD, Fairchild and Agilent also reported involvement in extensive designing work they reported having their most sophisticated operations abroad. Over 50 per cent of firms in all four clusters had evolved to at least level 4 technological capabilities - across human resource, process technology and product technology. The electronics cluster in Penang enjoys a high agglomeration of firms engaged in component manufacture and the assembly of telecommunications and industrial electronics products. The assembly of consumer electronics and disk drives became important in the 1980s and 1990s but were relocated to China, Philippines, Indonesia and Vietnam as labour shortages became a problem.

Of the four clusters, the incidence of firms having level 4 knowledge intensity capabilities was highest among electronics firms in Penang. Overall, 83.7 per cent of the sampled electronics firms in Penang were either fully or majority foreign owned in 2008. Foreign firms dominated participation in levels 4 and 5 activities in Penang. Most national firms were engaged in level 3 knowledge-intensive activities. Vitrox is among the few national firms engaged in level 4 activities. A number of foreign TNCs (e.g. Intel, Motorola, Agilent, Advanced Micro Devices, Fairchild and Alterra) were engaged in level 5 designing activities that generated patents in the United States.

Whereas the salmon and button clusters of Los Lagos and Qiaotou were fairly integrated with firms engaged both in the value chain and in complementary activities, the automotive and electronics clusters of Buenos Aires and Penang, have often faced wide swings in production focus. In addition, firms in Buenos Aires, Los Lagos and Penang have had to also contend with macro-volatilities generated by wide swings in prices and exchange rates. The majority of firms in the salmon and button clusters enjoyed designing and adaptive engineering support from national meso-organizations located within proximate distance. Button firms in Qiatou in addition also enjoyed R&D support in the development of new designs and material mixes from universities and R&D labs located in other provinces in China. The level 5 designing activities of automotive firms in Buenos Aires and electronics firms in Penang take place in-house in foreign TNCs. Although some collaboration was reported in 2008 between salmon firms in Los Lagos (Vidal, 2008), automotive firms in Buenos Aires (Bernat, 2008) and electronics firms in Penang (Rasiah, 2011) with universities they were strong and related to their prime activity only between a handful of firms and particular universities, but in most cases were *ad hoc* and did not relate to the core aspects of their operations.

Table 5: Technological intensity of firms in clusters, 2006

a) Latin America

Buenos Aires: Automotive cluster

	HR	PT	RD
1	50(100)	50(100)	50(100)
2	50(100)	50(100)	50(100)
3	50(100)	50(100)	50(100)
4	7(14)	7(14)	7(14)
5	2(4)	2(4)	1(1)
6	0(0)	0(0)	0(0)

Los Lagos: Salmon cluster

	HR	PT	RD
1	50(100)	50(100)	50(100)
2	50(100)	50(100)	50(100)
3	50(100)	50(100)	50(100)
4	12(24)	12(24)	5(10)
5	2(4)	3(6)	0(0)
6	0(0)	0(0)	0(0)

b) Asia

Penang: Electronics cluster

	HR	PT	RD
1	103(100)	103(100)	103(100)
2	103(100)	103(100)	73(71)
3	85(83)	85(83)	66(64)
4	69(67)	69(67)	57(55)
5	35(34)	35(34)	11(11)
6	1(1)	1(1)	1(1)

Qiaotou: Button cluster

	HR	PT	RD
1	100(100)	100(100)	100(100)
2	100(100)	100(100)	100(100)
3	75(75)	75(75)	72(72)
4	39(29)	37(37)	33(33)
5	3(3)	3(3)	2(2)
6	0(0)	0(0)	0(0)

Note: Figures in parentheses refer to percentage of total firms.

Source: UNIDO, *Survey of Dynamic Industrial Locations*, 2007-08

Economic synergies

Technological upgrading helped stimulate growth in exports, employment and wages (see Figure 1), and improvement in skills (see Figure 2) in the four clusters. Only firms in the button cluster showed low skills intensity, which is a consequence of the knowledge-based activities confined to external universities and labs.

Exports drove growth in clusters, while the differentiation and division of labour arising from technological upgrading drove export growth, with the latter providing circular push to support the former. Export growth continued to be strong in all four clusters over the period 2000-2006 (see Figure 1).

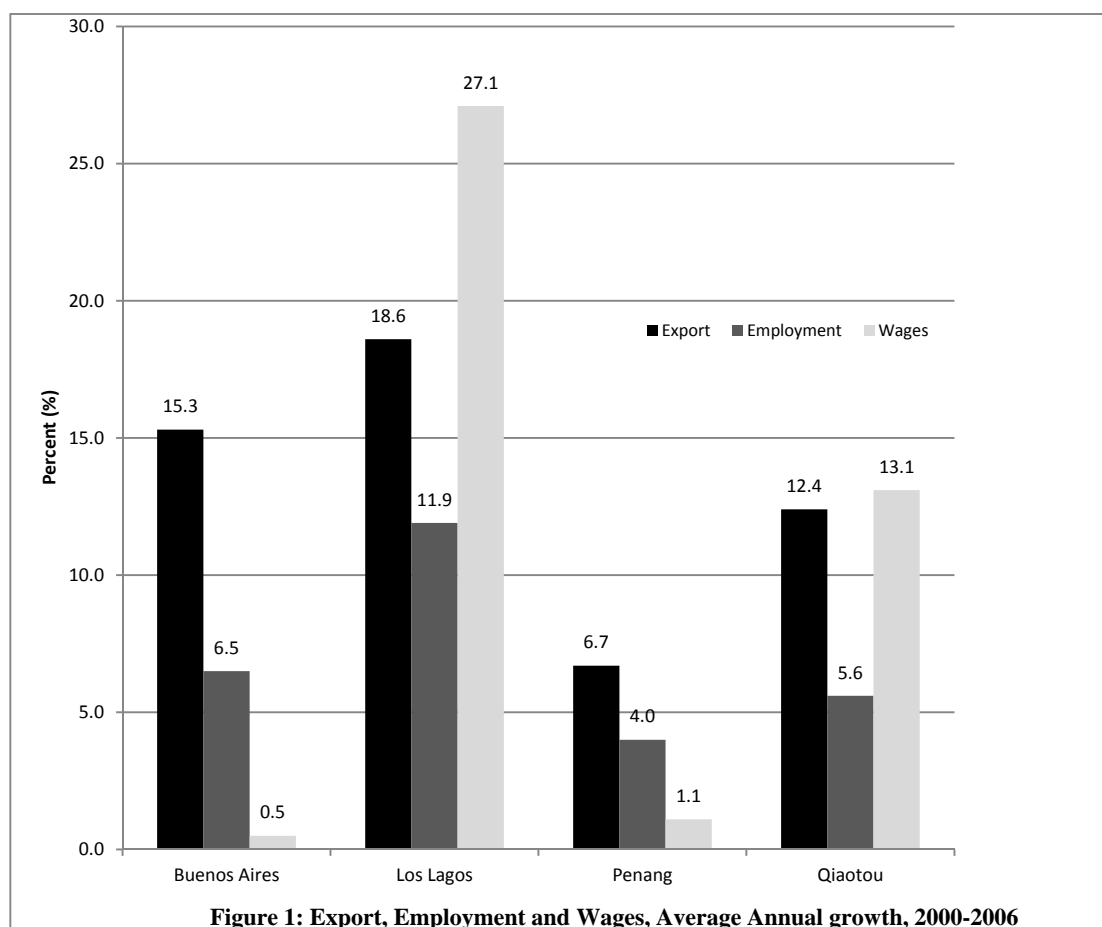
The salmon cluster of Los Lagos (18.6%) and automotive cluster of Buenos Aires (15.3%) have recorded the strongest export growth over the period 2000-2006. Whereas, institutional arrangements that required exports to compensate for imports within the MERCOSUR agreement were important in the growth of regional automotive exports from Buenos Aires, the expansion in supermarket coordinated demand has been instrumental in salmon exports from Los Lagos to global markets.

The Qiaotou (12.4%) and Penang (6.7%) clusters also enjoyed strong export growth demonstrating that foreign demand has been the spearhead of cluster development in these locations. Although a saturation in human capital supply has slowed down new firm relocation, it has not stopped existing firms from continuing to expand exports from Penang (see Chandran, 2008; Rasiah, 2011). Similarly, despite selling the bulk of buttons to export-oriented clothing firms in China, exports have continued to grow strongly in Qiaotou (Kong, 2008).

The agglomeration of export-oriented activities has driven strong differentiation and division of labour so that employment has recorded strong growth in all four clusters (see Figure 1). Employment in Los Lagos grew at an annual average rate of 11.9 per cent over the period 2000-2006. Employment in the automotive cluster of Buenos Aires and button cluster of Qiaotou grew at 6.5 and 5.6 per cent per annum respectively, and the electronics cluster of Penang grew at 4.0 per cent per annum in the period 2000-2006.

All the locations examined enjoyed positive real wage growth over the period 2000-2006. Los Lagos and Qiaotou in the samples enjoyed the highest annual average wage increments of 27.1 and 13.1 per cent, respectively over the period 2000-2006 (see Figure 1). Rapid technological upgrading helped these regions capture significant shares of exports in global markets for salmon and buttons produced.

Average annual real wages grew slower in Penang (1.1%) and Buenos Aires (0.5%) clusters over the period 2000-2006. Whereas high inflation slowed down real wage growth in Buenos Aires, the lack of technological capability building and an influx of foreign unskilled labour restricted the capacity of Penang to support rapid real wage growth. The liberalization of import restrictions reduced the automotive component manufacturing of national firms to low value added activities thereby keeping real wage growth low in Buenos Aires (Bernat, 2008).



Source: UNIDO, *Survey of Dynamic Industrial Locations, 2007-08*

While a low skills base has often been the point of integration of particular locations in global value chains, sustained technological upgrading and economic synergies have often driven improvements in skill-intensities.¹⁰ The data collected allowed the inclusion of skilled direct workers into the skills category, though skills intensities vary with industry.¹¹

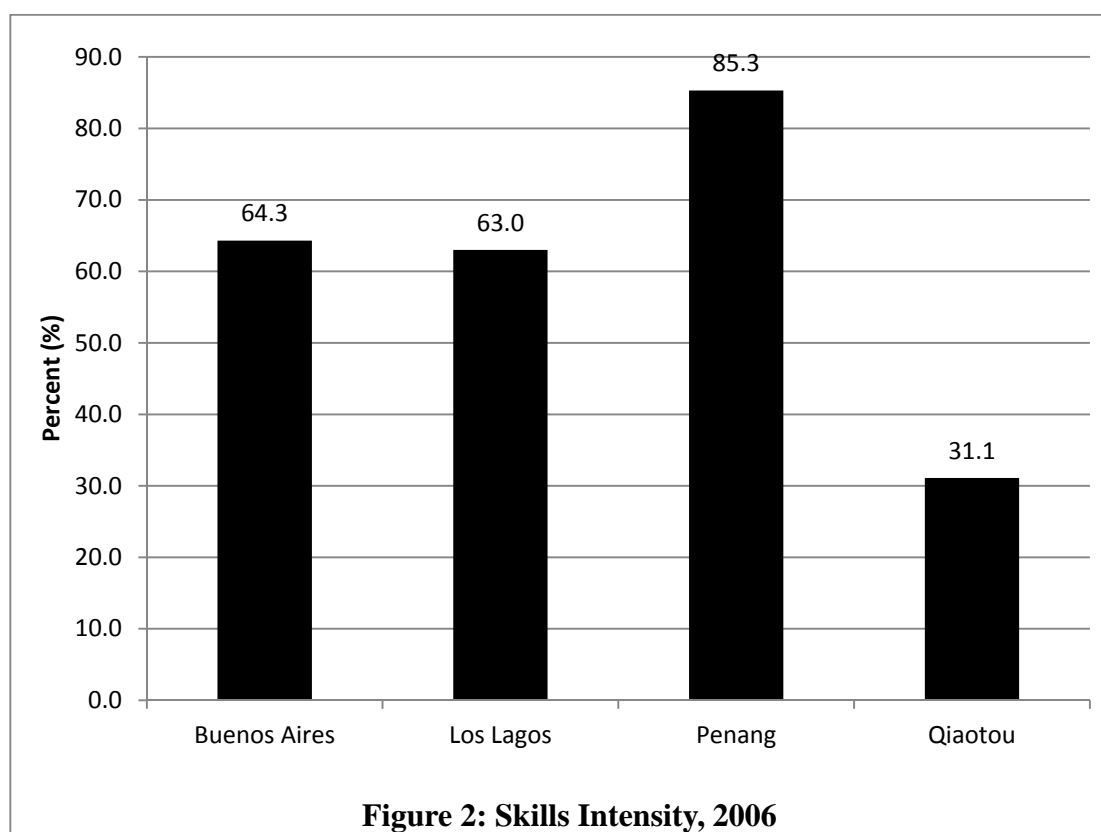
Skills intensity percentages of the automotive cluster in Buenos Aires and the salmon cluster in Los Lagos were 64.3 and 63.0 per cent respectively in 2006 (see Figure 2). Whereas in-house and external training were important in the development of skills of workers among salmon producers in Los Lagos, the industry association officials considered TNCs as the prime source of skills training among automotive firms in Buenos Aires.

Skill-intensity was highest in Penang's electronics cluster at 87.7 per cent. The long period of production experience and rising pressure for upgrading since the introduction of flexible production techniques from the 1980s has been pivotal in raising skills levels in Penang. Especially integrated circuits (IC) assembly and test activities, which forms the nucleus of the electronics cluster in Penang, became knowledge intensive since the 1980s (see Rasiah, 1988).

Skill intensity levels were lowest in Qiaotou (31.1 per cent) whose button firms are limited to just manufacturing. Button firms in Qiaotou specialize in manufacturing in a division of labour where the highest value added activities of testing, designing and R&D take place in meso organizations either in Zhejiang or other provinces in China.

¹⁰ Skill intensity was calculated as a percentage of professional, managerial, supervisory and clerical personnel, and skilled workers among direct workers as a share of the workforce.

¹¹ Skill-intensity is typically estimated by dividing the sum of managers, professionals, and technical, clerical and supervisory personnel by the total workforce. We included skilled direct workers in the numerator.



Source: Computed from UNIDO, *Survey of Dynamic Industrial Locations*, 2007-08

6. Drivers of upgrading

All four clusters have generated knowledge to support substantial technological upgrading and economic synergies. Whereas foreign sources of knowledge still drive innovation in the salmon cluster of Los Lagos, and TNCs in the automotive cluster of Buenos Aires and the electronics cluster in Penang, national meso organizations have assumed that role in the button cluster of Qiaotou. We discuss in this section the drivers of technological upgrading and economic synergies in the four clusters.

Automotive cluster of Buenos Aires

While the automotive industry in Buenos Aires was started earlier, it began to expand strongly from the 1950s under an import-substitution (IS) regime that was characterized by several agreements between the Argentinian government and TNCs, which became the basis of the relocation of subsidiaries to supply the domestic market. The production of motor vehicles in Argentina grew rapidly from 100,000 units in 1961 to 200,000 units in 1968 and 294,000 units in 1973 (Bernat, 2008: 3). Employment in the industry grew from 10,000 in the late 1950s to 57,000 in the mid-1970s.

After a long period of contraction from the mid-1970s till the end of the 1980s arising from falling demand domestically as the macroeconomic environment took a hit from economic mismanagement, three important developments drove the rejuvenation of the industry. First, the Reconstruction of the Automotive Industry Decree of 1991 underlined benefits and obligations of automobiles manufacturers operating in Argentina, which allowed producers to import parts and finished vehicles tax free when compensated through exports, and if they met investment criteria associated with new technology and the launching of new models. The Decree also restricted imports of automotive parts, which boosted domestic producers.

Second, it was established within the MERCOSUR agreement between Argentina, Brazil, Paraguay and Uruguay in 1995 for a commercial compensation exchange involving automotive goods between members of the union so that imports of such items from one member country will be compensated by reciprocal exports of related items to that country, which drove the leading global producers to retain operations in both Argentina and Brazil (Bernat, 2008: 4). Third, the Convertibility Regime that was introduced in the early 1990s encouraged increases in domestic incomes, which helped raise demand for cars in the country, though the 2001 crisis also hit hard the firms owing to the collapse of the convertibility mechanism.

Whereas the contracting plants such as Mercedes began to expand operations again from 1991, new TNCs such as Volkswagen, General Motors, Peugeot, Fiat, Ford, Renault and Toyota relocated operations from the mid-1990s as a consequence of the new policy developments. Automobile production in Argentina subsequently grew from 262,000 units in 1992 to 446,000 units in 1997 (Bernat, 2008: 4). Production of automobiles in the Buenos Aires cluster rose from 188,000 units in 1992 to 232,000 units in 1997. Employment in the cluster increased from 18,000 in 1991 to 26,000 in 1997. While the MERCOSUR proved superior to the IS regime of the past, exchange rate fluctuations often created considerable problems, which weakened the industry over the period 2001-2003. Production of automobiles in the Buenos Aires cluster fell to 90,000 units in 2003 leading to the Argentinian government suspending the MERCOSUR agreement, which led Katz to call for effective macro-micro coordination to insulate Argentinian firms from external shocks.¹² The situation reversed from 2004 as an appreciating Brazilian Real and a depreciation in the Argentinian peso in 2003 made exports attractive. The changed environment also attracted the relocation of Honda into the Buenos Aires cluster.

The TNCs have been the main source of knowledge flows to the Buenos Aires automotive cluster. While the TNC assemblers imposed stringent standards they also played a critical role by transferring technology to the national suppliers through quality assurance certifications, and quality and inventory control systems. The suppliers that enjoyed strong relationship with the TNCs had in possession ISO 14001 and quality standard (QS) certifications. Through networking that shows a blend of competition and cooperation, the TNCs assisted in the acquisition of state of the art precision machinery and equipment by the suppliers. The TNCs often encouraged their domestic suppliers to share knowledge in developing their engineering capabilities so as to participate effectively in generating the innovations essential to perform original equipment manufacturing and original design manufacturing operations. The strong training regimes essential to stay competitive in the automotive industry practiced by the TNCs was extended to the first-tier suppliers who also undertook profound training and retraining of staff. Several TNCs also regularly trained personnel from suppliers whenever a new model was to be launched (Bernat, 2008).

TNCs have been the major drivers of knowledge flows into both the introduction of new models, assembly of vehicles and component production (see Bernat, 2008). Indeed, a distinct division of labour has emerged between automotive suppliers tied to TNCs and the independent automotive component manufacturers. While the Reconstruction of the Automotive Industry Decree and the MERCOSUR agreement have played important roles in TNCs deploying high technology operations in Argentina, the global restructuring of the industry also helped, which has driven TNCs to introduce best practices in all major locations. However, while new model designs have emerged in the main assembly locations, none of their operations are sophisticated enough to undertake level 6 technological depth operations. Also, contrary to the believe that a freer world will lead to the concentration of production among the most efficient (on the basis of good infrastructure, market size and best incentives) locations, automotive production appears to be dispersed over locations with reasonably large markets because of one, capabilities that have been evolved from past policies, and domestic demand.

Salmon cluster of Los Lagos

Although Union Carbide's subsidiary, Domsea Farms Chile, began farming salmon in 1974 (Katz, 2006: 195), the mass development of the industry in Southern Chile started as part of the government's initiatives from the post-Allende era to alleviate poverty. The Public Works Corporation (CORFO) was very much

¹² Communication with Jorge Katz whose argument is consistent with Keynes' (1936) and Stiglitz's (2002) accounts of the disruptive consequences of leaving financial instruments to market forces.

instrumental in the growth of the next salmon farming firm in 1974, Lago Llanquihue, which exported the fish to France since 1978. Following severe problems that derailed Domsea's operations, FC bought its facilities in 1981 to create Salmenes Antartica, which became the first firm to produce over 1 thousand tons of the fish in 1998 (Maggi, 2006: 113). The geo-political support the industry received from the United States, Canada, Scotland and Norway to transfer salmon farming technology to Chile is also important (Chile, 2003).

The transformation of Salmon from ornamental fish to a commercial food-based export good started with the import of salmon eggs, vats, breeding technology and other inputs from United States, Canada, Scotland and Norway. Salmon was identified among a number of resource-based sectors by FC that included timber and wine for development through sector specific policies at a time when Chile under the Pinochet regime had abandoned tariff-based import-substitution policies. The Chilean government through the Institute of Fishing Industries and the Fishing and Hunting Division of the Department of Agriculture, and CORFO and FC, intervened extensively to promote the industry in the poorer regions of Southern Chile. Interviews with officials of FC show that Chile's participation in salmon farming to become the second largest exporter in the world after Norway owes much to public-private initiatives that brought together government agencies, financial institutions, and development and experimentation laboratories.¹³ CORFO and FC carried out several experiments to adapt technology imported from the United States, Canada, Scotland and Norway. Foreign agencies and firms that relocated support through Nichiro Chile, the Japan International Cooperation Agency, and the University of Washington were among the early pioneers in the transfer of knowhow to Chile (Katz, 2006: 194). Nichiro Chile is a Japanese firm that started to cultivate salmon farming in Chile to supply the Japanese market. The FC also incubated firms that were sold to private owners once they matured.¹⁴

Financial regulation and deregulation was not only aimed at financing enterprises by government authorities of Chile but also at insulating enterprises in the country from external shocks (Bank of Chile, 2004). The impact of the debt crisis of the early 1980s led to the enactment of a new banking law in 1986 to check moral hazard and systemic risks, and regulations' governing financial institutions that were amended to strengthen the financial sector (Bank of Chile, 2004: 11-12). The government introduced capital controls to insulate the economy from volatile financial fluctuations, which included a 3-year stay on foreign direct investment.¹⁵ Salmon firms were some of the firms that benefited from the capital control measures that shielded them from disrupting technological upgrading.

The United States' Food and Drug Administration, and the Food and Agriculture Organization (FAO) assisted in the improvement of quality standards in the 1980s. A blend of competition and cooperation helped the firms amortise investments by pooling and undertaking bulk purchases of inputs from abroad. CORFO and FC played important roles in building the bonds of trust between the firms and meso organizations. The major breakthrough into global expansion started when supermarket chains took control of salmon value chains. Chilean Salmon enjoyed this connectivity from the 1990s when large retailers got into the monitoring of production and into the marketing of the fish. This development provided the impetus for large scale production as markets became large and more secure. Also, the large supermarkets helped connect the salmon producers with the major standards organizations regulating sanitation standards in Europe.

While some large private farms absorbed and adapted from abroad some of the best practices, the FC and CORFO played a wide role on a national scale to attract best practices from Canada, Scotland, United States and Norway, which were adapted and introduced in especially the smaller salmon farms in Chile. In fact, interviews with farm officials, and officials from FC and CORFO showed that much of the adaptation of fish cultivation, virus treatment and feed meal experimentation were started in the public-private organization of FC before it was absorbed by the salmon farmers and processors. The prime focus of FC has been on experimentation and adaptation without deep involvement in R&D. Having strong links with

¹³ We had the benefit of talking to officials from firms, FC and Corfo to make the point.

¹⁴ FC played the role of supporting the private sector, selling its successful incubators whenever there were buyers (Chile, 2003).

¹⁵ The three-year stay was reduced to one year in 1995 (Bank of Chile, 2004: 12).

the frontier salmon farming systems in Canada, United States, Scotland and Norway helped as best practices were fairly easily imported, adapted and introduced in Chile.¹⁶ The best practices included the size of breeding vats, feeding materials, frequency of feeding, sterilizations before use of the vats and the organization of process flow in both the vats, and when it matures in the sea, and processing factories.

As competition increased in the second half of the 1990s with sharply falling prices, the firms began to merge and expand vertically to expand participation into hatchery, cultivation, and processing, and in some cases also in the production of feed meal, fish eggs, and several other inputs. Demand for high quality, quick delivery, and transparency in the production process against a fall in profit margins meant that fewer but highly integrated large firms began to dominate the export market. Some family firms became suppliers to these large firms.

Salmon production has since the 1990s also become highly concentrated worldwide (Iizuka and Katz, 2006). The average firm became more capital- and technology-intensive with skill intensities growing strongly as workers were exposed to higher daily targets, automation and technical tasks. The expansion in size is also a consequence of a lack of growth in independent knowledge suppliers to support rapid productivity growth. In fact, salmon firms in Chile, as well as, FC and CORFO, still rely considerably on R&D undertaken in Scotland, Canada, United States and Norway.

Overall, the origin of a commercially viable salmon farming and processing industry in Los Lagos started very much with the role of government to absorb its farming from the typical salmon farming countries of the United States, Canada, Scotland and Norway. The government, through FC and CORFO played major roles to experiment, adapt and introduce farming methods until they became successful in the 1980s by when they had begun to sell these ventures to private owners, which differs considerably from Perez's (2005) who argues that the industry association was instrumental in the development of salmon farming in Chile. Our view was also shared by the senior researchers interviewed from the government hatchery labs in Puerto Mont who explained that the government's role was central here, though, their objectives was always to support private firms. The early acquisitions were by Chilean firms, including family firms with national firms dominating salmon farming in the 1980s. The government also helped with improvements to the infrastructure in the farming provinces. Hence, the prime drivers were the Chilean government and the foreign providers of the knowledge with the mobilizational role of the United States – who also enticed Canada, Scotland and Norway to support the exercise.

As competition from abroad and domestically intensified, farming lakes and shores became congested, the industry began to face vertical integration as the small and medium firms were not able to compete as prices crashed, and to meet the quality standards demanded by export markets. It is since the late 1990s that foreign firms began to acquire brownfield firms to participate in increasing numbers in salmon production in Chile. Nevertheless, the drivers of technological change from the late 1990s still included FC and CORFO who continued to adapt and adopt best practices developed in the R&D labs in Norway, Scotland, United States and Canada, and the supermarket chains that became the main wholesale link for producers in Los Lagos. While such strategies facilitated the upgrading of salmon farming activities to levels 3 and 4 technological capabilities, the lack of strong participation in R&D activities domestically left Chilean salmon farming vulnerable to the vicissitudes of new virus outbreaks. Hence when ISA struck in 2008 the Chilean innovation system lacked the capability to generate the new knowledge essential to cure the disease at the expense of a huge contraction in production as output fell by almost 60 per cent from its peak of 700 thousand tons in 2006 with close to 20,000 job losses (Iizuka and Katz, 2011: 269).

Electronics cluster of Penang

The electronics cluster began in Penang following the relocation of Clarion and National Semiconductor in 1971. The government earmarked the Bayan Lepas Free Trade Zone to attract labour-intensive export-oriented manufacturing activities through the provision of tax holidays and guaranteed the repatriation of profits, tariff-free imports and exports, leasing of subsidized land and security (see Rasiah, 1988). Interventions to offer the additional relative benefits of decomposing and relocating labour-intensive

¹⁶ See Pietrobelli (1998) for an early account of this experience.

segments from more secure developed sites to less secure and developed sites, such as Penang, were important.¹⁷ Since the domestic market was small, the relocation of electronics assembly to Penang in the 1970s was meant completely for export.

Over the period 1972-80, the government provided tax holidays and good basic infrastructure in the export processing zones. A change in government policy over the period 1981-85 that saw an ending of tax holidays began to discourage new investments. However, the economic crisis of 1985-86 drove the government to renew the tax holidays and devalue the Ringgit to attract new TNCs into the country (see Malaysia, 1986). The government used the macro-micro coordination initiative again when capital controls were introduced in 1998 to shield the country from currency attacks (Rasiah, 2000). The government also allowed exporting firms to handle transactions in US dollars. A strong US economy helped boost electronics exports during the 1997-98 Asian financial crisis. The government subsequently introduced fiscal stimulus in 2009 following a massive contraction in exports during the global financial crisis of 2008-09 and assisted with the repatriation of foreign workers back to their home countries as a recession gripped the electronics cluster (Malaysia, 2010).

Government incentives and grants have played an important role to stimulate upgrading. Incentives were given since 1988 to stimulate upgrading, which was augmented by further efforts to promote technological upgrading following the implementation of the Action Plan for Industrial Technology Development (APITD) in 1991. The APITD gave rise to important meso organizations to drive training, testing, R&D (including the provision of grants), government-funded venture capital and high tech infrastructure (see Malaysia, 1996). The late 1980s and 1990s saw the promotion of clustering in Penang when consumer and industrial electronics firms were attracted to use components produced in the Island. However, the relocation of consumer and industrial electronics firms did not raise the technological intensity of the cluster as the forward integration involved less knowledge-intensive activities than integrated circuits production. The government's Second Industrial Master Plan of 1996 included Penang's clustering as one of its main promotional targets (Malaysia, 1996). However, the lack of human capital, and a mechanism to vet and appraise investment and production restricted its capacity to drive upgrading.

The government also strengthened basic infrastructure and security in the export processing zones, including the construction of an excellent airport to fly in wafers and other components and fly out assembled products from Penang, which was strengthened further since the 1980s with the expansion of highways, broadband internet service and bridges to facilitate the employment of employees across the Penang Straits. As the industry became increasingly knowledge-intensive, the Penang Skill's Development Centre (PSDC) was started in 1989 with the government imposing only a nominal rent on its building, while the TNCs donated computers and other machinery and equipment. The PSDC subsequently upgraded its services to include designing through federal grants. The Penang Development Corporation played a critical role in coordinating the demands of the TNCs with the regulatory institutions of the federal government. The entrepreneurial Chinese community took advantage of growing demand for proximate sourcing of precision machine tools and plastics products to evolve a sizable group of suppliers over the 1980s and 1990s. A number of engineers who had gained experiential knowledge working in the TNCs left in this period to start machine tool, plastic injection moulding and electronics firms (see Rasiah, 1995).

The skill-intensity of electronics firms in Penang rose strongly with increasing demand. However, the lack of human capital to participate extensively in designing and R&D restricted the capacity of Penang to upgrade extensively into level 5 technological capabilities *a la* Korea and Taiwan. Electronics firms also lacked access to R&D from meso organizations in Penang. Hence, the handful of TNCs that managed to obtain R&D grants from the government internalized their designing activities without much links to domestic R&D labs and national universities. Only Motorola reported having its most sophisticated plant in Penang. Also, all firms engaged in levels 5 and the one firm engaged in level 6 knowledge-intensive activities are foreign TNCs. National firms are only engaged in levels 3 and 4 activities in Penang.

¹⁷ Indeed, as Hymer (1960) had argued, the relative benefits of host-sites against home-sites were important in the relocation of electronics TNCs in Penang.

Button cluster of Qiaotou

Two brothers began distributing buttons in Qiaotou in 1978 that were purchased from Huangyan in Hubei, which then snowballed into the opening of over 300 stalls by 1982 (Kong, 2008: 3). Markets dominated the initial growth of the industry. Faced with rising costs, exhaustion of raw materials and competition from the emerging economies, Italian button producers responded to requests by Qiaotou's entrepreneurs to outsource the manufacturing stage of button production over the period 1982-84. Chinese clothing firms supplying export markets using foreign brand names were an important catalyst in stimulating the relocation of button manufacturing in Qiaotou.

Button sales began expanding rapidly from 1984 as production in Qiaotou soared to account for over 4,000 stalls, 28,000 varieties of buttons, 14,000 employees, and over RMB2.6 billion in sales by the mid-1990s (see Kong, 2008: 4). Operations in this period were still largely driven by markets. However, the limits of market-based coordination became obvious from the late 1990s when national firms in Qiaotou were not able to upgrade technologically. Government support and social networks helped provide the spur for technological upgrading from 2000. However, the button manufacturers remained mainly confined to levels 3 and 4 activities as the designing and R&D support evolved in design centres, R&D labs and universities that supplied these inputs through strong social networks established by the Yongjia County local government. The world class quality of button designs and materials saw leading buyers of garment value chains visiting Qiaotou exhibitions to seek new designs to import. Over 160 international brands began to source their buttons from Qiaotou by 2005. In fact, the entire cluster of activities from button materials and design to complementary activities related to machinery and components, resins and dyes and other inputs had evolved in Qiaotou by 2006 (see Rasiah, 2012).

While a long history of entrepreneurial experience was important, effective coordination between government, the entrepreneurs and markets was instrumental in the development of the cluster, as well as, provide the firms with testing, training, materials and designing, and R&D support. The local government of Yongjia took measures to stimulate upgrading the infrastructure in the county by focusing on the construction of an information channel that included industrial conferences to connect the button firms to the whole country, develop a special industrial zone, establish incubation facilities for new firms, and mapping and filling the missing components of the button cluster so as to generate a complete button ecosystem, and to support new brand development and knowledge diffusion.

A button and slide fastener Industrial Park was built in Qiaotou to locate the firms with basic infrastructure connected with excellent roads, water and electricity supply, telecommunication, pollution disposal centres, and internet cables. The local government also offered subsidized land and regulation fees for industrial operations with standard factory buildings. To assist firms to resolve collective action problems in fields, such as, training, testing and technological upgrading, the government either attracted or collaborated with meso organizations, such as, universities to support training, standards testing, designing and R&D. Also, the local government encouraged firms and universities to collaborate through science and technology projects. In 2006, the local government encouraged Wenzhou Mailida and Dongda Integrated Chemicals to work with Huanan Technical University to develop new technologies to raise button quality and recycling button material waste productively. New button technologies that were developed from such collaboration – both product and process – helped raise value added of the industry by 2008 (Kong, 2008: 15).

The county government also established the pump and valve industrial science and technology innovation centre with collaboration from Lanzhou Technical University located in Gansu Province in Northwest China, which provides technology information and human resource support for button manufacturing firms in Qiaotou.

The stable exchange rate regime of China ensured that the small button firms in Qiaotou were not drowned by sudden destabilization from plummeting or soaring exchange rates. While markets drove the initiation of the industry in Qiaotou, the Yongjia local government in collaboration with the firms and meso organizations played a critical role in supporting technological upgrading in the industry.

7. Conclusions

The evidence presented in this paper is convincing enough to support the heterodox argument that industrial policy and in particular institutional change is essential to drive technological upgrading. However, consistent with evolutionary arguments the evidence also shows that technological upgrading in particular clusters is conditioned by industry and location specificities. The Buenos Aires and Penang clusters were characterized by the origin and expansion of TNCs that transformed the technological intensity of their operations as competition drove upgrading at host-sites. Important institutional changes were made to the regulatory environment to provide both the incentive and the pressure to upgrade operations to high value added activities. Qiaotou has enjoyed upgrading until R&D and designing, though, such activities are confined to meso organizations located outside the button firms. The government agencies of FC and CORFO have continued to access foreign sources of knowledge to drive firms' participation in level 4 knowledge activities with some firms' participating in level 5 activities. The most sophisticated technological operations in Buenos Aires and Penang involve designing by TNCs.

All four clusters have continued to enjoy strong growth in exports, employment and wages. Skill intensity levels of firms in Penang, Buenos Aires and Los Lagos was high. It was low in Qiaotou because of the organization of economic activity where the button firms researched only specialized in manufacturing with intermediary organizations supplying knowledge generated from designing, testing and R&D. Government-business coordination initiatives in the four locations have produced institutions and intermediary organizations that are uniquely industry and location specific to stimulate upgrading. The networking with meso organizations of universities, training, testing and designing centers and R&D labs organized by Yongjia County has helped support the production of the world's latest buttons in Qiaotou. However, state intervention in Los Lagos, Buenos Aires and Penang was limited to pressuring firms to participate in cutting edge process technologies and designing. Whereas the lack of supply of experienced engineers was reported as a barrier for further upgrading in Buenos Aires and Penang, the reluctance of government to stimulate adaptive engineering capabilities to new knowledge creation has restricted firms' movement to the frontier in Los Lagos. Whereas government policy not to support R&D through grants appears intentional in Buenos Aires and Los Lagos owing to the risks and uncertainty associated with R&D, the inability to mobilize a critical mass of quality engineers is the prime barrier to further technological deepening in Penang.¹⁸

Federal governments in the four clusters have also played important roles to prevent volatile macro disruptions from undermining technological upgrading by either pegging currencies or imposing capital controls during turbulent times in the external environment. Whereas buttons firms in China enjoyed stable exchange rates as the government avoided foreign calls to float the RMB, the governments of Chile and Malaysia imposed capital controls to stem volatile external disruptions in the 1980s and 1990s respectively. Argentinian firms perhaps were the hardest hit as the appreciating peso destroyed firms in the period 2001-03 but the government still responded by suspending the free commercial interchange under the Mercosur Agreement.

Therefore, we conclude that the analysis in the paper compliments industrial policy arguments on the need for interventions in markets to target technological upgrading by focusing on the importance of industry and location specific characteristics. Firms' movement up the technology ladder in particular clusters depends very much on how institutional change supports upgrading. Industrial policy and institutional change have played critical roles with their significance varying over time and industry and location specificities. The only constant in the development of the clusters has been the focus on technological upgrading through effective coordination between firms, intermediary organizations and macro institutions.

¹⁸ Interviews showed that national firms enjoyed R&D grants to undertake knowledge-intensive activities from 1990, but foreign firms were only approved similar grants after 2005.

References

- Abramowitz, M. (1956). Resource and Output Trends in the United States since 1870, National Bureau of Economic Research, *American Economic Review*, 46(2): 5-23.
- Amsden, A. (1985). The division of labour is limited by the rate of growth of the market: The Case of the Taiwanese Machine Tool Industry, *Cambridge Journal of Economics*, 9: 271-284.
- Amsden 1989 *Asia's Next Giant: South Korea and Late Industrialization*, New York: Oxford University Press.
- Aslesen, H.W. (2007) "The Innovation System of Norwegian Aquacultured Salmonids", TIK Working Paper on Innovation Studies, No. 20070606.
- Bank of Chile (2004) "Institution Building in a World of Free and Volatile Capital Flows: A Case Study of Chile", paper prepared by the Central Bank of Chile Working Group for the 11th Asia Pacific Economic Community Finance Ministers Meeting, December 2004, downloaded on September 12, 2012 from www.bcentral.cl/eng/studies/economic-policy-papers/.../dpe12.pdf.
- Becattini G., Bellandi M., De Propris L. (eds.), 2009, *A Handbook of Industrial Districts*. Cheltenham: Edward Elgar.
- Bernat, G. (2008). *Greater Buenos Aires Cluster: A Summary*, Background paper prepared for the Industrial Development Report 2009, United Nations Industrial Development Organization, Vienna.
- Best, M. (2001) *The New Competitive Advantage*, Oxford: Oxford University Press.
- Brusco, S. (1982) "The Emilian Model: Productive Decentralisation and Social Integration", *Cambridge Journal of Economics*, 6(2): 167-184.
- Chandran, G. (2008) "The Electronic Cluster in Penang", Background paper prepared for the Industrial Development Report 2009, United Nations Industrial Development Organization, Vienna.
- Chakravarty, S. (1993) *Development Planning: The Indian Experience*, New York: Oxford University Press.
- Chile (2003) *Aquaculture in Chile*, Santiago: Technopress.
- Coase, R. (1937). Nature of the Firm, *Economica*, 4(16): 386-405.
- Coe, N.M., Hess. M., Yeung, H.W, Dicken, P. and Henderson, J. (2004). Globalizing' regional development: a global production networks perspective, *Transactions of the Institute of British Geographers*, 29:468-484.
- Coe, N.M., Dicken, P. and Hess, M. (2008). Global Production Networks: realizing the potential, *Journal of Economic Geography*, 8(3): 271-295.
- Cripps, F. and Tarling, R. (1973) *Growth in Advanced Capitalist Economies, 1950-1970*, Cambridge: Cambridge University Press.
- Fransman, M. (1986). International competitiveness, technical change and the state: the machine tool industries in Taiwan and Japan. *World Development*, 14(12), 1375-1396.
- Freeman, C. (1988) "Japan: A new national innovation system?", G. Dosi, C. Freeman, R. R. Nelson, G. Silverberg and L. Soete (eds), *Technology and Economy Theory*, London: Pinter;
- Gereffi, G. (2002). International Competitiveness in the Global Apparel Commodity Chain, *International Journal of Business and Society*, 3(1): 27-60.
- Gereffi, G., Humphreys, J. and Sturgeon, T. (2005). The governance of global value chains, *Review of International Political Economy*, 12(1): 78-104.

- Gershenkron, A. (1952) "Economic Backwardness in Historical Perspective", Granovetter, M. and Swedberg, R. (eds), *The Sociology of Economic Life*, Boulder: Westview Press.
- Giuliani, E., Pietrobelli, C. and Rabellotti, R. (2005), "Upgrading In Global Value Chains: Lessons From Latin American Clusters", *World Development*, Vol.33(4), pp. 549–73.
- Giuerrieri, P. and Pietrobelli, C. (2006) "Old and New Forms of Clustering and Production Networks in Changing Technological Regimes: Contrasting Evidence from Taiwan and Italy", *Science, Technology and Society*, 11(1): 9-38.
- Giuerrieri, P., Ianmarino, S. and Pietrobelli, C. (2001) *The Global Challenge to Industrial Districts: SMEs in Italy and Taiwan*, Cheltenham and Lyme: Edward Elgar
- Helleiner, G. (1973). Manufactured Exports from Less Developed Countries and Multinational Firms, *Economic Journal*, 83(329): 21-47.
- Hirschman, A.O. (1970) *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States* Cambridge: Harvard University Press
- Hobday, M. (1995) *Innovation in East Asia: the Challenge to Japan*, Cheltenham: Edward Elgar.
- Hymer, S. (1960) "The International Operations of National Firms: A Study of Direct Foreign Investment". PhD Dissertation. Published by MIT Press in 1976, Cambridge.
- Iizuka, M. and Katz, J. (2011) "Natural Resource Industries, 'Tragedy of the Commons' and the Case of Chilean Salmon Farming", *International Journal of Institutions and Economies*, 3(2): 259-286.
- Kaldor, N. (1967) *Strategic Factors in Economic Development*, Ithaca: Cornell University Press.
- Katz, J. (2000). Structural Change and Labor Productivity in Latin American Manufacturing Industries 1970-96, *World Development*, 28(9): 1583-1596.
- Katz, J. and Stumpo, G. (2001). Structural Change, Productivity and International Competitiveness, *Cepal Review*, 75: 131-152.
- Katz, J. (2006) "Salmon Farming in Chile", Chandra, V. (ed), *Technology, Adaptation and Exports: How some developing countries got it right*, Washington DC: World Bank, pp 193-224.
- Keynes, J.N. (1890) *The Scope and Method of Political Economy*, Cambridge: Batoche Books.
- Keynes, J.M (1936) *The General Theory of Employment, Money and Interest*, Cambridge: Macmillan Cambridge University Press.
- Kim, L. (1997). *From imitation to innovation*. Cambridge: Harvard Business School Press.
- Kong, X.X. (2008) "The Button Cluster in Qiaotou", Background paper prepared for the Industrial Development Report 2009, United Nations Industrial Development Organization, Vienna.
- Lall, S. (1992) Technological Capabilities and Industrialization. *World Development*, 20(2): 165-186.
- Lall, S. (1996) *Learning from the East Asian Tigers*, Basingstoke: Macmillan.
- Lundvall, B.A. (ed) (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive*, London: Pinter Publishers.
- Maggi, C.C. (2007) "The Salmon Farming and Processing Cluster in Chile", Pietrobelli, C. and Rabellotti, R. (eds), *Upgrading to Compete: Global Value Chains, Clusters and SMEs in Latin America*, Cambridge: Harvard University Press. pp 109-140.

World Economic Review

- Malaysia (1986) *The Fifth Malaysia Plan 1986-1990*, Kuala Lumpur: Government Printers.
- Malaysia (1996) *The Seventh Malaysia Plan 1996-2000*, Kuala Lumpur: Government Printers.
- Malaysia (2006) *Ninth Malaysia Plan 2006-2010*, Kuala Lumpur: Government Printers.
- Malerba, F., Nelson, R. Orsenigo, L. and Winter, S. (2008). Public policies and changing boundaries of firms in a "history-friendly" model of the co-evolution of the computer and semiconductor industries. *Journal of Economic Behavior & Organization*, 67(2), 355-380.
- Markusen, A. (1996) Sticky Places in Slippery Space: A Typology of Industrial Districts. *Economic Geography*, 72(3): 293-313.
- Marshall, A. (1920). *Industry and Trade*, London: Macmillan.
- Marshall, A. (1890) *Principles of Economics*, London: Macmillan.
- Mazuzan, G.T. (1994). *National Science Foundation: A Brief History*, Arlington: National Science Foundation.
- Narula, R. and Dunning, J. (2000). Industrial Development, Globalisation and Multinational Enterprises: New Realities for Developing Countries, *Oxford Development Studies*, 28(2): 141-167.
- Nelson, R. (2008). Economic development from the perspective of evolutionary economic theory. *Oxford Development Studies*, 36(1), 9-22.
- Nelson, R. and Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press.
- North, D. (1990) *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
- Perez, A. (2005) "Cluster formation, institutions and learning: the emergence of clusters and development in Chile", *Industrial and Corporate Change*, 14(4): 651-677.
- Pietrobelli C., (1998) *Industry, Competitiveness and Technological Capabilities in Chile. A New Tiger from Latin America?*, London and New York: Macmillan and St.Martin's
- Pietrobelli, C. and Rabellotti, R. (eds), 2007, *Upgrading to Compete: Global Value Chains, Clusters and SMEs in Latin America*, Cambridge: Harvard University Press.
- Piore, M. and Sabel, C. (1984) *The Second Industrial Divide: Possibilities For Prosperity*, New York: Basic Books.
- Porter, M.E. (1990). The Competitive Advantage of Nations, New York: Free Press.
- Pyke, F. and Sengenberger, W. (1992) *Industrial districts and local economic regeneration*, Geneva: International Labour Organization.
- Rasiah, R. (1988). The Semiconductor Industry in Penang: Implications for the New International Division of Labour Theories, *Journal of Contemporary Asia*, 18(1): 44-67.
- Rasiah, R. (1994) "Flexible Production Systems and Local Machine-Tool Subcontracting: Electronics Components Transnationals in Malaysia", *Cambridge Journal of Economics*, 18(3): 279-298.
- Rasiah, R. (1995) *Foreign Capital and Industrialization in Malaysia*, Basingstoke: Macmillan.
- Rasiah, R. (2000) "International Portfolio Equity Flows and the Malaysian Financial Crisis", *Journal of Contemporary Asia*, 30(3): 369-401.

Rasiah, R. (2006) "Human Capital and Cluster Synergies: The Case of Penang and Jalisco", in: A. J. Scott and G. Garofoli (eds.), *Development on the Ground: Clusters, Networks and Regions in Emerging Economies*, London: Routledge.

Rasiah, R. (2010) "Are Electronics Firms in Malaysia Catching Up in the Technology Ladder?", *Journal of Asia Pacific Economy*, 15(3): 301-319

Reinert, E. (1994) "Catching-up from way behind: A third World perspective on First world history", Fagerberg, J., Verspagen, B. and Tunzelmann, N. (eds), *The Dynamics of Technology, Trade, and Growth*, Aldershot: Edward Elgar.

Rodrik, D. (1997) "TFPG Controversies, Institutions, and Economic Performance in East Asia," NBER Working Paper No. 5914, National Bureau of Economic Research, Cambridge.

Rosenberg, N. (1982) *Inside the Black Box: Technology and Economics*, Cambridge: Cambridge University Press.

Rosenstein-Rodan, P.N. (1943). Problems of Industrialisation of Eastern Europe and South-Eastern Europe, *Economic Journal*, 53(210/211): 202-211.

Saxenian, A.L. (2006). *The new argonauts: regional advantage in a global economy*. Cambridge: Harvard University Press .

Schmitz, H. (2004) *Local Enterprise in the Global Economy: Issues of Governance and Upgrading*, Cheltenham: Edward Elgar.

Scitovsky, T. (1954) "Two concepts of external economies", *Journal of Political Economy*, 62(2): 143-151.

Stiglitz, J.E. (2002) *Globalization and Its Discontents*, New York: W.W. Norton and Company.

Veblen, T. (1915) *Imperial Germany and the Industrial Revolution*, New York: Macmillan.

Vidal, L. (2008) "The Salmon cluster in Los Lagos", Background paper prepared for the Industrial Development Report 2009, United Nations Industrial Development Organization, Vienna.

WTO (2011) International Trade Statistics, Geneva: World Trade Organization.

Yap, X.S. and Rasiah, R. (2013) "TSMC's switch from latecomer to first mover in logic chips", *mimeo*.

Young, A. (1928). Increasing Returns and Economic Progress, *Economic Journal*, 38(152): 527-54.

Transacting Without Pricing, Pricing Without Transacting

A case for disconnecting the valuation function from the exchange function of financial markets

Jacques-Olivier Charron

Associate researcher, IRISSO, Paris-Dauphine University, France

Abstract

The paper argues in favor of a radical disconnection between the exchange function and the valuation function of financial markets. It defends it from an investee's point of view, mainly on the grounds that valuing financial assets is a matter of judgment. Financial assets do not have any intrinsic utility; hence allowing markets to price them implies their price changes are determined by expectations. These latter rely on information that can be variously framed and interpreted, hence no a priori and stable valuation criteria can prevail. Consequently, investees bear the consequences of judgments that do not meet basic criteria of justice. Insulating the valuation function from the transaction function of financial markets through the creation of independent valuation institutions could be a way to tackle this problem.

Key words: financial markets, valuation, judgment, investor, investee.

Introduction

Financial markets have traditionally been considered, at least since Walras, as the "purest" form of markets, the "real" markets closest to the model of "pure and perfect" ones. What makes them so is the way prices are determined on it: order books exhibit, for a given financial asset at a given moment, lists of potential buyers and sellers, each of them specifying the quantity it is willing to buy or sell and at what price it is willing to do so. The matching of supply and demand, that allows trades to be made, can be realized through the kind of machines Fabian Muniesa called "walrasian robots" (Muniesa 2000, 2003). What makes it so "pure" is that, at least apparently, really nothing else than supply and demand determines the price.

The question we are going to start with is the following: for actors who buy or sell on these markets, is the price the most important issue?

1. It's the return, stupid

What is a financial asset here for? Basically, a financial asset is a claim¹ that has been made liquid through a legal, social and institutional framing: a firm's ownership, that is a claim on its profits, or a debt, that is a claim on the debtor's wealth, becomes a set of financial assets when it is made negotiable on a market. A financial asset is here, first and foremost, to be negotiable, to be liquid, which requires the establishment of specific markets (Orléan 1999).

But then, the very development of these markets has made them become investment opportunities as such, more liquid than any others. Liquidity attracts investors, because it allows them to get cash at the very moment that suits them the best. Financial markets, definitely, are an investor's world. And what is the most important issue for an investor? Return. In finance, God has a name: it's called Return.

Return is not a price: it's a price difference, a price change. When you hear that a stock has risen

¹ In Minsky's terms, a financial instrument is "a commitment to pay cash at some time or if some event occurs" (Minsky 1986, p. 69)

by 5% at \$50, the most important information here is “+5%”, not “\$50”; it’s the return, not the price. When Nassim Taleb famously revived an old debate on the shape of statistical distributions (Taleb 2007), the distributions he was talking about were not price distributions but price changes distributions, as was already the case, for example, back in 1965, in Eugene Fama’s thesis (Fama 1965).

The important issue on financial markets is not so much to know how prices are determined on these markets, but how price changes are determined. That is clearly not the same question: instead of wondering how supply and demand are matched at a given moment, we have to wonder how supply and demand change over time, that is to understand the dynamics of financial markets.

If we treat this as a practical question, not as a theoretical one, the answer is relatively simple: once a trade has been made at a given price, the next one will be clinched at a higher price if sell orders at this price can be matched with buy orders at the same price on the order book, which will generally happen if more buy orders appear on it, which signals a willingness to buy strong enough for the actors for them to be ready to pay more. Now, if you want to find an explanation for this “willingness to buy” (or to sell), the first one that will probably come to your mind is to say these actors just expect a price rise, expect a positive return that will allow them to sell later at a profit. Expectations, it seems, are just what determines price changes².

Most real-life investors, especially professional ones, would no doubt balk at such a simplistic description, for they just don’t see themselves as vulgar speculators, but, so the story goes, as sophisticated risk managers. This is not only a story, far from it, for most of them effectively base their investment strategies (or those of their robots...) on one form or another of risk modeling. What we can simply point out here is that these models rely, among other things, on specific assumptions on price changes distributions. Risk-obsessed investors do not cease to be return-obsessed: when they abide by a passive strategy (by, say, buying ETFs...), it’s because they think an active one would yield, in the end, a smaller return. They are risk-obsessed *because* they are return-obsessed. Which means, in our view, that return expectations ultimately determine price changes.

2. So why bother about prices?

There may be many ways to describe a financial crisis, but my guess is the most straightforward (and purely descriptive) is the following : a sharp and unexpected fall of the price of financial assets of some kind, that makes most people think and realize these assets had been grossly overvalued. That means a financial crisis is, at first glance, a valuation crisis.

That’s not really the case from an investor’s point of view: in this perspective, what is grave here is neither the price fall nor the previous overvaluation *as such*: it is just the unpredictability of the fall. Hence the classical blame on economists: why didn’t they see this coming? Again, price is not the main concern for investors: it’s return, its predictability, hence the predictability of price changes.

It is a major concern, though, for investees. I used the term investee (first in Charron 2010) not only to designate “the company in which an investment is held” (definition seen on <http://www.encyclo.co.uk/define/investee>, accessed 08/30/2011) but, in a much broader sense, any kind of actor, organization, state... that is valued by a financial market, for example through the pricing of its shares (for a listed firm) or through the interest rates applied to its debt (for a state). I meant here to make a symmetry discernible: investors value, investees are valued, the first ones are subjects, the second ones objects. If we consider the valuation function of financial markets, that is, the way prices are determined on it, investors are the actors who effectively perform it. But the actors that bear the consequences of these operations are the investees; and if we take in account the decisive influence of these valuations in the kind of financialized capitalism we have been experiencing since (roughly) 30 years, it doesn’t seem much exaggerated to say that *all the rest* of the society is made of investees. Of course, a given organization can

² This is consistent with Kaldor’s definition of speculation : “the purchase (or sale) of goods with a view to re-sale (re-purchase) at a later date, where the motive behind such action is the expectation of a change in the relevant prices relatively to the ruling price and not a gain accruing through their use, or any kind of trans-formation effected in them or their transfer between different markets.” (Kaldor 1939, p. 1)

be both an investor and an investee at the same time, but the distinction can be made analytically. Even if you are both an investor and an investee, what you expect from financial markets as an investor is not what you expect as an investee. Whereas return matters only for investors, price does for investors, hence for firms, states, societies.

3. But what does determine prices, then?

We said previously that return expectations determine price changes, that is... return. Orléan (1999) called this the “self-referential” nature of financial markets (and Soros (2008) “reflexivity”). Price changes on financial markets are determined by changes in supply or demand that in turn are determined by expectations, which implies the “law” of supply and demand doesn’t apply to them. This comes from the very nature of financial assets: they do not have any intrinsic utility, as consumer goods do.

The “law” of supply and demand may be, or even is, theoretically and empirically questionable, but it nevertheless does make sense on consumer goods markets; on financial markets, it is not enough to say this law doesn’t work: it just doesn’t make sense on them. When the price of a financial asset or a financial asset class rises, it can provoke a rise in demand, fueled by optimistic expectations. It’s not necessary the case, of course, but it can occur in a way that won’t show up in consumer good markets, simply because these latter are supposed to be consumed, hence not resold, hence price changes expectations do not play such a role on these markets. Moreover, the development of futures markets have made the supply and demand of financial assets nearly infinite, whereas the supply and demand of consumer goods, which are bought for themselves because they have an intrinsic utility (at least a perceived one), are always constrained by various finite factors: the purchasing power of consumers and the price of raw materials, to name two major ones.

So, even if *apparently* nothing else than supply and demand determines prices on financial markets, price determination there has nothing to do with the “law” of supply and demand which characterizes, among other features, “perfect” markets. Price determination is completely subordinate to the market dynamics that makes price changes depend from expectations.

The way financial assets are priced is decisively determined by the way they are exchanged. Just allowing markets to price financial assets implies their price changes are determined by expectations, which in turn is caused by the fact that financial assets do not have any intrinsic utility.

4. A question of judgment, which requires criteria

When we say that a financial asset is “overvalued” or “undervalued”, it clearly assumes there is some kind of optimal valuation of this asset, which corresponds to what is generally called its fundamental value.

Then, to determine the fundamental value of a financial asset is not a matter of science: it is a matter of judgment (Bhidé 2010). It is so, not because we haven’t discovered yet a scientific solution, the right model, the right formula, but because of the very nature of the object we are talking about. A financial asset is nothing more than a promise, and nothing less. What they have is an expected utility, which is not a scientific matter but a matter of judgment.

The problem is that, on actual financial markets, the way these judgments are made simply do not meet the most basic criteria of justice.

When you are judged, you have the right to know on what grounds, to know which criteria are applied. For investees, the most important is not to know how prices are established but what the criteria of pricing are.

Now, what does the way financial markets operate mean in terms of valuation criteria? It means that valuation changes over time depending on what price changes investors expect and how they “translate” these expectations into investment strategies. We cannot even say that these changes depend on information (about valuation, supposedly...), because information is necessarily interpreted and framed, and can be so in different ways that can also change over time. Consequently, there are simply no *a priori* valuation criteria for financial assets. Some may be found *a posteriori*, but again nothing guarantees their

stability. Zajac & Westphal (2004) provide a good example of this; studying market reactions to stock repurchase plans, they show that negative reactions to such decisions prevailed before the 1980's, these plans being interpreted as an inability to find profitable development opportunities to be funded, and that, then, positive reactions to the *same* decisions, the same informations, prevailed later because they were seen through the lens of agency theory, through another interpretative frame³.

Financial instability, then, is not primarily caused by some kind of irrationality, but by the very way financial exchanges function, and more specifically by the simple fact that pricing is made through trading and trading through pricing.

From an investee's point of view, what makes financial markets problematic and (at least potentially) harmful is first and foremost the absence of stable and publicly known valuation criteria; it entirely lies, then, in their valuation function.

As a matter of fact, financial markets have at least two functions: an exchange function (they make financial claims liquid), a valuation function (they price these claims). One could mention other ones, for example a capital allocation function, but we will concentrate on exchange and valuation, simply because it is the interplay between these two ones that, in our view, determines the dynamics of financial markets and hence their crises.

5. Disconnection

Here comes our basic idea: what is needed from an investee's point of view and, one should say, from the society's point of view are stable, publicly known and publicly debated and established valuation criteria. This cannot be achieved through a market mechanism, and this is why the valuation function should be insulated from the others through the establishment of independent valuation institutions.

How would it work? Prices of financial assets would be determined once or twice a year by a valuation institution. In the meantime, they would not move. Anybody willing to buy or sell a given quantity of such an asset would have to find a counterparty willing to (respectively) sell or buy it *at this price*. Any transaction done at another price would be illegal, null and void. A possible way to enforce this would be to rule that an investor that would have bought an asset at an "illegal" price would not be recognized as its legal owner. To be a member of the valuation body would require both a professional legitimacy, acquired through diploma and experience, and a democratic one, acquired through direct election or appointment by an elected institution. At this stage, I won't go further in the details, simply willing to defend the principle of this institutional design.

6. Objections to disconnection

I will now try to answer to the two main objections that will probably come to your mind:

-why would an institution be better than a market at pricing?

-why would experts be better as a crowd of investors at doing this?

You could say these are two different ways to ask the same question. I nevertheless chose to expose these two ways separately because each of them corresponds to a highly popular and successful argument in favor of the rule of markets: the Hayek's argument, the Surowiecki's argument.

Other sources could and should be discussed, but Hayek's article "The use of knowledge in society" (Hayek 1945) is probably the best known and the most influential text explaining why the market price system is the best one to establish prices. Here's the rub: its reasoning doesn't apply to financial markets. The first reason for it is that it deals with the allocation of scarce resources, whereas financial assets cannot be considered so, except in extreme cases (when a single investor makes a "corner" on a market, for example). But the main reason appears when we see what is at the core of the argument: speaking of the individual agent (or, in his terms, the "man on the spot"), Hayek explains:

"it does not matter for him why at the particular moment more screws of one size than of

³ See also Brière (2005) for examples on bond markets.

another are wanted, why paper bags are more readily available than canvas bags [...]. All that is significant for him is how much more or less difficult to procure they have become compared with other things with which he is also concerned, or how much more or less urgently wanted are the alternative things he produces or uses. [...] It is in this connection that what I have called the economic calculus proper helps us, at least by analogy, to see how this problem can be solved, and in fact is being solved, by the price system." (Hayek 1945, p. 525).

To be consistent, this reasoning requires that price changes reflect changes in the relative importance, in terms of a relative "difficulty to procure", of what is priced. Now, as we saw before, this is not the case on financial markets, where price changes are determined by expectations. Moreover, the problem Hayek deals with is how to convey the relevant information, hence it completely ignores the fact that information is interpreted and framed, whereas this is of decisive importance on financial markets, where expectations are often based on different ways to frame the same information.

Surowiecki is not an academic economist, but the "wisdom of crowds" argument has proved influential, for example through the development of prediction markets. Surowiecki's argument, precisely, is centered on the ability of "crowds", that is of non-coordinated and independent individual agents, to guess or to predict specific outcomes, but also to make decisions more generally. If we focus on financial markets, we will first recall that, for investors, the point is not to guess some "true" value but to maximize their return; their expectations, through investment strategies, determine this return. Investors *are* the market, to the point that the price is not something that stays before them at a distance, but something they make happen. The famous example, told by Francis Galton, of a crowd guessing the weight of an ox clearly has very little to do with a situation where, so to speak, the ox would not exist without the crowd and this crowd makes its weight vary. In this case, the verbs "to guess" or "to predict" are clearly improper.

So, if we consider price determination on financial markets as a problem, it is clearly not, in Surowiecki's terms, a "cognition problem". This is not so clear-cut for this author, who, on several occasions in his book, seems to treat it this way, particularly when he stresses the inability of most professional investors to beat the market, which means their predictive performance is what is at stake. It is not the case if what they face is, again in Surowiecki's terms, a "coordination problem", where their decisions contribute to the making of the reality they are trying to evaluate. He seems conscious of that when, dealing with bubbles and krachs, he does provide an insightful analysis, partly based on Keynes, taking in account the role played by interdependent opinions and expectations, rightly explaining how it makes a difference with non-financial markets. But then what he recommends is for investor to make their decisions *independently*. Surowiecki's problem, in our view, is that he doesn't go as far as considering interdependence as *structural*. If he did, then he would have to recognize, as we do, that investor's decisions are heavily constrained by the features of the interdependence they are inevitably involved in (Elias 1978, 1983). For us, to behave independently, for an investor, is not an option, at least not a realistic and profitable one. The interplay between trading and pricing on a market for goods that don't have any intrinsic utility implies, as we saw previously, that prices will depend on expectations produced without a priori stable criteria.

Conclusion

Precisely, from an investee's point of view, the point is neither to guess what the price will be some seconds, days or months ago nor to urge investors to make their decisions independently, but to know on which criteria they are judged. Focusing on the way investors produce estimates and make decisions means you are embedded in an investor's point of view. This is only one side of the story, where success, whatever it takes, is the only criterion: we do not only make decisions, we also bear the consequences of decisions made by others. And in this case we have the right to know of their rationale and to make sure they are based on much more precise criteria. To take an extreme example, when you are arrested, your most basic human right is to know what you are accused of precisely, to know your rights, to know which legal procedure will apply and to know what kind of penalty you incur. Listed companies do not even get this from markets: when their stock price varies, it can be so for various reasons, these reasons can change over time,

the way their performance is estimated can also change and so forth.

Financial markets may be better than any body of experts at serving certain investor's needs, notably liquidity. But they are not at pricing in a way that meets the basic needs of investees for stability and justice. Independent valuation institutions would, in our view, do it better and, by doing so, would also probably serve long-term interests of investors.

References

Bhide, A. 2010, *A call for judgment: sensible finance for a dynamic economy*, Oxford University Press.

Briere, M. 2005, *Formation des taux d'intérêt*, Economica.

Charron, J-O. 2010, *Disambiguating efficiency: a question of points of views, transparencies and research designs*, Presentation for the Finance in question / Finance in crisis conference, CRESC, Manchester University.

Elias, N. 1978, *What is sociology ?*, Hutchinson (first published [in German] in 1970)

Elias, N. 1983, *The court society*, Blackwell (first published [in German] in 1936)

Fama, E. January 1965, 'The behavior of stock market prices', *The Journal of Business*, Vol. 38, n°1, pp. 34-105.

Hayek, F. A. September 1945, 'The use of knowledge in society', *The American Economic Review*, Vol. 35, n°4, pp. 519-530.

Kaldor, N. October 1939, 'Speculation and economic stability', *The Review of Economic Studies*, Vol. 7, n°1, pp. 1-27

Minsky, H. 1986, *Stabilizing an unstable economy*, McGraw Hill.

Muniesa, F. 2000, Un robot walrasien : cotation électronique et justesse de la découverte des prix, *Politix* 13 (52), pp. 121-154.

Muniesa, F. 2003, *Des marchés comme algorithmes : sociologie de la cotation électronique à la Bourse de Paris*, PhD thesis in sociology, Ecole des Mines de Paris.

Orlean, A. 1999, *Le pouvoir de la finance*, Odile Jacob.

Soros, G. 2008, *The new paradigm for financial markets : the credit crisis of 2008 and what it means*, PublicAffairs.

Surowiecki, J. 2004, *The wisdom of crowds*, Doubleday.

Taleb, N. 2007, *The black swan*, Random House.

Zajac, E & Westphal, J. June 2004, 'The social construction of market value : institutionalization and learning perspectives on stock market reactions', *American Sociological Review*, vol. 69, pp. 433-457.

Merit Regulation via the Suitability Rules

Ronald J. Colombo¹

Professor of Law, Maurice A. Deane School of Law, Hofstra University, USA

Abstract

The philosophy underpinning federal securities regulation in the United States is one of disclosure. This has been the case since the inception of federal securities regulation in 1933,² and continues to be the case with Congress's most recent enactments on the subject, contained in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010.³

In the wake of the financial industry's collapse in 2008, and the recession it helped spark, some have questioned whether this paradigm remains advisable.⁴ They have suggested the introduction of merit regulation into the U.S. securities law regime, whereby the government would not merely mandate certain issuer disclosures, but would also prevent the offering of securities deemed too risky.⁵ Although not revolutionary (as several American states, and nations such as China, have a merit component to their securities laws), the concept of merit regulation is indeed largely alien to the scheme of U.S. federal securities regulation.⁶ As such, it would be a transformative development.

There is, however, a far more modest way of approximating the same result. And it builds upon our existing regulatory infrastructure: suitability rules. Via enhancements to the suitability rules, policymakers can achieve much of what merit regulation promises, without the significant, accompanying drawbacks. Properly enhanced, such rules could provide a system that safeguards investors from unsuitably risky investments on a case-by-case basis, thereby depriving neither corporations, nor investors, of mutually beneficial opportunities that might be fully appropriate for them despite their inappropriateness for others. It could also furnish an additional tool by which authorities could regulate systemic risk.

Like Caesar's Gaul, this Article is divided into three parts. Part I will describe the disclosure-based federal securities regulatory regime that prevails in the United States today. It will highlight the limitations of this regime, as underscored by merit-regulation proponents seeking its reform. Part II will describe merit regulation, both in theory and in practice. It too will end with an articulation of the drawbacks associated with such an approach. Part III will describe the "suitability rules" component of U.S. securities law, as they are currently formulated. Part III will also demonstrate how the suitability rules can be utilized to essentially achieve the desideratum of merit regulation without the costs associated therewith.

Key words: suitability, merit regulation, securities regulation, broker regulation

1.1 The disclosure based regime of U.S. federal securities regulation

Prior to 1933, securities regulation in the United States was largely a matter of state concern – much like corporate law.⁷ As will be discussed in Part II, state securities regulation (commonly referred to as the "blue sky laws") followed primarily a merit-based approach.⁸ Pursuant to this approach, "securities proposed to be sold in a state [must] be submitted to an administrative agency for review as to their 'merit' or intrinsic

¹ I thank my colleague J. Scott Colesanti for his helpful assistance with this project, along with the World Economics Association (WEA) for including this paper in its November 2012 Conference "Rethinking Financial Markets"

² Securities Act of 1933, ch. 38, 48 Stat. 74 (1938) (current version at 15 U.S.C. §§ 77a-aa (2006)).

³ Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 171(b)(4), 124 Stat. 1376 (2010).

⁴ See Joan MacLeod Heminway, *The Best Of Times, The Worst Of Times: Securities Regulation Scholarship And Teaching In The Global Financial Crisis*, 5 J. BUS. & TECH. L. 59, 68 (2011).

⁵ See *id.*

⁶ Robin Hui Huang, *The Regulation of Securities Offerings in China: Reconsidering the Merit Review Element in Light of the Global Financial Crisis*, 41 HONG KONG L. J. 261, 261 (2011); Jennifer J. Johnson, *Private Placements: A Regulatory Black Hole*, 35 DEL. J. CORP. L. 151, 156 (2010).

⁷ 1 FEDERAL REGULATION OF SECURITIES §1 (1995).

⁸ See *infra* Part II.

worth.”⁹ Despite a generation of experience with such an approach,¹⁰ the federal regulation of securities took a decidedly different tack.¹¹ Instead of following the states’ merit-based approach, the U.S. Congress adopted a disclosure-based regime of securities regulation when it moved into action.¹²

Prompting federal action was the stock market crash of 1929, and the practices leading up to it (especially as they came to light in subsequent investigations and inquiries).¹³ And it is easy to see why. Quantitatively, the losses were staggering. “The aggregate value of all stocks listed on the NYSE on September 1, 1929, was \$89 billion.... In 1932, the aggregate figure was down to \$15 billion.”¹⁴ Qualitatively, the cover was ripped off of corporate practices that were roundly condemned as unscrupulous and immoral.¹⁵ Among other things, prior to the crash, a nation hungry for speculative securities was willingly fed by promoters whose practices ranged from hype and puffery to downright misrepresentation.¹⁶ “[I]nvestors, who had been given little information about the securities they had invested in, were allured by promises of easy wealth and became victims of widespread fraud and manipulation.”¹⁷

Congress’s initial response was the 1933 Securities Act.¹⁸ Pursuant to the 1933 Act, an issuer of securities is required to make certain, specific public disclosures before selling its securities.¹⁹ These disclosures are made to the SEC via a publicly available registration statement – and summarized in a prospectus to be distributed to prospective investors before or at the time of their securities purchase.²⁰ To buttress the credibility of these disclosures, strict anti-fraud rules are also contained in the 1933 Act, making it far easier (in many cases) for a defrauded investor to recover from an unscrupulous issuer than had been the case under state law.²¹

The 1933 Act was followed up by the 1934 Securities Exchange Act, which added to the volume of required regulatory disclosures.²² Under the 1934 Exchange Act, the issuer of a publicly traded security is obliged to issue periodic reports well after an offering of securities: annual reports, quarterly reports, and periodic reports triggered by certain specified occurrences.²³

Additional securities legislation flowed out of Congress throughout the 1930s, but this legislation was more targeted in its focus, as can be gleaned from the names of the acts in question: the Public Utility Holding Company Act of 1935, the Trust Indenture Act of 1939, the Investment Company Act of 1940, and the Investment Advisors Act of 1940.²⁴ Further, in these acts, Congress demonstrated a greater willingness to engage in aggressively substantive lawmaking – moving beyond simply disclosure and antifraud rules.²⁵ But, to the extent that this is so, these acts, and their specialized applicability, represent the proverbial exceptions that prove the rule: the overall approach to the regulation of securities issuance and trading in the United States is set forth in the 1933 and 1934 Acts, which are firmly disclosure-based pieces of legislation.

A driving force behind the federal approach was Louis D. Brandeis, who famously remarked that “Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.”²⁶ In other words, enhanced disclosure would lead to better securities industry practices, by making unsavory practices more difficult to conceal or get away with.²⁷

Other factors were at work as well, however. The British experience with securities regulation, with which Congress was also familiar, was disclosure based, giving Congress something other than the blue-sky

⁹ Jonathan R. Macey & Geoffrey P. Miller, *Origin of the Blue Sky Laws*, 70 TEX. L. REV. 347, 348 (1991).

¹⁰ See *id.*

¹¹ See Louis Loss, Joel Seligman, & Troy Paredes, 1 FUNDAMENTALS OF SECURITIES REGULATION 45-46 (2011).

¹² See *id.*

¹³ See Ronald J. Colombo, *Buy, Sell, or Hold?*, 73 BROOK. L. REV. 91, 119-120 (2007).

¹⁴ Loss, *supra* note 11, at 45.

¹⁵ See Colombo, *supra* note 13, at 119-121.

¹⁶ See Macey, *supra* note 9, at 355.

¹⁷ See Susanna Kim Ripken, *The Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, 58 BAYLOR L. REV. 139, 144 (2006).

¹⁸ See Loss, *supra* note 11, 45-46.

¹⁹ See *id.* at 57.

²⁰ See *id.*

²¹ See *id.* at 47.

²² See *id.* at 58.

²³ See *id.* at 58-59.

²⁴ See 60-74.

²⁵ See 60-61.

²⁶ *Id.* at 46.

²⁷ See Ripken, *supra* note 17, at 151.

model to consider.²⁸ Then there was the prospect of federal merit regulation in practice: would not an enormous and potentially unworkable new apparatus be required to implement such a regime?²⁹ Further, Congress was attempting to thread the needle and devise a regulatory regime that would simultaneously protect investors without impeding corporate access to the capital markets.³⁰ It found a disclosure-based approach best suited to this delicate balancing act.³¹ Congress also feared the signaling effect that federal merit regulation might bring about, as it wanted to “avoid the implicit approval by the federal government of the merits of any securities offered for sale to the public.”³²

Finally, there was the economic argument that the provision of adequate information via disclosure would lead to increased “transparency and efficiency in the securities markets.”³³ This, in turn, produces “increased price stability and diminished market volatility.”³⁴ Although disclosure may not be a panacea, it was seen as something coming very close to one.

In sum, then, the “main goal of the securities laws [was] to provide sufficient disclosure to enable investors to make informed decisions about the securities they buy and sell.”³⁵

Over time, policymakers have remained fairly faithful to this mission. The two most recent and comprehensive reforms of the securities laws, the Sarbanes-Oxley Act of 2002³⁶ and the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010³⁷ attest to this. Although each contains an aggressive dose of substantive regulatory reform, each also generally carries on the tradition of regulating through disclosure. Sarbanes-Oxley, for example, requires public companies to disclose *whether* they have at least one “financial expert” on their board of directors – but does not actually require the presence of one.³⁸ Dodd-Frank directed the SEC to promulgate rules requiring public companies to disclose their efforts to avoid the purchase of “conflict minerals” – without prohibiting the actual purchase thereof.³⁹ Indeed, it has been observed that “[t]here is a recurrent theme throughout [the federal securities laws] of disclosure, again disclosure, and still more disclosure.”⁴⁰ As Susannah Kim Ripken observed:

The disclosure of material information is said to do everything from producing more transparent and efficient markets, to making corporate executives behave more honestly and diligently, to decreasing investor risks and protecting the public interest.

In the wake of continued financial crises since the Great Depression (and especially in the wake of the 2007-2009 recession), some have questioned the effectiveness of the American disclosure-based system of securities regulation.⁴¹ They echo arguments raised long ago by William O. Douglas, expressed in the 1930s, that a disclosure-based would be simply too simplistic for the complex world of modern finance:

The whole business [that is, the 1933 and 1934 acts] is essentially a “nineteenth-century piece of legislation” that unrealistically envisages a return to “Main Street business.” This

²⁸ See Charles H.B. Braisted, *State Registration of Securities: An Anachronism No Longer Viable*, 78 WASH. U. L. Q. 401, 405 (2000).

²⁹ See *id.* William O. Douglas, who favored a merit-based approach, was apparently fully aware of what it would entail yet not dissuaded: he acknowledged that it would require a “government agency ... ‘a thousand fold more complex than the ... Interstate Commerce Commission.’” See Loss, *supra* note 11, at 48.

³⁰ See Braisted, *supra* note 28, at 405.

³¹ *Id.* (“Congress chose the disclosure philosophy as the best protection for public investors because it allowed each investor to make his or her own investment decision based on full information, without imposing an unreasonable restraint on legitimate business finance.”) James Cox observes that “there is a clear inconsistency between professing obedience to capitalism and allowing civil servants to dictate what ventures may raise funds in capital markets.” James D. Cox, *Regulatory Duopoly in the U.S. Securities Markets*, 99 COLUM. L. REV. 1200, 1200 (1999). Although there is truth in this, I question whether the New Deal policymakers behind the 1933 and 1934 Acts can be said to have professed “obedience to capitalism.”

³² Braisted, *supra* note 28, at 405.

³³ Ripken, *supra* note 17, at 153.

³⁴ See *id.* at 154.

³⁵ See *id.* at 144.

³⁶ Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (2002).

³⁷ Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 171(b)(4), 124 Stat. 1376 (2010).

³⁸ See Ripken, *supra* note 17, at 144-45.

³⁹ Sarah A. Altschuller, Amy K. Lehr, & Andrew J. Orsmond, *Corporate Social Responsibility*, 45 INT’L LAW. 179, 183-84 (2011).

“Conflict minerals” include tantalum (coltan), cassiterite (tin), wolframite (tungsten) and gold. The sale of conflict minerals, it is believed, helps armed groups fund the purchase of weapons and allows them to continue hostilities in the [Congo].” *Id.*

⁴⁰ Ripken, *supra* note 17, at 145 (alteration in original).

⁴¹ See *id.* at 139-148.

explains, among other things, the “great reliance placed on truth about securities, as if the truth could be told to people who could understand it – a supposition that might be justified if little units of business were seeking funds and people were buying shares with the modicum of intelligence with which they are supposed to buy wearing apparel or horses.” We cannot “turn back the clock” to simpler says, said Douglas. We must perfect a plan for control of our present forms of organization so as to harness the “instruments of productions not only for the ancient purpose of profit but also for the more solely evolving service in the sense of the public good.” ... “The control needed is one which would combine regulation by industry with supervision by government.” ... Ultimately, there must be some form of control over access to the capital market, Douglas believed.⁴²

More specifically, the critiques of the U.S. disclosure-based system of securities regulation can be divided into two categories: those criticizing the system from the standpoint of investor protection, and those criticizing the system from the standpoint of systemic risk.

With regard to the question of investor protection, the critics of the disclosure approach mistrust the “prudence of investors.”⁴³ They question whether investors, even if armed with all necessary and relevant information, will make investment decisions that are sound and reasonable.⁴⁴ Indeed, they question whether investors will even be able to capably understand the information disclosed to them.⁴⁵ In short, these critics seek to protect investors from themselves. Their perspective is unquestionably a paternalistic one, but one also one rooted in historical experience.⁴⁶ “Exhibit A” in support of this position has been the financial crisis of 2008.⁴⁷

The critique also draws support from the observation that financial instruments have grown increasingly (and incredibly) complicated.⁴⁸ Indeed, it has been stated that “some structures are getting so complex that they are incomprehensible.”⁴⁹ This calls into question the utility of disclosure as a means of investor protection.⁵⁰

The second line of attack concerns systemic risk. Simply put, the capital markets are deemed simply too vital and complicated to be left, fundamentally, in private hands.⁵¹ As per Douglas, “a more thoroughgoing and comprehensive control is needed.”⁵² The capital markets “should be lodged ‘in the hands not only of the new self-disciplined business groups but also in the hands of government agencies whose function would be to articulate the public interest with the profit motive.’”⁵³ This concern is not so much about protecting investors from themselves, but about protecting everyone from those who would invest imprudently. It is predicated upon the notion that an individual’s (or, more likely, an institution’s) poor investment decisions can impose negative externalities upon others. Once again, “Exhibit A” in support of this contention is the financial crisis of 2008.⁵⁴ The crisis demonstrates vividly how the disclosure approach

⁴² See Loss, *supra* note 11, at 48.

⁴³ Daniel J. Morrissey, *The Road Not Taken: Rethinking Securities Regulation and the Case for Federal Merit Review*, 44 U. RICH. L. REV. 647, 649 (2010).

⁴⁴ See *id.* at 681.

⁴⁵ See *id.*

⁴⁶ See *id.* at 648.

⁴⁷ See *id.* The degree to which the financial crisis of 2008 serves to condemn the U.S. disclosure-based system of securities regulation is an interesting question – and one that is a bit beyond the scope of this Article. On the one hand, most of the financial instruments and transactions at the heart of the crisis fell outside of the disclosure rules. See *id.* 683. Due, in large part, to the sophistication of the actors in question, much of the activity that precipitated the crisis was exempt from the myriad rules and regulation requiring disclosure. See *id.* This suggests that, if anything, the crisis provides a more trenchant indictment of the exemptions from the disclosure regime – rather than an indictment of the disclosure regime itself. That said, the actors involved were typically quite sophisticated – and either had access to (or could have likely obtained) whatever information would have been provided to them had the transactions not been exempt from the disclosure requirements. Thus, it is unlikely that things would have been significantly different had these transactions been fully registered and accompanied by the disclosures required of non-exempt transactions. See *id.*

⁴⁸ See Steven L. Schwarcz, *Rethinking the Disclosure Paradigm in a World of Complexity*, 2004 U. ILL. L. REV. 1, 12 (2004).

⁴⁹ *Id.*

⁵⁰ See *id.*

⁵¹ See Morrissey, *supra* note 43, at 681.

⁵² *Id.* (quoting William O. Douglas, *Protecting the Investor*, 23 YALE L. REV. 521, 522-23 (1934)).

⁵³ *Id.*

⁵⁴ See *id.* at 683.

to regulation leaves each of us exposed to harm and fallout resulting from the poor investment choices of others.⁵⁵

2. Merit regulation

As already indicated, when the United States enacted a regime of federal securities regulation, the road not taken was that of merit regulation.⁵⁶ “Merit regulation” is what generally characterized the state regulation of securities at the time of the federal securities laws’ promulgation.⁵⁷ (Indeed, merit regulation continues to characterize much of the state regulation of securities today – an area of regulation persists to the degree that it has not been preempted by federal legislation.⁵⁸)

The American experience with merit regulation (and, moreover, with securities regulation generally) was commenced in the early part of the twentieth century.⁵⁹ Prior to that time, securities transactions were not subject to any specialized body of law or regulation.⁶⁰ In 1911 Kansas enacted the first law in America regulating the sale of securities – and by the Great Depression, every state had followed suit.⁶¹ The Kansas law, like many, was enacted in response to widespread securities fraud in that state.⁶² As a result, these state securities laws were soon given the name “blue sky laws,” because they were passed to combat the efforts of securities fraudsters to sell building lots in the blue sky to gullible investors.⁶³

As the moniker “merit regulation” suggests, this approach to securities regulation entails “review by a state securities commissioner (or administrator) to determine whether the quality of a given issue of securities was [or is] adequate for sale in that state.”⁶⁴ As one commentator put it: “Merit regulation seeks to foster fairness, to regulate the riskiness of investments offered, to prevent fraud, and generally to increase investor confidence.”⁶⁵ Of course, the same can be said (and is said) of disclosure-based regulation. The difference being is that the disclosure approach allows investors and the market to determine whether a given offering is fair or excessively risky, whereas it is a government official who makes that determination in merit regulation.

Merit-regulation does not purport to screen for only the highest-quality investment options – quite the contrary, it endeavors merely to “assure that all securities will be of at least minimum quality.”⁶⁶ That is, the securities must simply be found fit for sale (and purchase) – they may still be generally undesirable for a host of reasons. The standard employed in determining the fitness of a security is ordinarily that the terms of the securities offering be “fair, just and equitable.”⁶⁷ Although the factors consulted in making this determination vary somewhat from state to state, a universally important factor is an evaluation of the portion of proceeds going to the issuer versus the portion going to underwriters.⁶⁸ Other factors consulted include “options and warrants to be issued in connection with the offering, cheap stock, the extent of the existing capitalization of the issuer, the promoters’ investment, dilution, the offering price, voting rights, loans to promoters and offering expenses generally.”⁶⁹ In short, the merit regulator is (in many states) empowered to “act[] as a negotiator in getting a better deal for investors.”⁷⁰ The merit regulator polices the transaction to protect investors from offerings that are unscrupulous, unfair, or simply too risky.⁷¹

Additionally, merit regulators have assumed the power to “directly intervene to require changes in the internal structure of a securities issuer, the relations among insiders and outsiders, and the terms of the

⁵⁵ See *id.*

⁵⁶ See *supra* text accompanying notes 7-12; see also Morrissey, *supra* note 43, at 647.

⁵⁷ See RICHARD A. BOOTH, FINANCING THE CORPORATION §9.13 (2012).

⁵⁸ See *id.* That said, not all state securities legislation is truly “merit regulation” – some take a different approach. See Roberta S. Karmel, *Blue-Sky Merit Regulation: Benefit to Investors of Burden on Commerce?* 53 BROOK. L. REV. 105, 105 (1987).

⁵⁹ See Booth, *supra* note 57, at §9.2.

⁶⁰ See Morrissey, *supra* note 43, at 677-78.

⁶¹ See *id.*

⁶² See Booth, *supra* note 57, at §9.2.

⁶³ *Id.*; see also James E. Ballowe, Jr. & Penelope Y. F. Tham, *State Blue Sky Regulation*, 816 PLI/CORP 433, 435 (1993).

⁶⁴ Booth, *supra* note 57, at §9.2.

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.*

⁶⁸ See *id.* “As a general rule, commissions in excess of 15 percent will not pass muster in a merit regulation state.” *Id.*

⁶⁹ *Id.*; see also Ballowe & Tham, *supra* note 63, at 443-46.

⁷⁰ Karmel, *supra* note 58, at 116.

⁷¹ See Ballowe & Tham, *supra* note 63, at 437.

offering.”⁷² As such, the merit regular has broad authority to demand changes to a corporation or its proposed offering before allowing a securities offering to proceed.

Further still, most merit-regulation jurisdictions impose licensing requirements on securities broker-dealers, and ban offerings by issuers who officers have run afoul the law.⁷³

As with the disclosure regime, merit regulation has its detractors. The screening of proposed securities offerings by government officials “clearly imposes burdens on capital formation.”⁷⁴ Not only must securities issuers convince the capital markets of their offering’s worthiness, but they have the added hurdle (and costs) of having to convince a regulator as well. With that hurdle comes the risk of regulator error: the prospect that a quality offering will be wrongfully rejected.⁷⁵ The specter of corruption and bias in the process must also be considered.⁷⁶

The paternalistic philosophy of merit regulation is also challenged.⁷⁷ Even if such paternalism could have at one time been justified, it cannot be so today (so goes the argument).⁷⁸ For merit regulation arrived on the scene during the era of the individual investor, but today is the era of the institutional and professional investor.⁷⁹ Today’s investors (many of them at least) may very well be more sophisticated than the regulators themselves. It seems like folly, therefore, to allow a regulator to dictate what offerings an investor can or cannot partake in from some consumer-protection rationale focused on fairness and appropriate risk.⁸⁰

Standing in obvious response to the anti-paternalism argument is the simple fact that even large, sophisticated, institutional investors made a series of tremendously bad investments in the run-up to the 2008 financial crisis.⁸¹ But in an important respect, this criticism misses the mark. The critics of merit regulation’s paternalism do not claim that the twenty-first century investor is infallible. Rather, these critics query whether a government regulator would be any less fallible.⁸² The point is a good one, and not to be dismissed lightly. There is no reason to believe that the best-and-the-brightest financial experts are inevitably drawn to Washington, D.C. In fact, given the substantial draw of lucrative private sector salaries, it is difficult to see why this would be the case. At best, therefore, one could hope that government regulators are on par with those whom they regulate in terms of expertise and ability.

There is, however, at least one important factor that may help level the playing field – or, perhaps, give the government’s regulators an advantage: access to information. Quality information is, arguably, one of the most precious commodities in our modern world – and especially so within the financial services industry and the capital markets. Armed with the new powers and tools by Dodd-Frank Act,⁸³ today’s regulators undoubtedly have greater access to critical information than their counterparts in private industry. Whereas a top flight private analyst has access to abundant publicly-available information, a government regulator has access to all that and much more: to nonpublic information procured via subpoena and other means.

Another line of criticism against merit regulation is that it failed to prevent the crisis of 1929.⁸⁴ After

⁷² Karmel, *supra* note 58, at 116.

⁷³ See Ballowe & Tham, *supra* note 63, at 441-42.

⁷⁴ *Id.*, at 106. *But see* Manning Gilbert Warren III, *Legitimacy in the Securities Industry: The Role of Merit Regulation*, 53 BROOK. L. REV. 129, 140 (1987) (“The argument that capital formation is impeded by state securities regulation, and, more particularly, merit regulation is ... highly suspect.”).

⁷⁵ *But see* Jay T. Brandt, *The Silver Lining in Blue Sky Laws: The Effect of Merit Regulation on Common Stock Returns and Market Efficiency*, 12 J. CORP. L. 713, 734 (1987) (finding that “the criticism regarding market inefficiency due to merit restrictions may be unwarranted”).

⁷⁶ On a related note, corruption in the traditional process of corporate chartering (which, like the merit regulation of securities, was done on a company-by-company basis) is what led, in part, to the promulgation of statutes of general incorporation (depriving state regulators of discretion in the granting or denial of corporate charters). Gregory A. Mark, *The Court And The Corporation: Jurisprudence, Localism, And Federalism*, 1997 SUP. CT. REV. 403, 414 (1997).

⁷⁷ Ballowe & Tham, *supra* note 63, at 448-49.

⁷⁸ *See id.* at 448-49.

⁷⁹ *See id.*

⁸⁰ *See id.*

⁸¹ Morrissey, *supra* note 43, at 681.

⁸² Rutheford B. Campbell, Jr., *An Open Attack on the Nonsense of Blue Sky Regulation*, 10 J. CORP. L. 553, 565-67 (1985); Mark A. Sargent, *Report on State Merit Regulation of Securities Offerings*, 41 BUS. LAW. 785, 853 (1986).

⁸³ Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 171(b)(4), 124 Stat. 1376 (2010); see So-Yeon Lee, *Financial Stability Oversight Council*, 30 REV. BANKING & FIN. L. 528, 534-35 (2011).

⁸⁴ Joel Seligman, *The Obsolescence of Wall Street: A Contextual Approach to the Evolving Structure of Federal Securities Regulation*, 93 MICH. L. REV. 649, 673 (1995).

all, by the time of the crisis, most states had adopted some form of merit regulation.⁸⁵ However, the crisis of 1929 was national in scope, and as such arguably beyond the realistic reach of state regulators to prevent or contain.⁸⁶

Perhaps the most forceful critique of merit regulation – on a federal level, at least – is its simple unworkability. It is difficult to imagine how the SEC, or any federal agency, could effectively provide a merit review of each and every proposed securities offering. An enormous increase in staffing and resources would be required – and even that fails to guarantee whether sufficient expertise would be available to adequately analyze the volume of offerings expeditiously enough.

3. Effectuating Merit Regulation via the Suitability Rules

A. The Suitability Rules

The “suitability rules” require that a broker, when recommending a securities transaction to a customer, does so based upon the informed belief that the transaction is “suitable” for the customer.⁸⁷ This requirement is not imposed by the federal securities laws, but rather by the securities exchanges themselves.⁸⁸ For brokers of the New York Stock Exchange, the relevant rule is FINRA Rule 2111.⁸⁹ Although not absolutely identical, the suitability rules adopted by all other securities exchanges in the United States substantially follow FINRA Rule 2111.⁹⁰

Rule 2111 reads, in its entirety, as follows:

(a) A member or an associated person must have a reasonable basis to believe that a recommended transaction or investment strategy involving a security or securities is suitable for the customer, based on the information obtained through the reasonable diligence of the member or associated person to ascertain the customer's investment profile. A customer's investment profile includes, but is not limited to, the customer's age, other investments, financial situation and needs, tax status, investment objectives, investment experience, investment time horizon, liquidity needs, risk tolerance, and any other information the customer may disclose to the member or associated person in connection with such recommendation.

(b) A member or associated person fulfills the customer-specific suitability obligation for an institutional account, as defined in Rule 4512(c), if (1) the member or associated person has a reasonable basis to believe that the institutional customer is capable of evaluating investment risks independently, both in general and with regard to particular transactions and investment strategies involving a security or securities and (2) the institutional customer affirmatively indicates that it is exercising independent judgment in evaluating the member's or associated person's recommendations. Where an institutional customer has delegated decisionmaking authority to an agent, such as an investment adviser or a bank trust department, these factors shall be applied to the agent.⁹¹

Rule 2111 is frequently, and advisably, read in conjunction with FINRA Rule 2090 – FINRA's “know your

⁸⁵ See *supra* text accompanying note 61.

⁸⁶ James E. Ballowe, Jr. & Penelope Y. F. Tham, *State Blue Sky Merit Regulation*, 816 PLI/Corp 433, 437 (1995); Thomas E. Geyer, *Viewing the Columbus Skyline: Incorporating Federal Law into the Anti-Fraud Standard of the Ohio Securities Act*, 28 U. Tol. L. Rev. 301, 305-06 (1997).

⁸⁷ See 1 CLIFFORD E. KIRSCH, *BROKER-DEALER REGULATION* §11:1.1 (2D ED. 2012).

⁸⁸ See *id.* That said, courts have held that a violation of the suitability rules can also constitute a violation of Rule 10b-5, the anti-fraud provision promulgated by the SEC under the 1934 Securities Exchange Act. See *id.*

⁸⁹ See FINRA Rule 2111 (available at http://finra.complinet.com/en/display/display.html?rbid=2403&record_id=13390&element_id=9859&highlight=2111#r13390).

⁹⁰ See Jennifer A. Frederick, *Not Just For Widows & Orphans Anymore: The Inadequacy Of The Current Suitability Rules For The Derivatives Market*, 64 FORDHAM L. REV. 97, 110 n.79 (1995). It should be noted that if a broker is primarily in the business of rendering investment advice, then the broker would instead be regulated by the Investment Advisors Act of 1940. See TREATISE ON THE LAW OF SECURITIES REGULATION, 5 Law Sec. Reg. § 14.16[1] (2012).

⁹¹ See FINRA Rule 2111 (available at http://finra.complinet.com/en/display/display.html?rbid=2403&record_id=13390&element_id=9859&highlight=2111#r13390).

customer” rule.⁹² This short rule reads, in its entirety, as follows:

Every member shall use reasonable diligence, in regard to the opening and maintenance of every account, to know (and retain) the essential facts concerning every customer and concerning the authority of each person acting on behalf of such customer.⁹³

Thus, suitability requires a broker to comprehend his or her client’s financial situation (“investment profile”) via the use of “reasonable diligence.” Further, the broker must restrict his or her investment suggestions to those that are “suitable” to the client in light of the client’s situation. Observe that what may or may not be suitable is not wholly determined by the client’s own wishes, but rather takes into account certain objective factors (“the customer’s age, other investments, financial situation and needs, tax status, investment objectives, investment experience, investment time horizon, liquidity needs, risk tolerance”) that go well beyond a client’s expressions of interest or desire. The broker is, therefore, expected to exercise a large degree of independent, professional judgment in determining whether to recommend a specific investment as suitable.

Although it is not clear from the Rule’s text itself, the broker must (of course) also be well-versed in the security, and/or the security transaction, that he or she is recommending.⁹⁴

Not surprisingly, a client’s wealth and income are not dispositive factors. FINRA has held that a dealer “may not rely exclusively on the client’s status as an accredited investor under Regulation D of the Securities Act for satisfying suitability obligations,” because this status alone does not adequately inform the entire suitability analysis.⁹⁵ Further, the suitability requirement does not automatically vanish when a broker’s client happens to be an institutional investor. As Rule 2111(b) explains, even in this situation the broker has important suitability obligations. More specifically, the broker must be satisfied that the institutional investor is capable of adequately evaluating the transaction in question, and the institutional investor must affirmatively declare that it is indeed exercising independent judgment with regard to the transaction. (If the institutional investor is acting through an investment advisor, or the equivalent, then those factors are applied to that advisor.⁹⁶)

B. The Federalization of Suitability to Achieve Merit Regulation

The suitability rules are well-positioned to serve as a means by which the benefits of merit regulation can, to a significant degree, be realized. Moreover, this realization can occur without the more serious drawbacks that often accompany merit-based regulation, as addressed previously.⁹⁷

As currently formulated, suitability rules (such as FINRA Rule 2111) vest a tremendous degree of discretion in the hands of brokers. But this need not be the case. The SEC could adopt guidance and mandates that would help brokers define certain investments as simply “suitable” or “unsuitable” for certain classes of investors defined by regulation.⁹⁸ In other words, the SEC could federalize the concept of suitability.

In so doing, the SEC would not be mandating regulatory approval for each and every proposed securities offering (along the lines of typical blue-sky merit regulation). Instead, the SEC would simply be promulgating broad guidelines that brokers would need to incorporate into their suitability analysis. From purely quantitative metrics such as earnings-to-price ratios, to more qualitative ones such as the issuer’s particular industry, the SEC could promulgate standards against which a particular security’s level of risk could be assessed. Based upon these standards, securities could be classified, by brokers, into particular “risk-bearing” categories themselves defined by the SEC. Thus, a security could be labeled “high risk” if,

⁹² See Kirsch, *supra* note 87, at §11:1.1.

⁹³ See FINRA Rule 2090 (available at http://finra.complinet.com/en/display/display.html?rbid=2403&record_id=13389&element_id=9858&highlight=2090#r13389).

⁹⁴ See TREATISE ON THE LAW OF SECURITIES REGULATION, 5 LAW SEC. REG. § 14.16[1] (2012).

⁹⁵ Kirsch, *supra* note 87, at §11:1.3.

⁹⁶ See *id.*

⁹⁷ See *supra* notes 74-86.

⁹⁸ This would not be entirely unprecedented, as the SEC has adopted its own suitability rule with regard to the trading of “penny stocks.” See 1933 Securities Act Rule 15g-9.

upon an examination of its particular characteristics, including those set forth as relevant by the SEC, the security possesses a great deal of investment risk.

Of course, brokers, advisors, and industry analysts already do this, and one might question the utility of the SEC's guidelines, deeming them duplicative. But for at least two inter-related reasons, this is not the case.

First, in the aftermath of the Dodd-Frank Act, the SEC has broader access to non-public information than ever before.⁹⁹ Via use of the powers contained in the Act, the federal government can acquire information deemed critical to the nation's economic health, and share that information with agencies (such as the SEC) situated to act upon it. Thus, in promulgating its standards regarding an investment security's risk, the SEC could draw upon this information, and provide guidance that could not otherwise be replicated by private industry professionals. Thus, the risk-bearing classification would be based upon standards that are derived, in part, from important nonpublic information.

Second, the SEC's guidelines would reflect public policy determinations, whereas wholly private risk assessments ordinarily do not. As will be fleshed out momentarily below, depending upon how the risk standards are calibrated, the SEC's guidelines could effectively place certain securities off-limits for certain groups of investors. Moreover, the SEC's guidelines may consciously do this to implement a policy of keeping particular types of securities out of the hands of certain types of investors.¹⁰⁰

Which brings us to the second prong of the approach: a classification of investors based upon particular, relevant characteristics. The SEC could promulgate an investor classification schema, pursuant to which investors could be placed into particular categories based upon their degree to tolerate and/or withstand risk. This too, is already done by brokers and other securities professionals, which once again begs the question: where is the value added by the SEC's roll-out of such a classification?

Unlike the classifications employed in the private market, the SEC's classification system need not be a simple exercise in lining up groups of investors, in order, between the poles "non-wealthy" and "wealthy" (with the assumption that the more wealthy the investor, generally speaking, the greater his, her, or its ability to withstand risk). Instead, the SEC's classification schema could (and should) take into account the principles embodied in the Dodd-Frank Act (and banking regulation in general) – namely, that certain entities are too systemically important to assume excessive levels of risk. Thus, the investor classification scheme should take into account the fact that certain investors, despite a high level of sophistication, and a substantial degree of wealth, are nevertheless placed into a lower-risk-tolerance classification on account of their systemic importance.

Finally, the SEC would furnish a "suitability matrix," which would tie these two variables together (a security's "risk bearing" classification, and an investor's "risk tolerance" classification). This matrix would identify which securities were suitable for which investors.

The matrix would not be a simple matching of "high risk" bearing securities with "high risk" tolerant investors, and so forth. Rather, it envisions a more complicated classification scheme, in which an investor's "risk tolerance" and a securities "risk bearing" classifications are not simple, one-dimensional characteristics. Rather, these classifications would themselves vary depending upon context. That is, the classifications would interact with one another, and thereby be further delineated. For example, a particular security could be deemed a "low risk" bearing instrument generally, but nevertheless "high risk" for a particular category (or categories) of investor. Or, from the opposite angle, an investor could be generally deemed to have a "high risk" level of tolerance, but that risk level drops to "low" or "intolerant" when confronted with securities of a particular type or bearing particular characteristics (related, perhaps, to an industry or some other factor that

⁹⁹ See *supra* note 83 and accompanying text.

¹⁰⁰ It should be noted that the SEC could take a more aggressive approach toward this same end. Instead of simply promulgating guidelines and standards, the SEC could take the next step and actually categorize each securities offering itself in accordance with these guidelines and standards. That is, the SEC could classify each offered security into a pre-established risk category, and make such classification binding upon the securities industry. This would come very close to implementing a federal blue-sky approach. However, there would be at least one significant distinction. Unlike typical blue-sky regulation, the SEC would not be authorized to block the underwriting of a securities offering based upon its substantive risk. Instead, the SEC would be limited to simply assigning the offering a risk classification which, as explained more fully below, would be cross-referenced with an investor's status and situation to determine the security's suitability. Such an aggressive approach would, however, require a tremendous amount of resources, and as such undermine one of the major benefits of the proposal over existing versions of blue-sky regulation, as discussed below.

would be singularly relevant to the investor classification in question).

Of course, not every possible interaction could be foreseen – and there are only so many categories into which securities and investors could be placed. The SEC's guidelines and matrix would not need to, nor purport to cover the entire field of potential securities transactions. Rather, the SEC would attempt to cover most of the field and, moreover, cover those transactions which the SEC believes most seriously need to be covered.

As ambitious as such an approach might initially seem, it does to a large extent build upon existing regulatory infrastructure and practices. At its heart lies the well-established suitability rules. The proposal essentially does little more than augment this privately administered system for regulating risk with federal guidelines and definitions.

Further, the prospect of classifying investors and securities is not entirely new to the SEC. For example, the SEC already classifies investors into certain categories, via its “accredited investor,” and “QIB” statuses.¹⁰¹ Of course, this is not as detailed a classification as the one proposed above (nor as comprehensive). Further, the “accredited investor” and “QIB” classifications are employed to ascertain the degree to which an investor may partake in the sale of a non-registered security – they do not purport to accurately assess the degree to which a particular investor can handle investment risk. Nevertheless, their use does underscore the point that classifying investors is not alien to U.S. securities laws.

Additionally, an argument can be made that the SEC has already adopted a securities classification scheme – albeit indirectly. For ever since the regulatory reforms of 2005, the SEC has classified *issuers* into the following four categories: “(1) the well-known seasoned issuer, (2) the seasoned issuer, (3) the unseasoned Exchange Act reporting issuer, and (4) the non-reporting issuer.”¹⁰² These issuer classifications are largely driven by the degree to which a particular issue is publicly known and widely followed in the financial markets.¹⁰³ Their purpose is to assess the amount of disclosure deemed necessary in such issuer's public offering of securities.¹⁰⁴ Thus, the securities of some types of issuers (such as “non-reporting issuers”) will require greater disclosure in a public offering than the securities of others types of issuers (such as “well-known seasoned issuers”).¹⁰⁵ This purpose differs from the proposed classification schema proposed above, which is not to determine the amount of disclosure necessary in a public offering, but rather to categorize the substantive riskiness of the security as an investment. That said, the practice of categorizing issuers as per the 2005 reforms furnishes a precedent pursuant to which the SEC has determined to treat different securities differently.

Further still, U.S. securities laws have long recognized a distinction between securities that are registered versus securities that are unregistered.¹⁰⁶ Depending upon where a security falls within this dichotomy, the entire regulatory disclosure regime may or may not apply to it.¹⁰⁷ Moreover, and more relevantly, their classification in this regard even determines which investors may purchase the security.¹⁰⁸

C. The Potential Benefits of Merit Regulation via the Suitability Rules

The proposal set forth above would have the SEC play a substantial (if not a leading) role in the assessment of suitability. The classification of both securities, and investors, would be made in accordance with SEC guidelines. Even the ultimate question of suitability would itself be governed by SEC rules, pursuant to which certain risk-classifications of securities would be deemed unsuitable (and thereby off-limits) to certain categories of investors. Effectuating merit regulation via the suitability rules, as proposed here, could offer significant benefits without the costs that usually accompany blue-sky-type merit regulation.

The primary benefit is that, as already noted, such an approach would allow the federal government and the SEC to bring its resources to bear upon the question of investment risk. These resources include

¹⁰¹ See William K. Sjoström, Jr., *The Birth Of Rule 144a Equity Offerings*, 56 UCLA L. Rev. 409, 443-44 (2008).

¹⁰² See Joseph F. Morrissey, *Rhetoric And Reality: Investor Protection And The Securities Regulation Reform Of 2005*, 56 CATH. U. L. REV. 561, 574 (2007).

¹⁰³ See *id.*

¹⁰⁴ See *id.*

¹⁰⁵ See *id.*

¹⁰⁶ See Stephen J. Choi, *A Framework For The Regulation Of Securities Market Intermediaries*, 1 BERKELEY BUS. L.J. 45, 79 (2004).

¹⁰⁷ See *id.*

¹⁰⁸ See *id.*; see also *supra* note 101.

increased access to vital information under the provisions contained in the Dodd-Frank Act.¹⁰⁹ This information, in conjunction with the information and analysis already publicly available via the work of research analysts and other industry professionals, should yield a more accurate assessment and classification of the riskiness of securities.

But unlike the typical approach taken by merit regulators at the state level, this proposal would not be nearly as resource intensive – for at least two reasons. First, under this proposal the SEC would be limiting itself to promulgating rules and standards for use in assessing suitability – in terms of both security/issuer evaluation and investor classification. This varies from the typical approach of merit regulation, pursuant to which the regulator would pass judgment upon securities offerings on an offering-by-offering basis. Second, the SEC would not even be applying its rules and standards governing suitability – securities brokers would, thus passing on this cost to them.

An added benefit of this approach is the fine-tuning that it enables – another departure from traditional merit regulation, which possesses a certain all-or-nothing quality. That is, the SEC would not be designating certain securities offerings as simply “off limits” to *all* investors, but rather would be effectively placing certain securities off limits to only certain groups of investors (due to the interaction of the security's risk assessment under the SEC's guidelines, and the investor's risk tolerance, also determined under SEC guidelines). This also puts the SEC in a particularly good position to implement public policy regarding both investor protection and issues of systemic risk.

Further fine-tuning would be possible if the SEC were to permit investors to petition for exemption to the applicability of its suitability rules and standards on a case-by-case basis. For example, although SEC standards might deem a particular security unsuitable for a particular class of investor with a given classification, that investor could be permitted an opportunity to explain why its situation does not call for application of the governing standard.

A third improvement that the proposal features over traditional merit regulation is that it reaches secondary market trading in addition to primary market offerings. The typical merit regulator assesses the merits of a security at the time of its offering to the public.¹¹⁰ Once the security is trading in the marketplace, among investors, there is ordinarily little or no further regulation of the security's merit.¹¹¹ By contrast, the instant proposal would establish rules, standards, and guidelines that would apply to every securities transaction involving a broker. Thus, it would not only govern the transaction between an issuer and investor, but would also ensure that trades between investors were suitable and consistent with public policy.

Conclusion

The financial crisis of 2008 has caused many to revisit the merits of merit regulation. The benefits of such an approach, in terms of more robust investor protection, and another tool to address systemic risk, might be as appealing as ever. But so are the approach's drawbacks, including its resource-intensiveness and the risk that certain offerings might be locked out of the capital markets altogether on the basis of a regulator's judgment (which could be erroneous or, worse, biased). Federal commandeering of the suitability rules, along the lines outlined here, offer a means by which the benefits of merit regulation can be largely obtained without many of the costs that the usual accompany such an approach.

Admittedly, much remains to be considered with this proposal, and many questions remain unanswered. As drafted, the proposal only applies to broker's transactions in which the suitability rules are implicated. As such, a large number of transactions, both of individual investors and institutional investors, would not fall under its umbrella. To have its intended effect, this proposal would need to be replicated as necessary to cover such transactions as well.

Thus, as is readily apparent, this proposal is simply an opening foray. It represents an effort to address a critical problem, and a problematic solution, by sketching a course of action that is hopefully

¹⁰⁹ See *supra* note 83 and accompanying text.

¹¹⁰ See Thomas E. Geyer, *Viewing The Columbus Skyline: Incorporating Federal Law Into The Anti-Fraud Standard Of The Ohio Securities Act*, 28 U. Tol. L. Rev. 301, 306 (1997).

¹¹¹ See *id.*

creative but not unrealistic, modest but not meaningless. My hope is that to the extent it resonates, others will add flesh to its bones and build upon whatever elements are deemed of value.

References

- Altschuller, S.A., Lehr, A.K & Orsmond, A.J., *Corporate Social Responsibility*, 45 INT'L LAW. 179, 183-84 (2011).
- Ballowe, Jr. J.E. & Tham, P.Y.F. *State Blue Sky Regulation*, 816 PLI/CORP 433, 435 (1993).
- Ballowe, Jr., J.E. & Tham, P.Y.F. *State Blue Sky Merit Regulation*, 816 PLI/Corp 433, 437 (1995)
- Booth, R.A. Financing the Corporation §9.13 (2012).
- Braisted, C.H.B. *State Registration of Securities: An Anachronism No Longer Viable*, 78 WASH. U. L. Q. 401, 405 (2000).
- Brandi, J.T. *The Silver Lining in Blue Sky Laws: The Effect of Merit Regulation on Common Stock Returns and Market Efficiency*, 12 J. CORP. L. 713, 734 (1987)
- Campbell, Jr., R.B. *An Open Attack on the Nonsense of Blue Sky Regulation*, 10 J. CORP. L. 553, 565-67 (1985);
- Choi, S.J. *A Framework For The Regulation Of Securities Market Intermediaries*, 1 BERKELEY BUS. L.J. 45, 79 (2004).
- Colombo, R.J. *Buy, Sell, or Hold?*, 73 BROOK. L. REV. 91, 119-120 (2007).
- Douglas, W.O. *Protecting the Investor*, 23 YALE L. REV. 521, 522-23 (1934)).
- Frederick, J.A. *Not Just For Widows & Orphans Anymore: The Inadequacy Of The Current Suitability Rules For The Derivatives Market*, 64 FORDHAM L. REV. 97, 110 n.79 (1995).
- Geyer, T.E. *Viewing the Columbus Skyline: Incorporating Federal Law into the Anti-Fraud Standard of the Ohio Securities Act*, 28 U. TOL. L. REV. 301, 305-06 (1997).
- Geyer, T.E. *Viewing The Columbus Skyline: Incorporating Federal Law Into The Anti-Fraud Standard Of The Ohio Securities Act*, 28 U. TOL. L. REV. 301, 306 (1997).
- Hui Huang, R. *The Regulation of Securities Offerings in China: Reconsidering the Merit Review Element in Light of the Global Financial Crisis*, 41 HONG KONG L. J. 261, 261 (2011);
- Johnson, J.J *Private Placements: A Regulatory Black Hole*, 35 DEL. J. CORP. L. 151, 156 (2010).
- Karmel, R.S. *Blue-Sky Merit Regulation: Benefit to Investors of Burden on Commerce?* 53 BROOK. L. REV. 105, 105 (1987).
- Kirsch, C.E. BROKER-DEALER REGULATION §11:1.1 (2D ED. 2012).
- Loss, L., Seligman, J. & Paredes, T. 1 FUNDAMENTALS OF SECURITIES REGULATION 45-46 (2011).
- Macey J.R. & Miller, G. P. *Origin of the Blue Sky Laws*, 70 TEX. L. REV. 347, 348 (1991).
- MacLeod Heminway, J. *The Best Of Times, The Worst Of Times: Securities Regulation Scholarship And Teaching In The Global Financial Crisis*, 5 J. BUS. & TECH. L. 59, 68 (2011).
- Mark, G.A. *The Court And The Corporation: Jurisprudence, Localism, And Federalism*, 1997 SUP. CT. REV. 403, 414 (1997).
- Morrissey, D. J. *The Road Not Taken: Rethinking Securities Regulation and the Case for Federal Merit Review*, 44 U. RICH. L. REV. 647, 649 (2010).
- Morrissey, J.F. *Rhetoric And Reality: Investor Protection And The Securities Regulation Reform Of 2005*, 56 CATH. U. L. REV. 561, 574 (2007).
- Ripken, S.K. *The Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, 58 BAYLOR L. REV. 139, 144 (2006).

Sargent, M.A. *Report on State Merit Regulation of Securities Offerings*, 41 BUS. LAW. 785, 853 (1986).

Schwarcz, S.L. *Rethinking the Disclosure Paradigm in a World of Complexity*, 2004 U. ILL. L. REV. 1, 12 (2004).

Seligman, J. *The Obsolescence of Wall Street: A Contextual Approach to the Evolving Structure of Federal Securities Regulation*, 93 MICH. L. REV. 649, 673 (1995).

Sjostrom, Jr, W.K. *The Birth Of Rule 144a Equity Offerings*, 56 UCLA L. Rev. 409, 443-44 (2008).

So-Yeon Lee, *Financial Stability Oversight Council*, 30 REV. BANKING & FIN. L. 528, 534-35 (2011).
TREATISE ON THE LAW OF SECURITIES REGULATION, 5 Law Sec. Reg. § 14.16[1] (2012).

Warren III, M.G. *Legitimacy in the Securities Industry: The Role of Merit Regulation*, 53 BROOK. L. REV. 129, 140 (1987)

Basel III and the Strengthening of Capital Requirement:

The obstinacy in mistake or why “it” will happen again¹

Bernard Vallageas^{2, 3}

Collège d'Etudes Interdisciplinaires, Faculté Jean-Monnet, Université Paris-Sud, France

Abstract

Since the financial liberalisation of the 80's, the Basel committee on Banking Supervision wants to strengthen banks' capital and other stable funding with the purpose of increasing banks' financial security, as stable funding increases the security of the non-financial business sector. But bank capital has nothing to do with bank security and with money creation. It is shown, instead, that increasing banks' stable funding entails a decreasing of the stable funding for the rest of the economy and securitization. Thus this strengthening is harmful and the way for the financial sector to work in the interest of the economy is to separate deposit banks and other financial institutions, to strengthen banks' control and to recognize they do not need capital, and therefore no owners.

Résumé

Depuis la libéralisation des années 80, le comité de Bâle veut renforcer le capital des banques et leurs autres ressources permanentes dans le but d'accroître leur sécurité financière, tout comme l'accroissement de leurs capitaux permanents renforce la sécurité financière des entreprises industrielles. Mais le capital d'une banque n'a aucun rôle dans sa sécurité ni dans la création monétaire. Au contraire nous montrons que l'accroissement des capitaux permanents des banques entraîne une décroissance de ceux des entreprises industrielles et de la titrisation. Aussi ce renforcement est nuisible et pour que le secteur financier travaille dans l'intérêt de toute l'économie, il faut séparer les banques de dépôt des autres institutions financières, renforcer le contrôle des banques et reconnaître qu'elles n'ont pas besoin de capital et donc pas besoin de propriétaires.

JEL: E42 E44 E58 G32 G38

Key words: Bank capital, bank ratios, deposit banks, Basel accord

Introduction

On September 6, 2009, after the 2007-2009 financial crisis, the deepest since 1929, the Group of Central Bank Governors and Heads of Supervision, chaired by the French president of the European Central Bank, Jean-Claude Trichet, met at Basel in the headquarters of the Bank for International Settlements (BIS 2009b). The Group decided to strengthen the capital requirement of banks. However, as Bardoloi (2003)

¹ Cf. Minsky 1982.

² Preliminary drafts of this paper have been presented at the workshop on the crisis organised by the Association pour le Développement des Etudes Keynésiennes and the Collège d'Etudes Interdisciplinaires at Université Paris-Sud, on December, 16, 2009, at the international conference “*Questioning the Crisis and Prospects for Change*” organised by the Action COST IS0902, Systemic Risks, Financial Crises and Credit, with support of the Centre d'Economie de l'Université Paris-Nord and of the Laboratoire d'Economie Dyonisien (Université Paris 8) on October 27, 2010 at the Maison des Sciences de l'Homme Paris-Nord, at the seminar organized by the Centre Lillois d'Etudes et de Recherches Sociologiques et Economiques (Université des Sciences et Techniques de Lille) on March 24, 2012, and at the international conference “*Political Economy and the Outlook for Capitalism*” organized by the Association for Heterodox Economics, the International Initiative for Promoting Political Economy, the French Association of Political Economy and the Association pour le Développement de l'Economie Keynésienne at Université Paris 1 on July 5- 7, 2012.. The author thanks the participants in these conferences for their observations and Mario Seccareccia for his English rewriting. He takes sole responsibility for any remaining mistakes.

³ This paper was first written for the book *Monetary economies of production: banking and financial circuits and the role of the State. Essays in honour of Alain Parguez, Rochon, L.-P. and Seccareccia, M.* (eds), Cheltenham: Edward Elgar. We wish to acknowledge the editors and the publisher for permission to print it here.

affirmed, that is only new wine in an old bottle. The old bottle (called "Basel I"), the institution of capital requirement for banks, had been invented in July 1988 (BIS 1998, BIS 2009a), some months after the October 1987 crash, by the group of central bank governors; and the new wine (called "Basel III") is the strengthening of this requirement: more quality, consistency and transparency of the tier 1 capital, capital conservation measures such as constraints on capital distributions, increase of the leverage ratio, introduction of a counter-cyclical capital buffer... After the September 6, 2009 meeting, the Basel Committee on Banking Supervision published two consultative documents (BIS 2009c,d) on the improvement of the capital requirement and launched a consultation process of the financial actors. In December 2010, it published two documents (BIS 2010a,b), the new wine, to implement this improvement.

All of these measures concern banks' capital, but before 1988 no one worried about banks' capital, and nevertheless there had occurred no major crisis since the Great Depression. Curiously the adoption of capital requirements had taken place at the same time as the liberalisation of the financial system and, yet, despite their adoption, they did not prevent financial crises, particularly the major 2007-2009 crisis. Moreover, the strengthening of these capital requirements was part of the bank rescue package of 2008 (Vallageas 2009).

The idea we want to develop in this paper is that capital requirements may be a good tool for non-financial business enterprises, but they are not appropriate for banks, and, if excessive, these requirements may even prevent the banking system from working correctly and could create dysfunctions in the real economy. Most economists, even heterodox colleagues⁴, consider that the strengthening of bank capital is a good thing, even if other regulations are necessary, and that, in any case, it cannot do wrong. Our idea is that it does do wrong. The decline of Keynesianism since the 80's has reinforced the neoclassical idea that money is neutral, that everything works as if money did not exist, so that banks are financial intermediaries that borrow money, but do not create it. Therefore the distinction between deposit banks, normally described as commercial banks, which create money, and other financial entities, which do not, has been abolished (e.g. in France by the law of January 24, 1984 and in the U.S. by the Gramm-Leach-Bliley Act of November 12, 1999 reforming the Glass Steagall Act of 1933). Finally banks are considered like industrial firms, they must be free, they ought not to be regulated, and they must have capital. As it is for industrial enterprises, capital would be tool to ensure security, and thus it should be reinforced. But that is wrong; banks do not behave like industrial firms.

The advocates of capital requirement have forgotten that deposit banks create money and, hence, in a first section, we will review how these banks create money and show why capital has no role in that creation.

In a second section we will show that two new ratios introduced as the new wine in the two December 2010 documents (BIS 2010a,b) are a very bad wine since they reveal an absolute misunderstanding of the way the banking system works and a total confusion between deposit banks and other financial entities. The adoption by financial entities of these two ratios, conceived originally for non-financial business firms, will worsen the financial situation of these latter firms. Their adoption will also give rise to an increase of securitization which will entail dysfunctions. Thus the strengthening of bank capital will do wrong. In a third and last section, we shall consider where bank capital comes from and what is its purpose, since it is neither for money creation nor for deposit security.

In the conclusion, we will outline some ideas for a reform of deposit banks.

1 Deposit banks use reserves of central bank money, and not their capital, to create money

Money creation is neither considered by orthodox theory nor by the originators of capital requirements. These original advocates only considered bank capital because they thought that it secures bank liabilities, just as industrial firms' capital secures industrial firms' liabilities. But, contrary to this belief, bank capital does not secure bank liabilities. To understand this point we must study a bank balance sheet and concentrate on the specificity of banks in the creation of money. Indeed only deposit banks, i.e. banks that receive demand

⁴With the notable exception of Toporowski (2009)

deposits, create money, and we shall focus on these institutions. First we shall compare deposit banks' balance sheets and industrial firms' balance sheets, and then we shall study how banks create money in a metallic money regime and in a modern regime of non-convertible notes.

1.1 The balance sheets of an industrial firm and of a deposit bank

We present (Figure 1) the balance sheet of an industrial firm as it is normally circulated at an annual meeting of shareholders before the allocation of profit that goes toward the payment of dividends. We remind the reader of the meaning of the different items. The capital measures the value of the shareholders' contributions. These contributions are assets, which, whatever their form, money or other, appear in the assets column and whose value is registered for the same amount in the capital account, because of the double-entry system. The reserves measure that part of the assets which are truly owned by the shareholders but which are not included in the value of capital (e.g. the retained earnings). The aggregate of capital and reserves constitutes the shareholders' equity (in red) and measures the part of the assets which truly belongs to the shareholders, the other assets being financed by debts. Therefore, the first function of shareholders' equity (in red) is to finance a part of the assets.

Figure 1 Balance-sheet of an industrial firm before allocation of profit

fixed assets	capital
	reserves
stocks	debts
current claims	
cash	profit

Its second function is to secure the debts, since the creditors can seize the assets belonging to the shareholders. It is this last function, that the proponents of capital requirement want to apply to banks. By doing so they think wrongly that they will secure banks' liabilities.

Indeed, as a bank is legally a firm, its balance sheet looks similar to that of an industrial firm (see Figure 2).

Figure 2 Balance sheet of a deposit bank

account at the central bank and notes <small>(reserve of central money)</small>	demand and term accounts
loans	
titles representative of loans	titles representative of debts
	reserves
tangible assets	capital

However, it is convention to set it upside down, i.e. equity and tangible assets are written at the bottom. In addition to this difference of presentation, there are two peculiarities: 1. tangible assets are very small (the only machines that a bank needs are computers and it does not need raw materials); 2. among the financial assets one must distinguish the account at the central bank (in green) called by monetary economists the reserve of central bank money. This account is not peculiar to banks, since any firm may in principle have an account at the central bank, although it is not common institutionally in Western countries.

So we see that banks have a balance sheet similar to that of other firms and the two peculiarities do not come from law or accounting principles but from the true nature of deposit banks.

It is these two peculiarities that the advocates of banks' capital requirement have neglected to consider, by confusing reserves of central bank money (in green) and reserves in the purely accounting sense (in red). By doing so, they have forgotten the necessity of the central bank.

To develop this point, we must first consider the principle of monetary creation by a deposit bank. We will distinguish two categories of monetary systems: 1. the metallic currency system; 2. the modern monetary system in which the central bank is absolutely necessary to allow the interbank payments and settlements.

1.2 The monetary creation by a deposit bank

a) Under a metal currency regime

In this system, some people deposit metal (say gold) or bank notes convertible into this metal at the deposit bank (item in green on the balance sheet, Figure 3). This gold or notes are written on the asset side of the balance sheet, but, as they are not the property of the bank, a debt of a same amount (called the initial debt) is written on the liability side. Then, as most deposits are stable, the bank takes the risk to lend more than it has received, and these deposits then appear as supplementary loans and supplementary debts. The deposits to the bank come from ordinary people and not from the central bank. So the system may work only with gold, without notes and a central bank. The initial deposits (in green on Fig. 3) are called reserves by monetary theorists and banks must wait for them before lending. Thus the reserves appear *ex ante*.

Figure 3 Balance sheet of a deposit bank under a metal currency regime

supplementary loans	supplementary debt accounts or titles
deposit of gold or convertibles notes reserve	initial debt accounts or titles
vault and shop	capital

Indeed the bank anticipates that only the k^{th} of its creditors (including initial and supplementary creditors) will withdraw gold or notes. So the total debt of the bank may multiply to k times the initial deposits. If more creditors want to withdraw their deposit, the bank will fail. So the ratio that matters is k , the credit or deposit multiplier. The bank's capital does not appear in this ratio. Nevertheless the bank balance sheet includes an item called capital, which measures the contribution of the bank's owners. With this contribution, the bank has bought some tangible assets (say, a shop and a vault to keep gold or the notes). These tangible assets

may be seized when the bank fails, but their value compared to that of their deposits is low and no classical theory of monetary creation has considered it: there is no place for a capital ratio in the anti-bullionist or Banking School literature and neither is to be found in the theory of monetary creation by the deposit banks, which developed during the inter-war period.

b) Under the modern system of a non-convertible central money.

It is the universal system nowadays. There is no more metal, and notes are non-convertible and are issued by a central bank. The standard balance sheet of a deposit bank is depicted in Figure 2. Notes and accounts at the central bank constitute the central bank money. That part of these notes and accounts which belongs to the deposit banks (in green on Figure 2) are called reserves by monetary theorists. Central bank money appears in two ways:

(A) by means of credit advances from the central bank to the State, but these advances may be forbidden, as in Euroland;

(B) by advances from the central bank to deposit banks. Indeed banks borrow central bank money either for the interbank payments or because their customers want notes.

So the reserves arise in two ways:

(1) either the State spends central bank money for the services of a supplier or an employee, who holds it in his own deposit bank;

(2) or the deposit bank borrows from the central bank.

We may notice that the increase in reserves entailed by the first method is not the true property of the bank, since the bank keeps these funds for the account of its customer, the State's supplier or employee. So the first method resembles the old system when the customer brought gold to the bank. But this method works only for the State's suppliers and employees and it cannot possibly work if the State is not allowed to borrow from the central bank as in Euroland.

With the second method, the deposit bank borrows from the central bank only when it is obliged, i.e. when people want notes or when the bank needs central money for interbank settlements. These situations occur only when banks have already lent, so the reserves appear *ex post* and the concept of credit/deposit multiplier metamorphosizes into that of credit divisor.

c) Common characteristics.

In both systems, the ratio that matters is a link between the reserves (in green) and the money created, therefore the item "capital" (in red) of the balance sheet does not appear. In both systems, the reserves do not belong to the banks but are borrowed either from a depositor (a depositor of gold in the gold regime and as a result of a State's supplier or employee depositing central bank money in the modern system) or from the central bank in the modern system.

In fact the separation between the two regimes is not absolute, since, for instance, during the 19th century deposit banks might use the discount window of the central bank to get central bank money for loans that they had already made. In which case, reserves might be *ex post*. But what matters is that in no case bank's capital is considered.

Hence, despite the uselessness of capital for money creation and for the security of banks deposits, the Basel committee wants to strengthen capital requirements in implementing Basel III.

2. The new ratios implemented by Basel III denote a non-comprehension of the way the banking system works, they will entail more securitization and so will be counterproductive

2.1 The principal ratios of the Basel III documents

The documents published by the Basel Committee on Banking Supervision, as well the two consultative of December 2009 (BIS 2009c,d) and the two final of December 2010 (BIS 2010a,b) continue to strengthen the capital ratio invented with Basel I in July 1988 and they introduce two new ratios, the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR).

The strengthening of capital ratio reflects the continuation of Basel I and II (BIS 1998 and 2009a).

Basel I has introduced an 8% ratio of capital on assets weighted according their nature called the “Cooke ratio” and Basel II has improved this ratio, now called the “McDonough ratio” by introducing the risk-weighting of assets. Basel III (BIS 2010a,b) now wants to add a capital conservation buffer, that is an increase of the ratio in periods of good conjuncture, which will reach 2.5% on January 2019. A countercyclical buffer, which could rise to 2.5%, could be added in period of excess credit growth and the committee is working on the possibility to have larger requirements for systemically important banks. Moreover, Basel III introduces a simple leverage ratio that is not risk-weighted.

Indeed one can wonder what the precise justification of these figures is. There is, for instance, no explanation of the initial ratio of 8% and it seems that the Basel committee wants to increase it without limit. Nevertheless the newly introduced Basel III ratio, the NSFR, seems to justify the importance of bank capital, but it is an application of the classical financial analysis originally conceived for industrial firms and which cannot be applied to banks.

2.2 The new ratios LCR and NSFR are transposed from the financial analysis of industrial firms and, at first glance, do not look arbitrary

Both ratios come from a simple idea developed for industrial firms: the assets are used to pay the financial commitments written on the liabilities side of the balance sheet. So the payment of each commitment must be warranted by an asset of the same maturity or shorter: roughly the short-term liabilities must be covered by short-term assets. The Basel III accords (BIS 2010a,b) suggest (we write “suggest” since the Basel accords are not compulsory before being enacted by each jurisdiction) to take stock of all the commitments to be paid in the next 30 days and to verify that they are covered by inflows coming from the assets during the same period. The stock of assets that are supposed to assure inflows in the next 30 days is called “*Stock of High Quality Liquid Assets*” (SHQLA). The LCR is the ratio of this stock divided by the “*total Net Cash Outflows over the Next 30 calendar Days*” (NCON30D). The ratio must be at least 1,

$$\text{i.e. LCR} = \frac{\text{SHQLA}}{\text{NCON30D}} > 1.$$

On the other hand, the stable funds (called “*Available Amount of Stable Funding*” (AASF) by the Basel documents (BIS 2009b, 2010a,b)), that is shareholders' equity and liabilities of which maturity is at least one year, may be covered by stable assets (called “*Required Amount of Stable Funding*” (RASF)), i.e. assets of which maturity may be longer than a year. On the basis of this, the Basel III accord suggests that the NSFR, of which the numerator is AASF and the denominator RASF, be at least 1,

$$\text{i.e. NSFR} = \frac{\text{AASF}}{\text{RASF}} > 1.$$

This new ratio seems to justify a ratio for the capital requirement. If the assets of maturity of one year or more are $x\%$ of all the assets, the funding of one year or more, i.e. the stable funding, must be $x\%$ of the whole funding. That gives a ratio of stable funding which is not arbitrary, while the ratio of 8%, as well as the ratio of the conservation buffer, are arbitrary.

We may notice that the Basel III documents (BIS 2009c,d, 2010a,b) say nothing about assets and liabilities of which maturity is medium, i.e. included between 30 days and one year.

If both ratios are satisfied, the bank's balance sheet could be as shown on figure 4.

Both ratios may be criticised from two points of view: (a) from an internal point of view and (b) from an external point of view.

- (a) From an internal point of view, we may show that even if these ratios were correct in their principle, they would be inefficient, which means they would not do wrong, but they would be only inefficient, in which case they must be complemented by other measures.
- (b) From an external point of view, we will show that these ratios denote that their policy supporters have not understood how banks work, and what their place in the economy is. Indeed the implementation of the ratios will worsen the financial situation of industrial firms. Moreover banks will increase securitization and such policy thus will be counterproductive. Consequently, these ratios could do harm.

Figure 4 Balance sheet of a bank showing is financial situation

SHQLA	NCON30D
	MEDIUM LIABILITIES
MEDIUM ASSETS	AASF
RASf	

3. Even if the LCR and NSFR were well conceived, they would be inefficient

Even if both ratios were easy to conceive and appropriate, their implementation would be much more delicate. Now one must define exactly what must be included in the numerator and denominator of each ratio. The Basel documents (BIS 2010a) give instructions to calculate these elements, but the instructions turn out to be arbitrary. First let us consider the assets that constitute the SHQLA, i.e. the numerator of the LCR. They are divided into “Level 1” and “Level 2” assets with a maximum of 40% of “Level 2” (BIS 2010a, p. 7). The “Level 1” assets are the assets on public entities (States, central banks, international organizations...), the “Level 2” assets are those issued by private entities. A minimum haircut of 15 % has to be applied to each “Level 2” asset. The figures of 40% and 15% are completely arbitrary and, as L. Randall Wray (2006) has already said for Basel II, these rules are only “rules of thumb that guide good banking practice”.

The same can be said about NCON30D, the denominator of the LCR. The way to calculate it is outlined below (BIS 2010a, p. 12):

*“The term net cash outflows is defined as the total expected cash outflows minus total expected cash inflows in the specified stress scenario for the subsequent 30 calendar days. Total expected cash outflows are calculated by multiplying the outstanding balances of various categories or types of liabilities and off-balance sheet commitments **by the rate at which they are expected to run off or to be drawn down.** Total expected cash inflows are calculated by multiplying the outstanding balances of various categories of contractual receivables **by the rate at which they are expected to flow in** under the scenario up to an **aggregate cap of 75%** of total expected cash outflows.” (Emphasis added)*

One can wonder why it is necessary to limit the cash inflows and why this limit must be of 75% of the outflows and not a different figure. The rates at which the unsecured liabilities are expected to run off are specified (BIS 2010a, pp. 12-17) and go from 5% for deposits “from natural persons covered by an insurance

and for which the depositors have an established relationship with the bank” to 100% for “unsecured wholesale funding provided by legal entity customers” through 25% for “unsecured wholesale funding with operational relationships”. The rates are arbitrary and the classification remains largely subjective. For secured liabilities, derivatives and inflows, the document gives also rates and classifications which are equally arbitrary and subjective.

Let us consider AASF, i.e. the numerator of the NSFR. The Available Stable Funding compounds not only true permanent funding but also deposits, even if they have a maturity less than one year, to the extent that a part of them is expected to be stable. Hence, each liability is affected by an ASF factor depending on its expected degree of stability. These factors are given in a table (BIS 2010a, p. 27). All the true permanent funds, i.e. shareholder's equity (as defined in our section 1 and which appear under the names of “tier 1” and “tier 2”) and the liabilities with a maturity of one year or more have an ASF factor of 100%, while deposits for less than one year have a lower ASF factor. Of course there is a relation between the outflows of the deposits in the next 30 days as considered for the LCR and their stability for one year or more as considered for the NSFR, and what we have said about the subjectivity of the classification and the arbitrariness of the rates remains, even if the rates are now called ASF factors.

As for the RASF, i.e. the denominator of the NSFR, it excludes not only the financial assets with a maturity less than one year, but also assets with a maturity of one year or more (called “marketable assets”) to the extent that there exists a market that allows to selling them before one year. Hence, the RASF includes all the assets with a maturity of one year or more, but it affects them with a “RSF factor” going from 100% for the non marketable ones to 5% for the marketable assets issued by public institutions. Of course the liquidity of the market, and thus the RSF factor, is very subjective.

To summarize the main problems encountered in the calculation of both ratios, we may say that it is difficult to distinguish stable and unstable deposits, i.e. long-term and short-term liabilities. It is also difficult to distinguish permanent and liquid assets, because of the existence of markets which allow for the sale of assets before their maturity. So ASF and RSF appear arbitrary. These difficulties are very peculiar to banks and make the assimilation of banks' finance to industrial firms' finance inappropriate.

The ratios and factors can only be derived from past experiences: the values suggested by the Basel committee (BIS 2010a) can only be normal values derived from history, even if a security margin is added. If only some banks depart from these normal values, the gaps can be covered by insurance. Unfortunately no insurance may cover a systemic risk, which can only be remedied by an intervention of the lender of last resort, the central bank. If it happened exceptional withdrawals on deposit accounts exceeding their normal values, the only remedy would be a supplementary issue of bank notes, and any dysfunctional behaviour of the money market, as it happened in 2008, would imply an expected increase of the RSF factor and an intervention of the central bank would be necessary.

As these ratios come from the past experiences and are very technical, only the men who have lived these experiences, that is bankers, can establish them. It is very significant that the Basel accords come essentially from the work of bankers. The Basel committee has published the 273 opinions (BIS 2010c) that it has received concerning the consultative documents (BIS 2009c,d). The distribution of these opinions is given in Figure 5.

It appears that 80% of the opinions have been given by representatives of the financial sector, that is the members of “financial capitalism”, and only 6% by the members of “industrial capitalism”, the audit and accounting firms working for both forms of capitalism. The last three groups (supervisors, academics and NGO) are supposed to work for the interests of the whole society, but in fact, a large part of the supervisors, especially from central banks, and even of academics, have strong links with “financial capitalism”. Thus “financial capitalism” is over-represented, “industrial capitalism” is highly under-represented, and the rest of society nearly not represented.

Figure 5: Distribution of the opinions received in the consultative documents

Banks and banks' professional organisms	66%
Insurances	3%
Other financial institutions (exchange entities, counsels...)	7%
Credit rating agencies	4%
Sub-total "financial capitalism"	80%
Industrial firms and industrial firms' organisms	6%
Audit and accounting firms	1%
Supervisors	6%
Academics	6%
Non-governmental organizations	1%
Total	100%

One must regret that a profession whose behaviour impacts strongly all of society is controlled only by itself with nearly no participation from the rest of society. Of course the Basel documents (BIS 2010a,b) are only suggestions that will be implemented after the legal processes work themselves out in the different jurisdictions, but we may fear that these processes constitute only rubber stamping.

4. The emphasis on LCR and NSFR reveal a lack of understanding of the place of banks and finance in the economy

Following the Basel documents (BIS 2010a) the LCR and the NSFR should be greater than one. We can distinguish, which is not done by the Basel committee, the deposit banks, which create money, and the other financial institutions, which do not. On the basis of this, we can show that, for the deposit banks, these ratios cannot be greater than one. On the other hand, the aggregate constituted by all the financial institutions, deposit banks and other, may have a NSFR greater than one, but this would imply that the aggregate constituted by all the other economic agents would have a similar ratio lower than one. This means that a situation for financial institutions that is qualified "good" by the Basel committee would imply a situation considered "bad" by classical financial analysis for the real economy.

a) LCR and NSFR can never be greater than one for deposit banks.

Financial institutions practice bank operations, among them loans and borrowings, as a normal activity. They have to be licensed and, in general, are not allowed to have other activities. Hence there is a legal distinction between financial and other institutions.

More peculiarly financial institutions are the only entities that are allowed to receive deposits from the public. When they receive demand deposits, or short-term deposits, that result from their loan-making activity, they create money or near money. We shall call deposit banks, or simply banks, the financial institutions that receive these deposits. Indeed most money creation comes from an exchange of debts between a bank and a non-bank entity: the former gives to the latter a bank debt (called loan), while the latter gives to the former an ordinary debt of the same amount. In the books of the bank, this debt exchange is written in two separate accounts: in a loan account on the assets side for the ordinary debt to the bank, and in a deposit account on the liabilities side for the debt from the bank. By its very nature the debt to the bank has a certain maturity: when someone borrows from the bank, it is for any motive (buying, paying a salary, lending to someone else...) and, whatever the motive, it will imply that the money will be given by the borrower to someone else. Therefore the bank will not be reimbursed immediately. At the opposite the debt from the bank has a legal maturity of zero, but, in fact, as the deposit has certain stability, it has a longer effective maturity of which expectation is given by the ASF factor. If there existed only one large bank in the

economy, the effective maturity would be exactly the same as that of the loan, since the deposit would remain in the same bank, even when it would be transferred from one depositor to another. Therefore it would be reimbursed exactly at the same moment as the bank loan.

In our actual world, with several banks, deposits have a zero legal maturity and an effective maturity at most equal to the maturity of the bank loans. So the LCR and the NSFR are at most equal to one, and searching to have these ratios greater than one would entail dysfunctions.

b) If the aggregate constituted by all the financial institutions has a NSFR greater than one, the real economy would be in a "bad" financial situation

Now we consider the aggregate constituted by all the financial institutions and we suppose it has achieved a NSFR greater than one. Since deposit banks have a NSFR less than one, that would imply that simple financial intermediaries would have a larger NSFR to compensate for the lower NSFR of the deposit banks. This NSFR > 1 for financial institutions will have implications for the real economy, i.e. the other part of the economy. To see what will happen in the real economy, we will suppose the economy is closed, and will refer to Figure 6 representing the two balance sheets of the two aggregates constituted by the financial institutions and by the real economy. To facilitate the comparison we have written the balance sheet of the financial sector in the same sense than for the real sector (i.e. equity and tangible assets on the top).

Within these balance sheets we shall consider only the legal maturities, that is before the effect of the ASF and RSF factors. The titles are all those representing loans of at least one year or shares held by the real economy on the financial institutions. The long-term loans are the loans of one year or more from the financial sector to the real economy corresponding to the borrowings of the real economy. The short-term loans and borrowings are similar but have a maturity less than one year. To simplify our demonstration we have not differentiated medium and short-term loans contrary the Basel documents (Basel 2010a) as shown in Figure 4. The deposits include all the liabilities of maturity less than one year from the financial sector to the real sector. The tangible assets are specific to each sector. The equity on the real sector held by the financial sector is included in the long-term loans from the financial sector to the real sector, while the equity on the financial sector held by the real sector is included in the titles held by the real sector on the financial sector. The loans and borrowings of each sector on itself are eliminated via within sector compensation.

Let us suppose the financial sector is in a "good" financial situation, as it is on Figure 5, i.e. its short-term borrowings (the deposits) are smaller than their short-term loans and thus the long-term funding is larger than the long-term assets, implying a NSFR greater than one. Therefore, by symmetry, the short-term borrowings of the real sector are smaller than their short-term loans (the deposits), which implies a "bad" financial situation. So the financial situation of the real sector is symmetrical to that of the financial sector, and the search of a "good" financial situation for the financial sector, i.e. a NSFR greater than one, implies a "bad" one, i.e. a NSFR lesser than one, for the real sector.

We may verify this with the use of national accounts. Let us consider the balance sheet of the whole French economy on December 31, 2009 published by the Institut National de la Statistique et des Etudes Economiques (INSEE 2011). As this is an open economy, we must oppose the situation of the French financial institutions to the aggregated situation of the French real economy with the rest of the world and verify that the situation of the former is the exact reverse of that of the latter. As the balance sheet is highly aggregated there is no distinction between long-term and short-term loans, but between loans which are represented by titles (titles other than shares) and those which are not (credits). If we suppose that the titles represent loans of one year or more and the credits shorter loans, we get the Figure 7 in billions of euros.

Figure 6 Aggregated balance sheets

assets		liabilities	assets		liabilities	
tangible assets of FI		shareholders' equity not held by RE	tangible assets of RE		shareholders' equity not held by FI	
long-term loans including equity on RE		titles including shareholders' equity held by RE	titles including equity on FI		long-term borrowings including equity held by the FI	
short-term loans		deposits	deposits		short-term borrowings	
Financial institutions (FI)			Real economy (RE)			

Figure 7 National accounts balance sheet

	French real economy + rest of the world (first sector)			French financial institutions (second sector)		
	assets	liabilities	balance	assets	liabilities	balance
tangible assets	11 901			247		
equity		11 534	-301		614	301
long-term loans and borrowings	8 992	9 058		5 254	5 188	
short-term loans and borrowings	5 879	6 180	301	5 694	5 393	-301
total	26 772	26 772	0	11 195	11 195	0

The fact that we consider all credits as short-term is fully arbitrary and, hence we shall also consider the opposite hypothesis, whereby all credits are long-term. This is shown in Figure 8.

Figure 8 National accounts balance sheet

	French real economy + rest of the world (first sector)			French financial institutions (second sector)		
	assets	liabilities	balance	assets	liabilities	balance
tangible assets	11 901		1 893	247		-1 893
equity		11 534			614	
long-term loans and borrowings	10 367	12 627		7 781	5 521	
short-term loans and borrowings	4 504	2 611	-1 893	3 167	5 060	1 893
total	26 772	26 772	0	11 195	11 195	0

Whatever the hypothesis we consider, the situations of both sectors are merely mirror opposites, as the short-term and long-term situations of each sector reflect one another. So a “long excess” (i.e. NSFR > 1) for one sector entails a “long deficit” (i.e. NSFR < 1) for the other sector.

Nevertheless the financial sector may improve its NSFR by considering the ASF and RSF factors without implying any change in the NSFR of the real sector. Indeed when a bank improves its ASF factor by considering that a part of its demand deposits is stable, this does not imply anything for the depositors who remain free to withdraw their deposits without notice. And when a bank improves its RSF factor by participating in a market for long-term assets, the legal maturities of the assets remain the same and so is the situation of the debtors. That explains why the financial analysis for industrial firms considers only the legal maturities and has not developed concepts similar to the ASF and RSF factors. Thus the financial situations of both sectors may be non symmetrical.

Therefore the Basel III accords will push the financial institutions, to which deposit banks belong, to use simultaneously four means to improve their NFSR:

- i. increasing the legal duration of debts;
- ii. decreasing the legal duration of financial assets;
- iii. creation of markets to decrease the effective duration of the financial assets by consideration of the RSF factor;
- iv. search for more stability in the deposits to increase the ASF factor.

The first two means above will deteriorate the financial situation of the real economy in the same manner that it will improve the state of the financial sector. The third method will increase securitization. The fourth method will include the conversion of deposits into titles, which is the object of the next section.

3. Where do the capital and other debt titles come from and what are they for, since they are neither for money creation nor to guarantee the deposits?

We have shown in I that capital is not useful for money creation and cannot be a warranty for the deposits, which are only guaranteed by the reserves of central bank money. Moreover we have shown in II that the excessive quest for capital and of other long-term resources will entail securitization, which is a source of financial crises, and a “bad” financial situation for the real economy.

Therefore, we can wonder if banks’ capital has any other purpose than to cover banks’ tangible assets, which are very few. To answer this question we must answer the preliminary question: from where do banks’ capital and other banks’ debt titles come? Indeed at the beginning of the money creation process there is no capital, there are only a loan account and a deposit account which are respectively debited and

credited of the same amount. Afterwards, as money circulates, the deposit accounts may change from one bank to another within the banking system or take the form of bank notes, but their global volume, when we consider the aggregate of the banking sector, remains equal to the amount of the outstanding loans. And there is not yet any room for capital or debt titles from the banks. If we abstract from the fact that a bank receives interests and commissions, the only way a bank can get capital or debt titles is by selling shares or titles to depositors within the same or at another bank. Thus bank capital and debt titles come from a process of conversion of deposits. Therefore the question of their usefulness is really the question of the utility of this conversion. For an individual bank, it is for fulfilling the Basel accord criteria, i.e. to improve its financial situation, but we have seen that such entails a "bad" financial situation for the real economy and so it seems that conversion of deposits into titles or shares is bad from a macroeconomic point of view.

Another difference between deposits on the one hand and titles and shares on the other is that the former are guaranteed in case of bankruptcy while the latter are not. Hence people who accept that their deposits be converted into titles want some advantages like an interest or a dividend higher than the interest on deposits, or, for the important shareholders, some power in the governance of the bank. The difference between remuneration of titles and interest on deposits must be equal to the premium paid to the guarantee fund or to the F.D.I.C., and thus for the bank the cost of titles and deposits is the same.

It remains that the titles have only two useful purposes: (1) the existence of the shares establishes the obvious fact that banks have owners like industrial firms, which of course is something very important to the shareholders, but may be less important for the economy. (2) As following the Basel criteria, the quantity of loans given by each bank is a multiple of its capital, that determines the quantity of money created by each bank, and therefore the global quantity of money. Thus governments give up any monetary policy for rules, which are international since the big banks are multinational.

Rules constitute the philosophy of the Basel accords: in a free market system, the executive power of the State must not intervene, so there are no monetary policy and no control of the banks. But the experience shows that if the financial sector is completely deregulated, economic crises occur. Hence the market needs rules, which must be written and enacted by the legislative power of the State. The main qualities of rules are that they ought to be simple and permanent, but the Basel accords are quite complex and changed a great deal from Basel I to Basel III, and the further elaboration of Basel III remains unfinished. These rules look partly arbitrary, they have not been elaborated democratically, and they entail dysfunctions in the real economy. Moreover they will be inefficient to prevent systemic crises which can only be remedied by an intervention of the central bank, i.e. an intervention of the executive power, since nobody can ignore that the central bank is a government-owned institution.

These rules allow banks to create all the money they want. First, it is well understood that, contrary to what is taught by the orthodox theory, the central bank cannot control *a posteriori* the quantity of money created by a bank: once a loan is issued by a bank, the central bank cannot, without provoking a crash, refuse to give bank notes or to transfer the money created to another bank, via the interbank market for funds. Secondly, by converting a part of their deposits into capital, which destroys money, the owners of a bank can always increase their potential of loans, which recreates deposits and then more money than the money just destroyed. Let us set the required coefficient of capital on loans is r . Suppose that the shareholders' equity in a given bank is K_0 .

Therefore this bank may issue the quantity of loans and deposits $M_0 = \frac{K_0}{r}$

The bank's owners can convert the part X of their own deposits into shares.

Their equity becomes $K_1 = K_0 + X$. and they can issue loans and deposits for an amount of

$$M_1 = \frac{K_1}{r},$$

M_1 being greater than M_0 .

By repeating this operation loans and deposits may increase without limit⁵.

This process is not automatic, which means that bank owners are not obliged to increase the volume of net deposits when they convert a part of their deposits into shares, but they may. Indeed they will only do so if they find creditworthy borrowers. As the potential increase of deposits is much larger than the real growth, that may happen only if the borrowers and the bankers anticipate an increase in prices, either of current goods or of assets.

Thus we see that the requirement of a capital ratio cannot be a tool to control the quantity of money, albeit his proponents have not thought about this control but only about the security of deposits. We have already said that capital cannot guarantee deposits and it appears now that bankers can increase bank capital and the quantity of money if higher prices are anticipated. But that will decrease security, since it may be very dangerous if the expectations are not realized. We may consider the recent crisis as an example of unrealized expectations of future houses prices.

4. Conclusion: Some ideas to reform deposit banks

We have shown that deposit banks do not need capital, and even capital requirement and high NSFR can lead to a “bad” financial situation for the real sector and to securitization which is another source of dysfunctions. Further the central bank cannot control the quantity of money *a posteriori* and must guarantee deposits.

Therefore two solutions may be considered:

- (i) the “100% money”, which has been suggested by prominent economists (Allais 1977, Fisher 1935) by which money creation is the monopoly of the central bank;
- (ii) the solution we prefer: deposit banks retain their ability to create money but this creation is controlled *a priori*. As they have no capital, or only to the extent they have tangible assets, they have no owners and are controlled by a public agency. As the deposits are guaranteed, they cannot go bankrupt (people would say “they are always saved”), thus they cannot be sanctioned, but a disciplinary board should be instituted to sanction their employees when they violate a regulation.

We must answer the question: how much money must be created and to whom must it be lent? To some this question may imply that money is exogenous to the system of production and that its quantity is controlled by the central bank as, for example, for the monetarists, by means of the deposit multiplier, or for the “100% money” theorists, for the fact that only the central bank creates money.

Both theories leave to the market the question of “to whom must it be lent?”, which means that money is lent to the most profitable activities, with this process determining the bank rate of interest. But experience shows that the most profitable activities may be only speculative and not productive. Post-Keynesian theory shows that money is endogenous, i.e. its quantity is not determined *ex ante* (i.e. before the production) globally by the sole monetary institutions, but at the beginning of each act of production in an exchange between the industrial firm and its bank. In fact for Post-Keynesian theory there is no conscious determination of a global quantity of money, this quantity being the result of individual borrowings from banks in order to finance production. Therefore the problem is not to determine the “good” quantity of money, but to verify that each act of marketable production may be financed by an adequate creation of money and that each act of money creation finances a marketable production and nothing else as a speculative operation. Hence it is necessary to separate deposit banks, which create money, from other financial institutions, with access to credits of the central bank being reserved to the deposit banks and to government. Further the monetary circuit theory (Vallageas 2013, chapter 6) shows that monetary creation is at least equal to the flow of incomes paid for the production of all goods and services, either for consumption or investment purposes. In order to prevent money creation for speculative purposes, we suggest that deposit banks finance only

⁵ From the idea that Monetary Authorities cannot control the quantity of money through the control of bank capital, Werner (2010a,b) concludes that they must use other control means. He quotes Wicksell (1907) to affirm that this control through bank capital is unfeasible.

those incomes and that they be obliged to finance all of them, the only condition being that the production be marketable⁶. Speculative activities and the purchase of investment goods (not to be confused with their production) would be financed by other financial institutions, if necessary. Therefore all the marketable production would be financed, which would lead to full employment. If we abstract from the costs of running banking institutions, that should be reimbursed, and following a suggestion of the Kansas City School (Forstater and Mosler 2004), bank interest would be zero, since it would not be necessary to limit production to a predetermined quantity of money, thereby eliminating the projects whose profitability would be less than the rate of interest.

The regulations that we propose will be certainly understood by some readers as a breach of the liberty of trade. But deposit banks are not industrial firms, since the former have the privilege to access central bank credits and to create money. Thus, as a counterpart of this privilege, they must work for the interest of society. Indeed liberty of trade means liberty to finance for any marketable production project and the reform we propose would guarantee this liberty.

References

Allais, M., 1977. *L'impôt sur le capital et la réforme monétaire*. Paris: Hermann.

Bardoloi, S., 2003. *Basel II: New Wine in an Old Bottle*. Risk World. [on line] Knoxville, TN: Tec-Com Inc. Available at: < <http://www.riskworld.com/news/03q2/nw03a103.htm> > [Accessed 1 June 2011].

Bank of International Settlements (BIS), 1998. *The Basel Capital Accord, International Measurement of Capital Convergence and Capital Standards* (July 1988, updated to April 1988) [on line] Available at: <<http://www.bis.org/publ/bcbsc111.pdf>> [Accessed 1 June 2011].

BIS, 2004. *International Convergence of Capital Measurement and Capital Standard. A Revised Framework*. June 2004 [on line] Available at: <<http://www.bis.org/publ/bcbs107.pdf>> [Accessed 1 June 2011].

BIS, 2009a. *History of the Basel Committee and its Membership* (August 2009) [on line] Available at: <<http://www.bis.org/bcbs/history.pdf>> [Accessed 1 June 2011].

BIS, 2009b. *Comprehensive response to the global financial crisis* (September 7, 2009) [on line] Available at: < <http://www.bis.org/press/p090907.htm> > [Accessed 1 June 2011].

BIS, 2009c. *Strengthening the resilience of the banking sector* (December 2009) [on line] Available at: <<http://www.bis.org/publ/bcbs164.pdf> > [Accessed 1 June 2011].

BIS, 2009d. *International framework for liquidity risk measurement, standard and monitoring* (December 2009) [on line] Available at: < <http://www.bis.org/publ/bcbs165.pdf> > [Accessed 1 June 2011].

BIS, 2010a. *Basel III: International framework for liquidity Risk measurement, standard and monitoring* (December 2010, <http://www.bis.org/publ/bcbs188.pdf> > [Accessed 1 June 2011].

BIS, 2010b. *Basel III: A global regulatory framework for more resilient banks and banking systems* (December 2010) [on line] Available at: < <http://www.bis.org/publ/bcbs189.pdf> > [Accessed 1 June 2011].

BIS, 2010c. *Comments received on the consultative documents "Strengthening the resilience of the banking sector" and "International framework for liquidity risk measurement, standard and monitoring"* (December 2010) [on line] Available at: < <http://www.bis.org/publ/bcbs165/cacomments.htm> > [Accessed 1 June 2011].

Forstater, M. and Mosler, W., 2004. *The Natural Rate of Interest is Zero* (Working Paper No 37, December 2004). Kansas City, MO: University of Missouri [on line] Available at: < <http://www.cfeps.org/pubs/wp-pdf/WP37-MoslerForstater.pdf> > [Accessed 1 June 2011].

⁶ Werner (2010a,b) agrees with Post-Keynesians that money is created by banks and that it should be created only for productive purposes. Therefore he concludes that an *a priori* control should be instituted or re-instituted, and he cites countries (Werner 1997, 2002, 2005), as Japan, where this control exists. Nevertheless he keeps the idea that Monetary Authorities should control *a priori* both the quantity of money and its distribution between productive activities, i. e. in fact marketable activities.

Fisher, I., 1935. *100% money; designed to keep checking banks 100% liquid; to prevent inflation and deflation; largely to cure or prevent depressions; and to wipe out much of the national debt*. New-York: Adelphi Institut National de la Statistique et des Etudes Economiques, 2011. Tableau Economique d'Ensemble au 31 décembre 2009 [on line] Available at: <http://www.insee.fr/fr/themes/comptes-nationaux/default.asp?page=archives/archives_cnat_annu.htm> [Accessed 1 June 2011].

Minsky, H. P., 1982. *Can "It" Happen Again?: Essays on instability and finance*. Armonk, NY: M. E. Sharpe.
Toporowski, Jan, 2009, "Enforced Indebtedness" And Capital Adequacy Requirements (Policy Note, 2009/7, May 2009). Annandale-on-Hudson, NY: The Levy Economic Institute of Bard College: [on line] Available at: <http://www.levyinstitute.org/pubs/pn_09_07.pdf> [Accessed 1 June 2011].

Vallageas, B., 2009. Analyse de la crise financière actuelle. Où vont-ils chercher ces milliards qu'ils n'ont pas? Comment l'Etat assureur en dernier ressort permet d'accéder à la Banque Centrale Européenne, prêteur en dernier ressort, *Revue de la Régulation* [on line] Available at: <<http://regulation.revues.org/index7544.html>> [Accessed 1 June 2011].

Vallageas, B, 2013. *Monnaie, Profit et Valeur. Un essai sur le circuit, l'économie monétaire de production et sur les liens entre la macroéconomie et l'analyse financière des entreprises*. Forthcoming.

Werner R. A., 1997. 'Towards a new monetary paradigm: a quantity theorem of disaggregated credit', with evidence from Japan. *Kredit und Kapital, Duncker and Humblot*, Berlin, 30, pp. 276–309.

Werner, R. A., 2002. Monetary policy implementation in Japan; what they say vs. what they do, *Asian Economic Journal*, 16, pp.111-51

Werner, R. A., 2005. *New Paradigm in Macroeconomics: Solving the Puzzle of Japanese Macroeconomic Performance*, UK : Basingstoke: Palgrave Macmillan

Werner, R. A., 2010a. *Comment received on the consultative document: Strengthening the Resilience of the Banking Sector* (Bank of International Settlements, April 16, 2010): [on line] Available at: <http://www.bis.org/publ/bcbs165/universityofsou.pdf> [Accessed 1 November 2012].

Werner, R. A., 2010b. *Towards Stable and Competitive Banking in the U.K. - Evidence for the ICB* (Centre for Banking, Finance and Sustainable Development, Policy Discussion Paper No 3/1 – 10, November 15, 2010). University of Southampton, UK: [on line] Available at: http://eprints.soton.ac.uk/342277/1/Werner_Soton_Towards_Stable_Banking_201011.pdf [Accessed 1 November 2012].

Wicksell, K., 1907. The influence of the rate of interest on prices, *Economic Journal*, 17, pp.213-20.

Wray, L. R., 2006. *Can Basel II Enhance Financial Stability? A Pessimistic View* (Public Policy Brief, Highlights N° 84A, May 2006). Annandale-on-Hudson, NY: The Levy Economic Institute of Bard College: [on line] Available at: <http://www.levyinstitute.org/pubs/hili_84a.pdf> [Accessed 1 June 2011].

Deconstructing the Theory of Comparative Advantage¹

Reinhard Schumacher

Department of Economic and Social Sciences, Universität Potsdam, Germany

Abstract

This article critically examines the theory of comparative advantage, which underlies the wide-spread support of worldwide trade liberalisations. Both the classical and neoclassical formulations of it are shortly discussed and its essential assumptions are scrutinised. These include the international immobility of capital and labour, balanced trade, the existence of an adjustment mechanism which is responsible for the transformation of comparative production advantages into absolute price advantages, full employment and the perception of international trade as a static and harmonious phenomenon. It is shown that all these assumptions are neither theoretically valid nor do they coincide with empirical research. The whole rationale why international trade exists according to this theory is deficient. The New Trade Theory, which claims to enhance the theory of comparative advantage, is unconvincing as a complement. It is concluded that the theory of comparative advantage should be dismissed. International trade theory, by relying on this theory, risks ignoring the most relevant and important elements with regard to international trade. The deficiencies of the theory of comparative advantage are especially crucial for trade policies that are derived from this theory, which is discussed with reference to the WTO and its ongoing Doha Round.

Key words: International Trade; Trade Theory; Comparative Advantage, Trade Policy, WTO

1. Introduction

Economists favour unrestricted international trade nearly unanimously. Polls in the last decades show that there is an overwhelming academic support for free trade among the profession (Kearl et al. 1979; Frey et al. 1984; Alston et al. 1992; Fuller and Geide-Stevenson 2003), which is unabatedly high. Theoretically, this support is underpinned mainly by the theory of comparative advantage. It shows why free trade is beneficial for all nations as well as the world as a whole and how free trade automatically leads to the realisation of those benefits. This theory dominates international economics or, more precisely, the theory of international trade. It is widely praised and has been vaunted as the “deepest and most beautiful result in all of economics” (Findlay 1987, p. 514) and as “an unassailable intellectual cornerstone” (Harrigan 2003, p. 86). Samuelson calls it the only proposition in social science that “is both true and non-trivial” (Samuelson 1972, p. 683). Some even argue that it has become “something of an article of faith” (MacDonald and Markusen 1985, p. 277) in modern economics. This theory is used to argue that it is of central importance for poor nations to open markets and to join a free trade regime in order to raise living standards. The whole WTO process of trade liberalisation that assumes that global economic welfare can be maximised through free trade has its theoretical foundations in the theory of comparative advantage.

The use and support of this theory, however, is surprising once its framework is considered. It rests on the assumptions that labour and capital do not move between nations, that there are no trade imbalances, that all resources are always fully employed and that international trade can be described by a comparative-static model. These assumptions are neither self-evident nor self-explanatory at the first glance. Rather, they evoke questions concerning their usefulness and adequacy. Therefore, this article tries to

¹ The author declares no conflict of interest. I would like to thank Peter Schmidt, Jorge Morales Meoqui and Norbert Häring for their very helpful comments and suggestions.

examine these assumption and with it the theory of comparative advantage.

Before this is done, a short introduction to the theory is given, which is closely linked to the name David Ricardo,² who formulates it in chapter 7 of his main work *On the Principles of Political Economy and Taxation*. Succeeding classical³ and neoclassical economists adopted this theory. The latter reformulated it to fit it into the marginal analysis framework, which is the fundament of neoclassical economics. This neoclassical or modern formulation of the theory of comparative advantage is the basis of the most widespread trade models and is used in textbooks on international economics. Both the classical and the modern formulation share the same basis and logic. They both consist of a normative dimension, which shows that free trade is advantageous for all participating nations, and a positive dimension, which describes how each nation specialises according to its comparative advantages automatically.⁴

Afterwards the assumptions will be scrutinised and the “magic of comparative advantage” (The Economist 2009, p. 13) disenchanting. I will focus on those assumptions that are, as will be shown, crucial and essential to the theory of comparative advantage regardless of its specific formulation. These assumptions are discussed and analysed concerning their theoretical and logical consistency as well as their empirical accuracy. The surface impression regarding the usefulness and adequacy is confirmed. It is shown that all these assumptions have both theoretical as well as empirical deficiencies. Hence, the conclusion is drawn that the theory of comparative advantage is all but irrelevant to international trade and is inadequate for explaining free international trade. However, these deficiencies are especially crucial for trade policies that are derived from this theory, which is discussed with reference to the WTO and its ongoing Doha Round.

2. Classical formulation

Ricardo⁵ is the first economist who distinguishes international trade from domestic trade (Blaug 1977, p. 126). He wants to show that international trade follows different rules than domestic trade (Ricardo 2004b, p. 133).⁶ This distinction is based on the assumption that labour and capital do not move between nations as they do inside a nation. The reasons for the immobility of capital are “the fancied or real insecurity of capital, when not under the immediate control of its owner, together with the natural disinclination which every man has to quit the country of his birth and connexions, and intrust himself with all his habits fixed, to a strange government and new laws” (Ricardo 2004b, p. 136). The immobility of labour also originates from the latter reason. Therefore, free international trade is determined, unlike free domestic trade, by comparative production advantages.

England and Portugal

Ricardo believes that free international trade is desirable and that nations and their populations benefit from it.⁷ He demonstrates this point with his famous England-Portugal example, in which both nations initially

² Many attribute the credits for its discovery to Robert Torrens. Since Ricardo's formulation is referred to most often by economists, I will concentrate on him. For an early discussion of the role of Torrens see Seligman and Hollander (1911). For a list of literature on the discovery controversy and a discussion of it, see also Ruffin (2002), Aldrich (2004) and Maneschi (1998, pp. 51-57).

³ In this article, the term ‘classical’ describes Ricardo and his successors, not his precursors like Adam Smith who did not use the theory of comparative advantage.

⁴ Despite the fact that the theory of comparative advantage is often acknowledged as a ‘pure’ theory of international trade, it relies on a monetary mechanism which is an essential part of it.

⁵ Ricardo's approach, as presented here, differs from the so called Ricardian model that can be found in most textbooks of international economics. The latter tries to reconstruct Ricardo's theory with neoclassical tools and assumptions that Ricardo did not use. Additionally, the classical formulations was formalised mainly by John Stuart Mill. However, I will focus on Ricardo who first gave a full account of this theory and whose name is associated with it.

⁶ Adam Smith before him developed no separate theory of international trade. He rather argues that it corresponds to domestic trade: “Were all nations to follow the liberal system of free exportation and free importation, the different states into which a great continent was divided would so far resemble the different provinces of a great empire” (Smith 1904, p. 41).

⁷ Ricardo's main object of study is the distribution of wealth among the different classes of a society (capital owners, workers, and landlords). His aim is to show that “the amount of value in a country” cannot be increased through international trade and that the

produce cloth and wine. To produce the same amount of cloth, England needs 100 labourers and Portugal 90 labourers. England needs 120 labourers to produce the same amount of wine that Portugal produces with 80 labourers (Ricardo 2004b, p. 135). These different labour requirements are due to their dissimilar circumstances that result from a nation's situation, climate and other "natural or artificial advantages" (Ricardo 2004b, p. 132) and are exogenously given.

With these "four magic numbers" (Samuelson 1972, p. 678), Ricardo shows that it would be advantageous for both nations if they specialised according to their respective comparative advantage and started trading with each other. England should specialise in the production of cloth and import wine from Portugal. The opposite applies to Portugal. As a result, they would both benefit.⁸ Due to the more efficient employment of labour and capital, "the amount and variety of the objects on which revenue may be expended" (Ricardo 2004b, p. 133) and "the sum of enjoyments" (Ricardo 2004b, p. 128) increase. The whole population, as consumers, benefits from international trade because goods become cheaper and available in larger quantities. There are no other economic gains from international trade. Dynamic developments like economic growth are not integrated into the theory of comparative advantage by Ricardo.⁹

Ricardo illustrates with his example that no nation needs to fear free international trade because it will be advantageous for both nations, even if one nation has a lower productivity in all goods and the other nation produces both goods more efficiently. This shows that absolute production costs are insignificant internationally, only comparative production costs matter. If the cost ratios are different in both nations, specialisation and trade will benefit both. However, the opposite is also true, namely when production cost ratios are equal in both nations no gains can be made by specialisation. In this situation trade will not take place at all because there would be no incentive for it. Different comparative production costs are "the essential and also the sufficient condition" (Cairnes 1874, p. 371) for the existence of international trade.

The price-specie-flow mechanism

Ricardo shows not only that free trade is advantageous for nations, but also that nations will benefit automatically because free international trade leads inevitably and even unintentionally to a specialisation according to comparative advantages. Here, Ricardo draws on the price-specie-flow mechanism, which is a simple version of the quantity theory of money and was developed by David Hume (1903a, 1903b).

Ricardo and succeeding classical economists present international trade as a form of barter. It is seen as "an actual trucking of one commodity against another" (Mill 1929, p. 583). Money is seen as neutral and has only one function in international trade, namely as a means of exchange in order to facilitate trade.¹⁰ As a corollary, trade must be balanced. This is an important presumption of the price-specie-flow mechanism.

According to this mechanism, changes in the quantity of gold (and silver), which was the means of payment at the time, have no real effect, only a price effect. Thus, absolute (gold) prices, wages, etc. depend on the quantity of gold that is available inside a nation. To illustrate how this mechanism underlies the theory

distribution of wealth is not influenced by international trade because neither the general rate of profit nor wages are generally affected by it (Ricardo 2004b, pp. 128-132). The latter point is directed against Adam Smith.

⁸ In Ricardo's example England "gains the labour of 20 Englishmen" and "Portugal gains the labour of 10 Portuguese" (Sraffa 1930, p. 541).

⁹ Ricardo expounds one exceptional case in which international trade leads to an increase in the general rate of profits, namely when international trade leads to a change in the level of wages (Ricardo 2004b, pp. 128-132). In Ricardo's theory, profits and wages are mutually dependent. If wages decrease profits will increase and vice versa (Ricardo 2004b, pp. 110-111). If the prices of wage goods (necessities of the labourers) decline because they can be acquired cheaper through foreign trade, wages will fall and profits will rise. This would lead to accumulation of capital and to economic growth in Ricardo's theory. But both models – the theory of comparative advantage and the profit-increasing model – are separated in Ricardo's theory: "In fact, Ricardo used two trade models which he never managed to integrate" (Gomes 2003, p. 44). The theory of comparative advantage does not include the profit-increasing model. Furthermore, Ricardo sees the latter only as an exceptional case (2004a, p. 25; 2004b, p. 133).

¹⁰ Hume argues that this neutrality of money is only true in the long run. In the short run a change in the quantity of money does matter. If money flows into a nation and thus the quantity of money is increased, then "every thing takes a new face: labour and industry gain life" (Hume 1903a, p. 293). The opposite is true when the quantity of gold and silver diminishes and is therefore "pernicious to industry" (Hume 1903a, p. 296). Ricardo does not adopt this part of Hume's theory in his theory of comparative advantages, nor do any of his classical successors. A reason why this short run effect is neglected or even rejected by both classical and neoclassical economists, might be that it has a mercantilist connotation as Petrella (1968, p. 366) suggests.

of comparative advantage I will consider Ricardo's example of England and Portugal again. Portugal can produce both cloth and wine with less labour. Supposing that this means the gold price of both Portuguese goods is cheaper when both nations start trading, consumers from both nations buy both commodities from Portugal. England has a trade deficit, while Portugal has a trade surplus. Gold flows from England to Portugal. Portugal's prices, expressed in gold, rise. Contrary, the quantity of money in England is diminished and her prices expressed in gold fall. English commodities become cheaper and England will improve her competitive situation. This leads to a situation in which England becomes competitive in one good and will be able to sell it cheaper than Portugal, namely cloth. The outflow of money from England will gradually slow down because cloths are now exported. Finally, the prices of both goods will adjust in a way that trade is balanced, the value of imports equals the value of exports and an equilibrium state is achieved in which both nations produce the commodity they have a comparative advantage in (Ricardo 2004b, pp. 138-40). Comparative production cost advantages are thus transformed into absolute money price advantages for the consumer. This transformation is significant and necessary: the "cost of production, though it may be, and generally is, the ultimate condition governing international exchange, is never in any case the proximate or immediate cause. That proximate or immediate cause is not cost, but price" (Cairnes 1874, p. 382). Since consumers buy a good from whoever sells it cheapest, comparative production cost advantages must be transformed into absolute price advantages.

This mechanism prevents trade from being unbalanced (Hume 1903b, p. 319). A (perpetual) trade surplus or deficit is thus theoretically impossible under free trade conditions. The volume of trade may change but international trade will always be balanced at least after some time of adjustment. Ricardo has this mechanism in mind when he says that, in a free trade system, "each country naturally devotes its capital and labour to such employments as are most beneficial to each" (Ricardo 2004b, p. 133) and "the exchanges could be no otherwise in every country than at par" (Ricardo 2004b, p. 330). Although each nation seeks to maximise its own advantage, it brings about the best possible outcome because labour is distributed "most effectively and most economically" (Ricardo 2004b, p. 134).

3. The modern formulation

Neoclassical theory, which is based on marginal analysis, succeeded classical theory as leading paradigm in economics. It adopted the theory of comparative advantages and fitted it into this paradigm. Therefore, it was disengaged from properties that were not compatible with neoclassical economics, but its basis was kept. The body of the theory remained the same, only the garb changed. This modern formulation underlies today's dominating theories of international trade. Neoclassical economic tools are used to demonstrate that free international trade is beneficial for all participating nations. The exchange rate mechanism is responsible that those benefits will be acquired automatically if free trade is prevails.

Benefits from trade

In order to show that nations benefit from free international trade, modern economics uses neoclassical microeconomic tools. They include opportunity costs, production possibility frontiers, social indifference curves and optimised production-consumption equilibria in autarky. The simplest illustration uses a 2x2x2 model with two nations, two goods and two factors of production. It is assumed that both goods are produced with both factors of production and that these factors can be substituted. In such a model, the overall amount of goods that a nation can produce is illustrated by the production possibility frontier. It describes simultaneously the possible consumption of a nation in the state of autarky.¹¹ The absolute value of the slope of this frontier equals the opportunity costs of one good expressed in the amount of another good – or the so-called marginal rate of transformation. Since neoclassical economics generally assumes increasing opportunity costs, the production possibility frontier is concave to the origin. Opportunity costs are used to

¹¹ A nation will always produce at a point on this frontier according to neoclassical assumptions. A nation cannot produce outside ('above') its production possibility frontier with the existing resources and technology. Since full employment of all factors is assumed, it cannot produce inside ('below') this frontier.

determine the relative prices in a nation. The exact production point is determined by demand, which in turn is determined by so-called social or community indifference curves.¹² The optimum production composition of the economy is determined by the point where a social indifference curve is tangent to the production possibility frontier.¹³

Comparative advantages are determined by comparing national opportunity costs at the respective optima. If each nation specialises according to its comparative advantage, the overall production increases and through trade the available quantity of commodities in both nations is higher than in the state of autarky. This means, national consumption increases beyond the respective production possibility frontier and each nation can reach a higher social indifference curve. Hence, the consumers' needs are satisfied to a higher degree. In this way, free international trade is beneficial for each nation.

There are several explanations why national opportunity costs differ. Neoclassical economists in general criticise Ricardo because he does not explain the reasons for a nation's comparative advantage and they want to fill this gap. The most famous model, the Heckscher-Ohlin model, assumes that it is the effect of different endowments of factors of production. In this model, each nation has a comparative advantage "in the production of commodities into which enter considerable amounts of factors abundant and cheap" (Ohlin 1933, p. 20) in this nation and will specialise accordingly. Other models consider the demand side as well.

The gains from trade are identically in both the classical and the neoclassical formulation, namely an increase in the overall output and consumption. There are no further gains from trade. Any possible dynamic changes and gains "are completely disregarded" (Heckscher 1949, p. 274). Consumption and production are at an overall maximum. If opportunity costs are equal in both nations and thus no comparative production advantages exist, international trade will not take place, as in the classical theory.

Contrary to classical formulation, the modern formulation assumes that profit and wage rates do change, and even equalise worldwide.¹⁴ But this is only seen as a corollary of free trade and not as a gain, since neoclassical economists, unlike Ricardo and other classical economists, do not give any normative judgement about changes in income (distribution) as long as the overall gain is positive, because winners could then compensate any losers.

Exchange rate adjustment mechanism

Like the classical formulation, the neoclassical formulation of the theory of comparative advantage contains an automatic adjustment mechanism. It has the same function, namely to transform comparative production advantages into absolute price advantages, because, ultimately, absolute price differences determine the international flow of commodities. Neoclassical economists assert the assumption of balanced trade. In today's world national paper money that is not backed by gold is the international means of payment. In case of floating exchange rates, the exchange rate adjustment mechanism is responsible for such a transformation. According to this mechanism, trade imbalances cause a shift in exchange rates. The exchange rate is solely determined by trade flows. The absolute level of money prices is internationally determined by the exchange rate of a nation's currency. Trade imbalances affect the demand for currencies and result in a change of the exchange rate. The currency of the nation that experiences a trade deficit – and thus an outflow of money – will be depreciated and the currency of the nation that has a trade surplus will be appreciated. Thus, the commodities that are produced in the deficit nation will become cheaper internationally while those from the surplus nation will become more expensive. "When exports become equal to imports in money value, the exchange rate will stop moving and equilibrium will exist" (Eicher et al.

¹² Viner (1931), who first combined this concept with the production possibility frontier, used the term "demand curve." These curves are convex to the origin and are composed of points at which consumer needs are equally satisfied. Such a curve can be interpreted either as "the representative citizen" or as "community indifference" (Chipman 1965, pp. 689-98).

¹³ At this point the resources are employed most advantageous for the economy and without trade, there exists no better employment. Consumer needs cannot be satisfied to a higher degree by any other production composition.

¹⁴ The neoclassical formulation was enhanced by the factor price equalisation theorem. This can be seen as "a corollary of the Heckscher-Ohlin formulation of comparative advantage" (Nayyar 2007, pp. 71-72). It was developed by Paul Samuelson (1948) and Abba Lerner (1952) and states that "factor prices will be equalised, absolutely and relatively, by free international trade" (Samuelson 1948, p. 169). As a result, commodity trade leads to the same result as if production factors were internationally mobile.

2009, p. 65). Anytime a trade imbalance exists, the equilibrium state will be restored through this exchange rate adjustment mechanism. Money is seen as neutral. Changes in the exchange rate do not change relative prices or long run allocation of resources. As a result, each nation will automatically specialise in the production of those goods, in which it has a comparative advantage and each nation will be able to “successfully compete in world markets” (Jones 1980, p. 235).

In the case that exchange rates are fixed, neoclassical theory reasons that trade is balanced via wage rate changes. The adjustment is “performed by the price-labor relation in each country” (Haberler 1929, p. 377). The demand for labour is responsible that wages in surplus nations rise. This makes its goods more expensive while wages in deficit countries fall. This wage adjustment mechanism is then responsible for the transformation of comparative advantages into absolute price advantages.

4. Evaluating the assumptions

As was shown, both formulations of the theory of comparative advantage differ in the tools they use. Their underlying structure, however, is the same. Krugman and Obstfeld argue rightly in respect of the theory of comparative advantage that “even though much about international trade has changed, the fundamental principles discovered by economists at the dawn of a global economy still apply” (2009, p. 23). Free international trade is beneficial to all participating nations because they can increase their overall production and more can be consumed if they specialise according to their comparative advantages. Furthermore, this specialisation does not take place by accident or political intention but automatically. A mechanism exists which transforms comparative production advantages into absolute price advantages.

All theories have to abstract from reality and simplify relationships in order to construct a useful and workable framework. The question arises, however, which assumptions are reasonable to make. It must be assessed which simplifications are appropriate. The ‘gains’ in form of clarity and predictability must be balanced against the ‘costs’ in form of relevance and realism. The danger is that simplifications are made which oversimplify the reality and that improper assumptions are made which misrepresent the real world. In this case a theory risks becoming irrelevant by ignoring the most interesting causations and problems. In order to be an adequate theory of international trade and to gain useful insights into this phenomenon, the theory of comparative advantage has to have a justifiable balance between these ‘gains’ and ‘costs’ while avoiding invalid assumptions.

As will be shown, five assumptions constitute the basis of the theory of comparative advantage, which are indispensable to it: first, capital and labour are immobile internationally; second, trade is balanced and an adjustment mechanism operates which prevents trade imbalances and at the same time ensures that comparative production cost advantages are transformed into absolute price advantages; third, all factors of production are fully employed; and fourth, international trade resembles a static phenomenon that excludes dynamic changes and gains. As a consequence of these assumptions international trade is portrayed as a harmonious phenomenon, which is the fifth assumption. These assumptions will be critically examined and assessed in respect to their theoretical validity, logic and empirical relevance.

International immobility of labour and capital

Ricardo states that domestic and international trade are regulated by different rules, the former by absolute and the latter by comparative production cost advantages. The modern formulation adheres to this differentiation. Both, classical and neoclassical economists, rest this difference on the assumption that labour and capital do not move between nations. Domestically, labour and capital are assumed to be mobile and will move wherever they can gain the highest wages or profits respectively. Internationally, only the produced goods can move freely. Capital and labour are trapped domestically. While classical economists assume that workers and capital are more or less bound to a nation by nature as Ricardo states or that the scale of these

movements is negligible (e.g., Cairnes 1874, p. 368),¹⁵ neoclassical economists first adopted this argument (e.g., Haberler 1930, p. 350)¹⁶ and later developed the factor price equalisation theorem that renders such movements superfluous (Samuelson 1948). This latter theorem states that international trade leads to the same result as if capital and labour were mobile. Hence, there is no necessity for their international mobility.

Ricardo is aware that the international immobility of labour and capital is an indispensable assumption. He devotes half of his explanation of the theory of comparative advantage to the discussion of it (Ruffin 2002, p. 734). He even expounds that if this assumption did not apply and labour and capital were able to move internationally, comparative advantages would not determine international trade (Ricardo 2004b, p. 136). He illustrates this with reference to his England-Portugal example by saying that “[i]t would undoubtedly be advantageous to the capitalists of England, and to the consumers in both countries, that under such circumstances the wine and the cloth should both be made in Portugal, and therefore that the capital and labour of England employed in making cloth should be removed to Portugal for that purpose” (Ricardo 2004b, p. 136). If capital could move internationally, English capital would move to Portugal in his example and leave England. This means that comparative advantages would be irrelevant.¹⁷ Therefore, international immobility is essential to the theory of comparative advantage. Otherwise there would be no reason why free international trade should be regulated by comparative advantages.

From a theoretical point of view, however, there is no reason why capital and labour should not be mobile internationally. The classical claim that workers and capital do not move beyond national borders because they have a natural inclination to stay in one’s homeland can hardly be justified theoretically. Similarly, the neoclassical assumption, that factors are trapped domestically has no theoretical basis. The assumption that factor prices equalise, cannot justify international immobility.

In practise, workers move in significant amounts between nations. Even at the end of the 18th and the beginning of the 19th century when Ricardo developed his theory, labour migration took place in large scale, for example from Europe to America (Oswald 2007, pp. 50-54). Today, labour migration “is truly a global phenomenon” (Stalker 2000, p. 31). Migration that is driven by disparity in wages, takes place between many regions. Capital, on the other hand, moves frequently between nations. With the decrease of transport and communication costs, capital has become ever more mobile. The assertion, that capital and workers might be mobile internationally, but that their numbers are negligible, cannot be maintained. Furthermore, there is no evidence that factor-prices equalise worldwide (Subasat 2003, p. 152). This neoclassical “idea that trade should substitute for migration involves a number of assumptions distant from conditions in the real world” (Stalker 2000, p. 33).¹⁸

Consequently, there is neither a theoretical reason nor empirical evidence that labour and capital do not move internationally. Rather, both are mobile to a certain degree – similar to the domestic level. There is no support for the claim that international trade should be determined by fundamentally different rules than domestic trade on these grounds. If capital and labour are assumed to be mobile to a certain degree, comparative advantages cannot determine the pattern of international trade. It is, therefore, not surprising that models that allow for labour and capital movement show that comparative advantages do not determine trade patterns (see, e.g., Brewer (1985); Jones (2000)).

To sum up, the assumption that labour and capital are not mobile internationally lacks a logical and theoretical justification and has no empirical support. It is not an appropriate simplification of reality.

¹⁵ Cairnes is aware that labour and capital move internationally and that at least capital is ‘cosmopolitan’. Yet he concludes that this international movement can be ignored because it represents only a small fraction of the overall capital and labour (Cairnes 1874, pp. 362-364).

¹⁶ Similar, Haberler argues that labour and capital might be mobile internationally but he purposely neglects this point and assumes immobility.

¹⁷ This would also mean that free trade is not necessarily advantageous to both nations because the “flight of capital and labor would be detrimental to England and its remaining population as a whole” (Anspach 1968, p. 515).

¹⁸ It is hardly imaginable that, for example, wage levels equalise in poor and rich nations solely as a result of free trade (Lutz and Lux 1988, p. 286). Pritchett calculates that in the period from 1870 to 1990, while the volume of international trade increased significantly, the income ratio of the richest to the poorest nations, measured in GDP per capital, increased by the factor five. The ratio increased from 8.7 in 1870 to 45.2 in 1990 (Pritchett 1997, p. 3).

Balanced trade and adjustment mechanism

A second crucial assumption is that trade is balanced, meaning that the value of imports equals the value of exports of each nation. This assumption is a corollary of the fact that trade is comprehended as barter. Ricardo insists that international trade takes place as if it “were purely a trade of barter” (Ricardo 2004b, p. 137), a presumption that is maintained by both later classical economists and neoclassical economists.

Balanced trade is essential for the theory of comparative advantage because the adjustment mechanism that results from it is responsible for the transformation of comparative production cost advantages into absolute price advantages. This transformation is vital, because consumers buy goods from whoever offers them at the lowest money price. Consumers are neither aware of nor interested in comparative production costs. Their main decision criterion is the price, given that competing goods are of an equal or at least similar quality. Since trade must be balanced, it is supposed that every single transaction, though performed independently, influences international trade as a whole. An adjustment mechanism must rebalance trade in case of imbalances – and even prevent trade balances a priori.

Theoretically, such an adjustment mechanism is necessary for the theory of comparative advantage.¹⁹ If the value of exports did not equal the value of imports there would be no reason why comparative production differences would be converted into real price differences and would determine international trade flows. The actual form of the adjustment mechanism is only of secondary interest. It can be the price-specie-flow mechanism, an adjustment via wages or via exchange rates. Important is the existence of such a mechanism. If trade were not balanced per se and no adjustment mechanism existed, there would be no reason why comparative advantage would be realised.

From a theoretical point of view, these adjustment mechanisms are problematic. The quantity theory of money, which Ricardo uses, assumes that money is neutral and neglects the velocity of a currency. However, neither is the velocity constant in practise nor is the quantity of money neutral to the real economy. Additionally, if money quantities change, interest rates are affected. If those effects are taken into consideration, there is no theoretical basis why changes should translate one-to-one into price level changes. Another fact, which is overseen by the price-specie-flow mechanism, is that inflation and deflation as a result of gold in- or outflows affect economies beyond mere price changes. A deflation means economic problems and distress for a state, its population and its businesses, that might be ruinous (Polanyi 2001, pp. 201-202). Empirically, the predicted dependence of price levels from the quantity of money cannot be proven (Stadermann 1996, pp. 78-79, 87-91). Though there is generally a positive correlation between the increase of the quantity of money and inflation (Hagen 2004), this correlation is not a 1:1 correlation, which the theory of comparative advantage not only suggests but also depends on.

The link between wage rates and the trade balance in a system with fixed exchange rates has similar theoretical problems and no empirical support. In comparing eleven nations Kaldor (1978) rather found an opposite link, namely that wage rates and export shares of a nation are positively correlated.²⁰ More recent examinations show mixed results (Reichel 2002, p. 333).

Likewise the adjustment mechanism via exchange rates in case of floating exchange rates has theoretical weaknesses. First, current accounts are not the sole determinant of exchange rates. The value of a currency is influenced by financial and currency markets on the basis of supply and demand (Harvey 1995). Supply and demand are in turn determined not only by trade factors but also by financial factors as well as rational and irrational expectations. Therefore, it must be recognised that exchange rates are mainly independent from trade imbalances.

But even if one allows for a moment that exchange rates are influenced by those imbalances, there are theoretical difficulties. According to this mechanism, a trade deficit of a nation leads to a depreciation of its currency. However, devaluation effects on the current account balance are ambiguous (Dornbusch 1995, p. 25). A devaluation leads to a fall in the demand of imported goods and a rise in the demand of exported

¹⁹ The barter assumption is not a necessary assumption for the theory of comparative advantage. However, it is hard to find any other justification why trade should be balanced all the time.

²⁰ This “‘perverse’ relationship between changes [...] in labour costs per unit of output and changes in export performance” (Kaldor 1978, p. 102) became known as ‘Kaldor paradox.’

goods because the former become pricier at home and the latter become cheaper abroad. However, the value of imported goods may increase, decrease or stay the same. The same is true for the value of exported goods. The overall effect for the trade balance is unknown in advance and depends on the price elasticity of supply and demand. Hence, even if the exchange rate reacted to trade imbalances as the theory predicts, these exchange rate changes could even increase trade imbalance.

Neoclassical economists have 'solved' this problem with the so-called Marshall-Lerner condition, which excludes cases in which devaluation does not lead to a reduction of a trade deficit.²¹ However, this only means that in order to sustain the adjustment mechanism via exchange rates, the theory departs further from reality since "it is widely acknowledged that there is no theoretical reason why a depreciation need have any particular effect on the trade balance" (Rose 1991, p. 301). Empirically, the Marshall-Lerner condition is not confirmed (Rose 1991).

To summarise, it exists rather a "remarkable relative inaction of trade balances with respect to exchange rate changes" (Brauer 2003, p. 12).²² Results of empirical studies do not indicate a general, unambiguous causal link between the exchange rate and the trade balance.²³

Unsurprisingly, balanced trade is only an exception in practise. Even strong defenders of the theory of comparative advantage like Krugman and Obstfeld have to admit that "in reality, a country's foreign trade is exactly balanced only rarely" (2009, p. 295). Internationally, trade imbalances that can be large and persistent are common (Milberg 2001, pp. 407-408).²⁴ Additionally, financial crises like the Asian crisis in the 1990s show "that balance of payment disequilibria are seldom benign and self-adjusting" (Felipe and Vernengo 2002, p. 71). It can be concluded that the "empirical evidence has been quite unsupportive of the comparative advantage hypothesis" (Shaikh and Antonopoulos 2013, p. 203).

The definition of international trade as barter trade is underlying the assumption of balanced trade. This, however, is not a useful simplification, but it leads necessarily to unrealistic analysis and predictions, mainly because it misrepresents the role of money, or rather is a "belittlement of money" (Dillard 1988, p. 306). A capitalistic world is not characterised by a barter economy but by a market economy. The main difference in the context of international trade is that "sales and purchases need no longer coincide. The seller does not necessarily have to buy at once" (Sardoni 1986, p. 422). Thus, money is not only a means of exchange. It is foremost a means of payment and is also used to store value, settle debts, transfer wealth and make extraordinary payments. Thus, contrary to the barter assumption of the theory of comparative advantage, money is not like any other commodity. It is rather of "practical importance [to possess] specifically money rather than any commodities at all" (Lapavistas 1996, p. 67). Moneymaking is the main purpose of every economic activity in a market economy. In the context of this article, it is especially important that "money as a store of value in a world of uncertainty does affect motives and decisions of wealth-holders and wealth-producers in a significant way" (Dillard 1988, p. 300). Therefore, nations can take an advantage if they have a trade surplus and thus accumulate capital domestically. Since money can be used as storage of wealth, it is obvious why a nation favours a trade surplus over balanced trade or a trade deficit. A trade surplus is generally seen as a characteristic of an economically successful nation (Pasinetti 1988, p. 140). A trade deficit on the other hand might mean that reserves are lost which worsen a nation's

²¹ A short version of the Marshall-Lerner condition is that it "indicates a stable foreign exchange market if the sum of the price elasticities of the demand for imports and the demand for exports, in absolute terms, is greater than 1" (Salvatore 2011, p. 563).

²² Even in cases where the Marshall-Lerner condition holds, exchange rate changes do not cause trade balances to react as the theory predicts (Brauer 2003, p. 12). There are empirical studies that support the Marshall-Lerner condition, but there are also studies that reject this condition. Therefore, there is no unambiguous empirical support for it, verifying the presumption that this condition may sometimes hold and sometimes not. But it cannot be inferred that the Marshall-Lerner condition should be used in theoretical models. For a list of empirical studies on the Marshall-Lerner condition see Shirvani and Wilbratte (1997). A more recent study that reinforces this ambiguous result can be found in Mahmud et al. (2004).

²³ For an overview of some empirical studies see Reichel (2002, pp. 328-29). A study of developing countries by Rose (1990) confirms these findings.

²⁴ Therefore, the term 'chronic' to describe a continuing trade deficits or surpluses is widespread. This label emphasises that there are no indicators which suggest that it is only a temporary phenomenon. Rather, chronic in this context is used "in the clinical sense of causality related to inherent systemic problems" (Preeg 2000, p. 2).

situation (Felipe and Vernengo 2002, p. 65).²⁵ It can be concluded that “the trade surplus country accrues liquid assets: there is no reason to assume these will be converted into non-liquid assets, much less into foreign-produced non-liquid assets. Saving is thus the mechanism which creates the possibility of [...] persistently unbalanced trade” (Milberg 2002, p. 242).

As was shown, there is neither a theoretical basis nor empirical support to assume balanced trade. This assumption is not a useful simplification but a distortion of the reality.²⁶ It ignores the influences of money and the (un)desirability of trade surpluses and deficits, which are important in the context of international trade.

By giving up this assumption, one would not only increase one’s understanding of international trade, but also have to drop the theory of comparative advantage. Without balanced trade and an adjustment mechanism in case of trade imbalances, there would be no transformation process which converts comparative production differences into absolute money price differences. Since trade imbalances are the norm and an adjustment mechanism neither exists in practise nor is it conclusive in theory, it has to be concluded that comparative advantages are not transformed to price differences and therefore, they cannot explain international trade flows.

Specialisation and gains from trade

The third assumption is that the gains that result from international trade are only static gains in the form of more goods being produced. Consumption is increased and the needs of the population are satisfied to a higher degree. This can be achieved because international trade leads to a more efficient use of the existing resources. Ricardo does not consider dynamic gains in his formulation of the theory of comparative advantage and he does not integrate growth into this theory. While Ricardo neglects technical progress as a result of trade, neoclassical models like the Heckscher-Ohlin model assume identical technology in all countries and thus exclude the possibility of technical change *ex ante*. In this model each nation is ‘naturally’ endowed with factors of production that are not affected by international trade, neither quantitatively nor qualitatively. Ohlin states that international trade only “mitigates the disadvantages of the unsuitable geographical distribution of the productive facilities” (1933, p. 42). As Bhagwati notes, neoclassical theory belongs “to the realm of ‘statics’” (1964, p. 4). An improvement in static efficiency of existing resources is the only benefit from international trade. These gains are merely once-and-for-all. New gains can only be achieved if trade is expanded since existing trade yields no further gains. Moreover, this allocation of resources can completely be reversed in case that international trade ceases or a state cuts its trade links. A nation can simply re-allocate its resource to the state that existed in autarky and easily return to the production level that it had before it entered into trade.

The theory of comparative advantage has never incorporated dynamic gains that lead to technical change and productivity growth which are not exogenously given but which are endogenous, i.e. the result of trade. Moreover, it is difficult to coalesce both, the theory of comparative advantage and endogenous development variables, because the former depends on static comparisons and predictable results. This is not changed by the so-called “dynamic comparative advantage”. In those models comparative advantages develop and change over time. This change, however, is not a result of trade itself but of a change in exogenous factors. These exogenous changes can come from population growth, industrial policies, the rate of capital accumulation (propensity to save) and technological inventions among others. However, the dynamics of these models originate from outside the realm of trade. They constitute models in which comparative advantages are influenced by exogenous changes rather than endogenous dynamic models of comparative advantage.

The underlying foundation behind these static models is what Buchanan and Yoon (2002) call

²⁵ However, a trade deficit is not necessarily bad for a nation. It “may be welcomed if it is due to a high rate of investment at home, which is developing resources that will yield a surplus of exports in the future to repay the debt. But a deficit that is due merely to competitive weakness is highly unfavourable” (Robinson 1977, pp. 1334-35).

²⁶ Stein argues in regard to neoclassical exchange rate theories: “Although most scholars are aware of the deficiencies of these models, the profession continues to use them wholly or partly because they do not have a logically satisfactory substitute” (Stein 1997, p. 228).

“Ricardian logic of trade” which is implied in the theory of comparative advantage. It is supposed that “specialization and subsequent trade become advantageous because of the inherent differences among potential trading parties” (Buchanan and Yoon 2002, p. 400). These differences between nations are not influenced by trade but remain constant after trade started. Hence, the source of comparative advantage is exogenously given, be it productivity differences (Ricardo), different endowments (Heckscher-Ohlin model) or something else. Trade “will occur and will be beneficial whenever countries’ relative prices would be different without trade” (Kenen 2000, p. 37).

In reality, the world and especially the industrialised nations are characterised by enormous technical growth that led to a huge increase in the standard of living and the wealth of the industrialised world. Dynamic gains are more important than mere static gains that arise from effective allocation of given resources. Skarstein calculates that the increase in production in Ricardo’s England-Portugal example amounts to 10% in wine production and 6.25% in the production of cloth. “Nothing more happens” (2007, p. 352).²⁷ This, however, is already an optimistic calculation since it assumes that both nations specialise fully and that both nations started with exactly the same output, which is not stated by Ricardo. Compared to the growth rates of developed nations in the last decades, this “once-and-for-all effect of specialisation under free trade is quite insignificant” (Skarstein 2005, p. 358). Additionally, states cannot easily return to autarky as the statement that these static gains are reversible suggests. Such a move would be costly for a nation and its economy because international trade also produces dependencies.

Since technological change and productivity growth are an important economic factor, there is no logical reason to disregard or even exclude them *ex ante* in a theory of international trade. Neither is there any reason for regarding dynamic changes only as exogenous to trade. Rather they should play a role in theory that corresponds to their significance in the real world. The Ricardian logic of trade risks confusing cause and effect. Nations differ, but trade and specialisation lead to even more differences which enhances trade further. Additionally, specialisation leads to productivity growth due to economies of scale that can be exploited and mechanisations. To assume that differences are exogenously given misses the dynamic developments that result from trade. In Ricardo’s example, productivity differences remain the same between nations after they specialised. This, however, is not very realistic. The neoclassical formulation assumes that factors of production are exogenously given. The factor endowment of a nation is even referred to as natural, which suggests that trade flows are determined by “nature” (Milberg 2001, p. 414). In reality these ‘endowments’ are not natural and change in consequence of trade. Labour and land can be influenced by education and fertiliser, for example. Capital, on the other hand, can hardly be seen as an endowment at all because it is nearly always produced (Steedman 1991, pp. 3-6; Subasat 2003, pp. 156-60).

A theory of international trade should include technical progress and dynamic gains that are endogenous to trade, because these gains are much more significant than any static gains (see also Steedman 1991; Williams 1929). They should even be a central issue. By ignoring or excluding them, the theory of comparative advantage disregards questions of central importance.

Full employment of capital and labour

The fourth assumption that is indispensable to the theory of comparative advantage is full employment of both labour and capital. Neoclassical models generally use this assumption. It is necessary for the concept of opportunity costs. If unemployment (or underutilised resources) exists, there are no opportunity costs, because the production of one good can be increased without decreasing the production of the other good. In this case relative costs of a commodity would stay undefined because the commodity could “be produced at no social cost” (Prasch 1996, p. 42). Since comparative advantages are determined by opportunity costs in the neoclassical formulation, these could not be calculated and this formulation would lose its logical basis. Ricardo and later classical economists assume that labour has a tendency towards full employment and that capital is always fully employed in a liberalised economy, because no capital owner will leave his or

²⁷ He comes to this result by arguing that England can produce an additional $\frac{1}{5}$ of its original output of cloth and therefore $\frac{1}{10}$ more than the overall original output of cloth with the 20 workers that England saves as a result of trade. The overall amount of produced wine would rise by 6.25 % correspondingly.

her capital idle but will always be trying to earn a profit from it. That there is no limit to the employment of capital is a consequence of Say's law which presumes that production is only constraint by resources and which is also adopted by neoclassical economists.

From a theoretical point of view, the theory of comparative advantage has to assume that either labour or capital is used at full capacity and resources constrain the production. There are two reasons, the realisation of gains from international trade and the adjustment mechanism. The theory of comparative advantage assumes static gains in form of a more effective resource allocation which can be seen as a consequence of the resource constraint approach. This cannot be reached unless employment of resources has the highest possible level domestically (Felipe and Vernengo 2002, pp. 54-55). If a nation's resources would not be fully employed, production and consumption could be increased domestically without participating in international trade. The whole rationale for the existence of international trade would vanish as well as the possible gains. In this case, a state could even gain more by abstaining from international trade and boosting domestic production because more labour and capital would then be employed and the national income would be increased. Furthermore, if unemployment is theoretically possible, it will also be possibly that international trade leads to job losses. In the case of job losses gains could not be unambiguously specified, because job losses might outweigh the gains (Shaikh 2007, p. 52).

Full employment (of labour) is also a necessary condition for the adjustment mechanism. If changing unemployment levels are allowed for, income can alter. Once income and thus demand can alter, the current account balance will rather be influenced by them than by price level or exchange rate changes. Demand effects are neither included in the quantity theory of money nor in the exchange rate adjustment approach.²⁸ Turnell concludes that "with unemployment allowed to exist in the model, the effect of the initial trade imbalance of the higher cost country is not to bring about price changes, but changes in income (employment) and/or real interest rates" (2001, p. 7). Thus, any adjustment mechanism that underlies the theory of comparative advantage no longer operates if unemployment exists (see also Çağatay 1994; Milberg 2002).²⁹

Theoretically, this assumption is problematic. Once it is allowed that money can also be used as storage of wealth (and not just as a means of exchange) one has to conclude that there is no tendency towards full employment of capital and of labour. The possibility of saving "creates the possibility of [...] underemployment" (Milberg 2002, p. 242). Hence, there is no theoretical justification for this assumption.

In practise, the "world is characterized by unemployment" (Felipe and Vernengo 2002, p. 54). Un- and underemployment of capital and labour is not a short run phenomenon but it is common and widespread. In the last decade between 175 and 200 million workers have been unemployed worldwide (ILO 2012). If underemployment is added, this figure rises to a much higher number. Even in the fifteen most economically liberalised nations, unemployment rates have ranged between 1.0% and 16.6% in the last two decades.³⁰ Similar, a nation has usually "productive capacity for more output than it can sell" (Robinson 1973, p. 15). It has to be concluded that unemployment and idle resources are rather the rule than the exception.

The assumption of full employment is, as was shown above, crucial to the theory of comparative advantage. Without this assumption, the supposed gains from international trade, namely a higher amount of available products that the population can consume, can be achieved without engaging in international trade. This means, that the explanation of the theory of comparative advantage why international trade takes place is itself invalid if unemployment to exist. Though it might be reasonable to use this assumption in other

²⁸ The reason is that international trade merely leads to a change in the composition of the production according to the theory of comparative advantage. National employment and income levels have the same level with and without trade in all nations (Turnell 2001, p. 8).

²⁹ It might be added that, strictly speaking, the realisation of gains and the adjustment mechanism are not dependent on full employment but on the assumption of a constant rate of unemployment is sufficient. However, once the possibility of unemployment is included it can hardly be argued, from a theoretical point of view, that the unemployment level will be constant at five per cent or twenty per cent. Rather, it must be assumed that the level changes.

³⁰ These nations include Australia, Canada, Chile, Denmark, Finland, Germany, Hong Kong, Ireland, Luxembourg, the Netherlands, New Zealand, Singapore, Switzerland, the UK and the United States. They were the fifteen most liberalised nations on average between 1990 and 2009 according to Economic Freedom of the World Report (Gwartney et al. 2011). The unemployment data is taken from IMF (2011).

economic models, it is inappropriate for the theory of comparative advantage because the whole motive of international trade collapses if this unrealistic assumption is given up.

International harmony

The fifth assumption is less an assumption in itself but a consequence of the other four assumptions. The theory of comparative advantage depicts international trade as a harmonious exchange. By definition, every nation must have a comparative advantage in the production of certain goods. Since these comparative production advantages are transformed through an adjustment mechanism into absolute price advantages, each nation can sell some goods, or at least one good, cheaper than all other nations and can thus “successfully compete in world markets, regardless of the degree of efficiency of its technology or resource-base” (Jones 1980, p. 235). There is no need for nations to focus on their competitiveness. Even a nation that lacks in productivity and is technologically backward is still competitive in the world market and can balance its imports and exports. As a consequence “international trade is not about competition, it is about mutually beneficial exchange” (Krugman 1997, p. 120).

In this way, international trade is described as an even-handed international division of labour. A “homeostatic view” prevails, in which a “natural pattern of specialization and trade” exists. In case of any deviation from this ‘natural’ order, “automatic forces tend to restore this natural pattern” (Krugman 1987, p. 41). In this idyllic picture, every nation produces those goods that it is able to produce comparatively best. Domestically, entrepreneurs as well as regions compete with each other, which leads to winners and losers. In contrast, international trade leads to “the happy result that all countries will be able successfully to participate in international trade in the sense that they will benefit from such trade and be able to generate export revenues equal to the value of imports” (Milberg 2004, pp. 56-57).

Underlying this view of international trade is the assumption that trade is always balanced which is the result of equating international trade with barter trade. Balanced trade and an adjustment mechanism lead to an equalisation in competitiveness among nations. Consequently, Ricardo assumes that both nations gain from the productivity growth in one nation because such an improvement “raises general prices in the country where the improvement takes place” (Ricardo 2004b, p. 141). According to the neoclassical formulation, productivity growth in one nation leads to an appreciation of its currency. Its higher productivity is balanced by a disadvantageous movement of the exchange rate. Nations are made, without any policy or other interference, equally competitive by exchange rate movements – independent of their technology, their resources, their workforce and their level of development. Free trade is beneficial from both a global and a national point of view. In this static analysis, all nations win. Moreover, the factor price equalisation theorem even postulates that differences in real wages are reduced and are finally eliminated. Differences in real earnings per head in all free trading nations will be reduced or even vanish. Therefore, neoclassical economists argue that poor nations have the opportunity to develop and catch up with developed nations. The wage level in poor nations would conform to the wage level of rich nations. Free trade alone has “the potential for development and convergence between rich and poor countries” (Kiely 2007, p. 15). Free trade is thus a great equaliser.

The world according to the theory of comparative advantage can be described by paraphrasing Karl Marx (1922, p. 31): from each nation according to its faculties, to each nation according to its needs!³¹ What Keynes states about mainstream economics in general, is certainly true in respect to the theory of comparative advantage: it “represents the way in which we should like our economy to behave. But to assume that it actually does so is to assume our difficulties away” (1936, p. 34). Obviously, this harmonic view does not reflect the real world. In a world that is dominated by market economies, companies compete internationally as well as nationally. Though I am dealing with international trade and the emphasis is on nations, it should not be forgotten that companies and not nations sell goods. Both do not build a unified subject when it comes to international trade. A nation has other aims than a firm. Ricardo is concerned with the distribution of income between the different classes. However, in the case of international trade he

³¹ Marx uses the original phrase to describe the “higher phase of communism” and not a market economy.

assumes that the interests of merchants and of consumers do not contradict with the interests of their nations. Neoclassical theory neglects different particular interests and is only concerned with the nation as a whole.

Companies export and import not for the benefit of their nations of origin but for the sake of their own profit. However, competition always means a rivalry between companies which can increase their own profits and their market shares by beating their rivals and squeezing them out of the market.³² Profit making and not consumption, as suggested by the barter assumption, is the motive of production in a market economy. If a firm can increase its profit by producing fewer goods, it will do so because of its pecuniary incentives. "The firm is dealing throughout in terms of sums of money. It has no object in the world except to end up with more money than it started with. That is the essential characteristic of an entrepreneur economy" (Keynes 1979, p. 89). The success of a company is measured by their profits and their capital accumulation.

Nations on the other hand also compete with each other, but in a different way than companies. This, however, is disguised by the theory of comparative advantage. Once its assumptions are rejected, one can also gain an insight into the competition between nations. If production factors are allowed to move internationally, nations will compete for capital (investments) and talent (highly-educated workers). If trade imbalances and unemployment are considered, a trade surplus becomes favourable for a nation, because it generally means competitive success, a higher level of production, a lower level of unemployment and an inflow of money, which can be invested as capital or stored as wealth. Successful export industries provide for employment and profits. Success in international competition is in the economic interest of the state (or to be more precise its leaders), because the mentioned effects lead to increases in the state revenue – either because the potential tax basis is increased or the welfare expenditure can be cut without decreasing the economic potential of the society. All this has consequences beyond the economic sphere, because nations also compete for power, such as political influence, military and bargaining power, which are all influenced by their economic success (Dunn 1994).

Contrary to the claims of the theory of comparative advantage that national competitiveness is always evened out, success in economic competition leads to further success, while failure often brings about more failure. Kaldor calls this phenomenon "polarisation process". This effect "is nothing else than the inhibiting effect of superior competitive power of industrially more efficient and dynamic countries, as compared to others" (Kaldor 1981, p. 597).³³ Therefore, some nations benefit more from free trade while others benefit less or might even suffer losses depending on their level of development. Kaldor concludes that "under more realistic assumptions unrestricted trade is likely to lead to a loss of welfare to particular regions or countries" (Kaldor 1981, p. 593). It is not surprising that a growing market share of a nation is "strongly positively correlated" with faster productivity growth and an increase in technological capability (Fagerberg 1996, pp. 40-41). Furthermore, the productivity and success of a production factor is not independent of other production factors, but not in the neoclassical sense of substitutability. Rather, production factors ordinarily complement each other. A lack of capital leads to higher unemployment, while more capital generally leads to increasing employment levels. A relative scarcity of capital does not lead to a competitive advantage in labour intensive goods, as the theory of comparative advantage states.

Additionally, the theory of comparative advantage assumes, in accordance with this harmonic view, that it does not matter in which goods a nation specialises. A nation benefits equally whether it specialises "in apples or Apple computers" (Çağatay 1994, p. 241). However, contrary to this assumption, in reality it has different consequences for a nation if it specialises in agriculture or in industry. Industrial nations are usually richer and economically more developed than agrarian nations. Theoretically, this conclusion is supported by Verdoorn's and Kaldor's growth laws which have wide empirical support. These three laws state that there are strong positive correlations (a) between economic growth and growth of the

³² Competition should not be understood in the neoclassical sense of perfect competition that likewise suggests a harmonic world. Rather, as Morgenstern points out, the meaning of competition "is one of struggle with others, of fight, of attempting to get ahead, or at least to hold one's place" (Morgenstern 1972, p. 1164). The neoclassical term 'free' or 'perfect' competition is thus "a complete misnomer" (Morgenstern 1972, p. 1171).

³³ This effect takes place internationally as well as domestically: "free trade is as much a mechanism for the concentration and centralization of international capital as free exchange within a capitalist nation is for the concentration and centralization of domestic capital" (Shaikh 1980, p. 227).

manufacturing sector (Kaldor 1967); (b) “between the growth of labour productivity and the volume of industrial production” (Verdoorn 2002, p. 28); (c) between growth of manufacturing and growth of the productivity outside manufacturing. Additionally, the terms of trade between agricultural goods and manufactured goods decline over time to the detriment of the former (Larrinoa Arcal and Maetz 2000).

5. Implication for International Trade Policy

As was shown, the theory of comparative advantage is an inadequate theory for the explanation and understanding of international trade because of its insufficient and deceptive assumptions. However, the most crucial consequences of a theory do not arise from theoretical inadequacy but from its utilisation in practical policies. If economic policies are based on a deficient theory, they can become harmful and lead to unintended consequences. That economic theories have huge impact on political decision-makers is unquestionable. As Keynes notes at the end of his *General Theory*, “the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood [...] soon or late, it is ideas, not vested interests, which are dangerous for good or evil” (1936, pp. 383-84). This is especially true for simple theories that have clear-cut predictions. They can easily be absorbed by policymakers and others outside the theory’s subject area. For policymakers, it is easy to refer to and rely on simplistic theories because these theories help to make one’s point and to convey and justify policy measure. More complex theories, which do not offer such unambiguous conclusions about a policy outcome, are less suitable thereto. This means that the danger of deficient theories is not only its theoretical falsity but also, and even more importantly, that it is used as the theoretical basis to solve policy issue.

It is not surprising, considering its widespread support, that the theory of comparative advantage was and still is influential in shaping international trade policies. International trade liberalisations, which comprise mainly the removal of trade barriers and which are pursued by the WTO, the IMF and the World Bank, are strongly influenced by it (Shaikh 2007; Shell 1996). It is argued that trade liberalisation is the best way in which “countries can benefit from comparative advantage-driven trade” (Kowalski 2011, p. 5). The WTO frequently refers to the comparative advantages as the theoretical justification to its commitment to free trade and to its attempt to achieve this objective. On its homepage the WTO bases “the case for open trade” on the theory of comparative advantage. It appraises this theory as “arguably the single most powerful insight into economics” (WTO 2013). Paul Lamy, the Director-General of the WTO, defended the theory of comparative advantage in a recent speech against criticism (Lamy 2010).

In contrast to the focus of WTO policies on trade liberalisation, the objectives of the WTO, which are stated in the preamble of its founding agreement (Marrakesh Agreement), include “raising standards of living, ensuring full employment and a large and steadily growing volume of real income and effective demand, and expanding the production of and trade in goods and services” (WTO 1994). However, the WTO is mainly concerned with free trade and reduction of trade barriers and less with the achievement of its other objectives (Ismail 2005; Rodrik 2001). This can be explained by the reliance on the theory of comparative advantage, which leads to the conclusion that the other aims, most notably national welfare, can be best achieved if trade is liberalised and protectionists policies are abolished. This means that other policy measures can be disregarded. In this way, free international trade becomes an end in itself and is not merely a mean to achieve other objectives. The other aims of the WTO are perceived only in the aftermath and as a result of trade (Rodrik 2001). Its recommendation and demands emphasise measures to reduce unrestricted trade such as abolishing tariffs and nontariff barriers and the elimination of subsidies (Shaikh 2007, p. 61).³⁴

As discussed above, the theory of comparative advantage neglects important issues of international trade. It is therefore not surprising that policies that rely on this theory also disregard these issues. Empirical studies do not support the predicted link of trade liberalisation and increasing living standards. Rodrik, for example, evaluates empirical studies and concludes “that there is no convincing evidence that trade liberalization is predictably associated with subsequent economic growth” (2001, p. 11). There are many countries that followed the recommendations of the WTO and liberalised their trade and that suffered as a

³⁴ An example of the neglect of its other objectives, which are stated in the preamble of the Marrakesh Agreement, is given by the Dispute Settlement Body of the WTO, which does not consider these objectives in its rulings over disputes on trade restrictions.

consequence of this policy. On the other hand, there are nations that had used protectionist trade policies to become successful in international markets (Shaikh 2007). This is not surprising considering the analyses above. One issue that is especially important in the ongoing and future trade policies is the assumption of the theory of comparative advantage, that trade compensates for international differences and is a great equaliser. It is assumed that free trade levels the playing field. This is especially crucial with regard to poor developing nations. It is uniformly argued, both by economists and politicians, that these nations possess a comparative advantage in agricultural goods and that they should therefore specialise in agricultural production. However, this disregards major ramifications and will misstate the opportunities of such a specialisation. As was shown above, it does matter whether a nation specialises in agriculture or manufacturing. Nations that specialise in agriculture have a lower growth rate. As Skarstein points out, “the industrial sector is the dynamic centre of technical change and productivity growth” (2007, p. 353) in any economy. This means that nations that specialise in agriculture will fall ever further behind industrial nations. This development is even amplified if a nation deindustrialises as a result of its specialisation in agriculture. Additionally, nations that export mainly agricultural goods are faced with tendency of declining terms of trade between agricultural goods and manufactured goods. This leads to trade deficits, which will not be balanced out automatically as the theory of comparative advantage suggests, but rather disadvantages these nations even further and leads to an accumulation of debts. Moreover, the fact that and the way in which nations compete with each other, as described above, is completely neglected. These negative developments of specialisation are not regarded by the theory of comparative advantage and trade policies that are derived from this theory risk to disregard them and to draw the wrong conclusions. If poor developing nations specialise in agriculture it might have some positive effects but the negative effects have to be taken into consideration, at least to be able to mitigate them. To get a better understanding of international trade and its ramifications neither economists nor politicians should rely on a misleading theory as the theory of comparative advantage.

6. Conclusion

After scrutinising the essential assumptions, on which the theory of comparative advantage is based – regardless of its specific formulation – it has to be concluded that these assumptions are inadequate and can be rejected on theoretical, logical and empirical grounds. Economic models are always a simplification from the real world and an abstraction from empirical phenomena. However, the assumptions made by the theory of comparative advantage are not valid and justifiable simplifications. Rather, this theory reproduces international trade in a way that is theoretically objectionable, contradicts empirical analysis and falsifies the reality.

The abandonment of the first two assumptions, international immobility of capital and labour and the existence of an adjustment mechanism that leads to balanced trade, leads to the conclusion that the presumption of an automatic transformation from comparative production advantages into absolute price advantages is untenable. Therefore, comparative advantages cannot determine international trade patterns. The abandonment of the third and fourth assumption, the static nature of the gains from trade and full employment, makes the claim invalid that free trade is necessarily beneficial. Additionally, the reason or motive for international trade, which consists of the benefits, is made obsolete. This questions the whole rationale behind the theory of comparative advantage. The scrutiny of the fifth assumption, harmony, emphasises the irrelevance of the theory of comparative advantage in respect to trade between capitalistic nations. This theory glorifies and misrepresents trade between market economies rather than describes it realistically.

To summarise, it has been shown that the theory of comparative advantage is not a useful and adequate theory that explains the patterns of and the underlying reasons behind free international trade. The ‘gains’ from its assumptions are predictability and simplicity. However, the ‘costs’ of these gains are too high for any theory, which consist of the irrelevance for the explanation of international trade. Free international trade between market economies does not take place as the theory of comparative advantage suggests. Therefore, this theory should be abandoned in favour of a theory that is logically and theoretically sound and

presents a justifiable simplification of reality. Such a theory would also take into consideration the relevant phenomena and problems that are connected to free international trade but disregarded by the theory of comparative advantage. Additionally, policies that rely on this theory are at least questionable and risk drawing wrong conclusion. This is emphasised by the scrutiny of the argument that poor developing nations should specialise in agriculture, which is a commonly brought forward and is also used in the WTO trade talks. This scrutiny shows that trade policies based on the theory of comparative advantage misstate the effects and implications of trade liberalisations.

Despite its theoretical and empirical problems and misassumptions, the theory of comparative advantage enjoys still a widespread acceptance in mainstream economics. It is “one of the least controversial ideas in economics” (The Economist 2011, p. 5). When difficulties arise, there are three ways of response. First, empirical difficulties are rejected on the grounds that states intervene and hinder trade to be free. The theory is defended and inconsistencies with the real world are blamed on political interferences. Second, theoretical as well as empirical difficulties are explained by short run phenomena that do not affect the long run predictions of the theory of comparative advantage. Third, economists complement it by modifying its assumption and formulate exemptions – as is done for example by the New Trade Theory.

All three responses, however, are deficient. Political intervention exists in the real world but that would not invalidate the theory of comparative advantage. Trade barriers such as tariffs can be added to the production costs and thus change comparative advantages of nations. The theory itself, however, would still hold in the face of those interventions. This was already recognised by Ricardo who argued that taxes hinder the most efficient use of the factors of production worldwide but do not render the theory of comparative advantage obsolete (Ricardo 2004b, p. 172). The second argument, that the theory of comparative advantage explains only long run developments but short run phenomena might contradict its predictions, is not convincing once it becomes clear that long run might mean seventy-five year or longer (Froot and Rogoff 1995, p. 1657). In such a period, international trade patterns will change significantly and allegedly long-term trends are all but insignificant. In this way, the theory itself reveals its irrelevance in explaining anything that matters in international trade. The modification or exemptions like economies of scale and ‘imperfect’ competition that complement the theory of comparative advantage where its shortcomings are obvious, amplifies this irrelevance.³⁵ International trade theory, by being based on the theory of comparative advantage, runs the risk of ignoring everything that is important in determining trade patterns. What Joan Robinson concluded nearly forty years ago is still true today, namely that “there is no branch of economics in which there is a wider gap between orthodox doctrine and actual problems than in the theory of international trade” (Robinson 1973, p. 14).

This article does not give judgement in favour or against free international trade – and it has no such intentions. Free trade as a phenomenon is not the object of study. Rather, the dominating theory that tries to explain this phenomenon, and on which present implementation of trade liberalising theories is based, is the object of critique. After deconstructing the theory of comparative advantage and its assumptions, it cannot be endorsed any longer. Not only were the assumptions dismissed in this article, but it was also shown that the theory of comparative advantage is wrong on its own terms. The obvious suggestion is to dismiss the theory of comparative advantage after nearly 200 years in order to get a better and sounder understanding of international trade.

³⁵ Additionally, the appearance of those exceptions like increasing returns contests rather than enhances the theory of comparative, because, according to it, every transaction influences international trade. An unproblematic coexistence of comparative and noncomparative advantage trade is theoretically more than questionable.

References

- Aldrich, J (2004), The Discovery of Comparative Advantage, *Journal of the History of Economic Thought*, 26 (3), pp. 379-399.
- Alston, R M, J R Kearl and M B Vaughan (1992), Is There a Consensus among Economists in the 1990's?, *The American Economic Review*, 82 (2), pp. 203-209.
- Anspach, R (1968), The Myth of Absolute Advantage: Comment, *The American Economic Review*, 58 (3), pp. 514-516.
- Bhagwati, J (1964), The Pure Theory of International Trade: A Survey, *The Economic Journal*, 74 (293), pp. 1-84.
- Blaug, M (1977), *Economic Theory in Retrospect*. 2nd ed, London: Heinemann.
- Brauer, H (2003), *The Real Exchange Rate and Prices of Traded Goods in OECD Countries*, Berlin: Springer.
- Brewer, A (1985), Trade With Fixed Real Wages and Mobile Capital, *Journal of International Economics*, 18 (1-2), pp. 177-186.
- Buchanan, J M and Y J Yoon (2002), Globalization as Framed by the Two Logics of Trade, *The Independent Review*, 6 (3), pp. 399-405.
- Çağatay, N (1994), Themes in Marxian and Post-Keynesian Theories of International Trade: A Consideration with Respect to New Trade Theory, In: Mark A. Glick (Ed.). *Competition, Technology, and Money: Classical and Post-Keynesian Perspectives*, Aldershot: Edward Elgar, pp. 237-250.
- Cairnes, J E (1874), *Some Leading Principles of Political Economy Newly Expounded*, London: Macmillan.
- Chipman, J S (1965), A Survey of the Theory of International Trade: Part 2, The Neo-Classical Theory, *Econometrica*, 33 (4), pp. 685-760.
- Dillard, D (1988), The Barter Illusion in Classical and Neoclassical Economics, *Eastern Economic Journal*, 14 (4), pp. 299-318.
- Dornbusch, R (1995), *Exchange Rates and Inflation*, Cambridge: MIT Press.
- Dunn, M H (1994), Do Nations Compete Economically?, *Intereconomics*, 29 (6), pp. 303-308.
- Eicher, T S, J H Mutti and M H Turnovsky (2009), *International Economics*. 7th ed, London: Routledge.
- Fagerberg, J (1996), Technology and Competitiveness, *Oxford Review of Economic Policy*, 12 (3), pp. 39-51.
- Felipe, J and M Vernengo (2002), Demystifying the Principles of Comparative Advantage: Implications for Developing Countries, *International Journal of Political Economy*, 32 (4), pp. 49-75.
- Findlay, R (1987), Comparative Advantage, In: John Eatwell, Murray Milgate, Peter Newman and Robert Harry Inglis Palgrave (Eds.), *The New Palgrave: A Dictionary of Economics, Volume 1: A - D*, London: Macmillan, pp. 514-517.
- Frey, B S, W W Pommerehne, F Schneider and G Gilbert (1984), Consensus and Dissension among Economists: An Empirical Inquiry, *The American Economic Review*, 74 (5), pp. 986-994.
- Froot, K A and K Rogoff (1995), Perspectives on PPP and Long-Run Real Exchange Rates, In: Gene M. Grossman and Kenneth Rogoff (Eds.), *Handbook of International Economics: Volume 3*, Amsterdam: Elsevier, pp. 1647-1688.

Fuller, D and D Geide-Stevenson (2003), Consensus among Economists: Revisited, *The Journal of Economic Education*, 34 (4), pp. 369-387.

Gomes, L (2003), *The Economics and Ideology of Free Trade: A Historical Review*, Cheltenham: Edward Elgar.

Gwartney, J, J Hall and R Lawson (2011), *2011 Economic Freedom Dataset*, published in *Economic Freedom of the World: 2010 Annual Report*, Economic Freedom Network, <http://www.freetheworld.com/2011/2011/Dataset.xls> accessed 18.03.2012.

Haberler, G (1929), The Theory of Comparative Cost Once More, *The Quarterly Journal of Economics*, 43 (2), pp. 376-381.

Haberler, G (1930), Die Theorie der Komparativen Kosten und Ihre Auswertung für die Begründung des Freihandels, *Weltwirtschaftliches Archiv*, 32, pp. 349-70.

Hagen, J v (2004), Hat die Geldmenge ausgedient?, *Perspektiven der Wirtschaftspolitik*, 5 (4), pp. 423-453.

Harrigan, J (2003), Specialization and the Volume of Trade: Do the Data Obey the Laws?, In: E. Kwan Choi and James Harrigan (Eds.), *Handbook of International Trade*, Malden: Blackwell, pp. 85-118.

Harvey, J T (1995), The International Monetary System and Exchange Rate Determination: 1945 to the Present, *Journal of Economic Issues*, 29 (2), pp. 493-502.

Heckscher, E (1949) [1919], The Effect of Foreign Trade on the Distribution of Income, In: Howard S. Ellis and Lloyd A. Metzler (Eds.), *Readings in the Theory of International Trade*, Philadelphia: Blakison, pp. 272-300.

Hume, D (1903a) [1752], Of Money, In: David Hume (Ed.). *Essays, Moral, Political and Literary*, London: Richards, pp. 289-302.

Hume, D (1903b) [1752], Of the Balance of Trade, In: *Essays, Moral, Political and Literary*, London: Richards, pp. 316-333.

ILO (2012), *Global Employment Trends 2012: Preventing a Deeper Jobs Crisis*, Geneva: ILO.

IMF (2011), *World Economic Outlook Database: April 2011*, Data retrieved from <http://www.imf.org/external/pubs/ft/weol/2011/01/weodata/download.aspx> accessed 18.03.2012.

Ismail, F (2005), Mainstreaming Development in the World Trade Organization, *Journal of World Trade*, 39 (1), pp. 11-21.

Jones, R W (1980), Comparative and Absolute Advantage, *Schweizer Zeitschrift für Volkswirtschaft und Statistik*, 116 (3), pp. 235-260.

Jones, R W (2000), *Globalization and the Theory of Input Trade*, Cambridge: MIT Press.

Kaldor, N (1967), *Strategic Factors in Economic Development*, Ithaca: New York State School of Industrial and Labor Relations, Cornell University.

Kaldor, N (1978), The Effect of Devaluations on Trade in Manufactures, In: Nicholas Kaldor (Ed.). *Further Essays on Applied Economics*, London: Duckworth, pp. 99-116.

Kaldor, N (1981), The Role of Increasing Returns, Technical Progress and Cumulative Causation in the Theory of International Trade and Economic Growth, *Économie Appliquée*, 34 (4), pp. 593-617.

Kearl, J R, C L Pope, G C Whiting and L T Wimmer (1979), A Confusion of Economists?, *The American Economic Review*, 69 (2), pp. 28-37.

- Kenen, P B (2000), *The International Economy*. 4th ed, Cambridge: Cambridge University Press.
- Keynes, J M (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan.
- Keynes, J M (1979) [1933], The Tilton Papers, In: Donald Moggridge (Ed.). *Collected Writings of John Maynard Keynes, Vol. 29: The General Theory and After: A Supplement*, London: Macmillan, pp. 76-106.
- Kiely, R (2007), *The New Political Economy of Development: Globalization, Imperialism, Hegemony*, Basingstoke: Palgrave Macmillan.
- Kowalski, P (2011), *Comparative Advantage and Trade Performance: Policy Implications*, OECD Trade Policy Working Paper, No. 121, <http://dx.doi.org/10.1787/5kg3vwb8g0hl-en>.
- Krugman, P R (1987), The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies, *Journal of Development Economics*, 27 (1-2), pp. 41-55.
- Krugman, P R (1997), *Pop Internationalism*, Cambridge: MIT Press.
- Krugman, P R and M Obstfeld (2009), *International Economics: Theory and Policy*. 8th ed, Boston: Pearson Addison-Wesley.
- Lamy, P (2010), *Comparative Advantage is Dead? Not at all*, http://www.wto.org/english/news_e/sppl_e/sppl152_e.htm accessed 21.01.2013.
- Lapavistas, C (1996), The Classical Adjustment Mechanism of International Balances: Marx's Critique, *Contributions to Political Economy*, 15, pp. 63-79.
- Larrinoa Arcal, Y F de and M Maetz (2000), Trends in World and Agricultural Trade, In: FAO (Ed.). *Multilateral Trade Negotiations on Agriculture: A Resource Manual*, available from <http://www.fao.org/docrep/003/X7351E/X7351E00.HTM>.
- Lerner, A P (1952), Factor Prices and International Trade, *Economica*, 19 (73), pp. 1-15.
- Lutz, M A and K Lux (1988), *Humanistic Economics: The New Challenge*, New York: Bootstrap Press.
- MacDonald, G M and J R Markusen (1985), A Rehabilitation of Absolute Advantage, *Journal of Political Economy*, 93 (2), pp. 277-297.
- Mahmud, S F, A Ullah and E M Yucel (2004), Testing Marshall-Lerner Condition: A Non-Parametric Approach, *Applied Economics Letters*, 11 (4), pp. 231-236.
- Maneschi, A (1998), *Comparative Advantage in International Trade: A Historical Perspective*, Cheltenham: Edward Elgar Pub.
- Marx, K (1922) [1875], *The Gotha Program*, In Daniel de Leon, New York: National Executive Committee, Socialist Labor Party, http://slp.org/pdf/marx/gotha_prgm.pdf.
- Milberg, W (2001), Decentering the Market Metaphor in International Economics, In: Stephen Cullenberg, Jack J. Amariglio and David F. Ruccio (Eds.), *Postmodernism, Economics and Knowledge*, London: Routledge, pp. 407-430.
- Milberg, W (2002), Say's Law in the Open Economy: Keynes's Rejection of the Theory of Comparative Advantage, In: Sheila C. Dow and John Hillard (Eds.), *Keynes, Uncertainty and the Global Economy: Beyond Keynes, Volume Two*, Cheltenham: Edward Elgar, pp. 239-253.
- Milberg, W (2004), The Changing Structure of Trade Linked to Global Production Systems: What Are the Policy Implications?, *International Labour Review*, 143 (1-2), pp. 45-90.

Mill, J S (1929) [1871], *Principles of Political Economy With Some of Their Applications to Social Philosophy*. 7th ed, London: Longmans, Green.

Morgenstern, O (1972), Thirteen Critical Points in Contemporary Economic Theory: An Interpretation, *Journal of Economic Literature*, 10 (4), pp. 1163-1189.

Nayyar, D (2007), Globalization and Free Trade: Theory, History, and Reality, In: Anwar Shaikh (Ed.). *Globalization and the Myths of Free Trade: History, Theory, and Empirical Evidence*, London: Routledge, pp. 69-84.

Ohlin, B (1933), *Interregional and International Trade*, Cambridge: Harvard University Press.

Oswald, I (2007), *Migrationssoziologie*, Konstanz: UTB.

Pasinetti, L L (1988), Technical Progress and International Trade, *Empirica*, 15 (1), pp. 139-147.

Petrella, F (1968), Adam Smith's Rejection of Hume's Price-Specie-Flow Mechanism: A Minor Mystery Resolved, *Southern Economic Journal*, 34 (3), pp. 365-374.

Polanyi, K (2001) [1944], *The Great Transformation: The Political and Economic Origins of Our Time*, Boston: Beacon Press.

Prasch, R E (1996), Reassessing the Theory of Comparative Advantage, *Review of Political Economy*, 8 (1), pp. 37-55.

Preeg, E H (2000), *The Trade Deficit, the Dollar, and the U.S. National Interest*, Indianapolis: Hudson Institute.

Pritchett, L (1997), Divergence, Big Time, *The Journal of Economic Perspectives*, 11 (3), pp. 3-17.

Reichel, R (2002), *Ökonomische Theorie der Internationalen Wettbewerbsfähigkeit von Volkswirtschaften*, Wiesbaden: Dt. Universitäts-Verlag.

Ricardo, D (2004a) [1815], An Essay on the Influence of a low Price of Corn on the Profits of Stock, In: Piero Sraffa (Ed.). *The Works and Correspondence of David Ricardo: Vol. IV Pamphlets and Papers 1815-1823*, Indianapolis: Liberty Fund, pp. 1-41.

Ricardo, D (2004b) [1817], On the Principles of Political Economy and Taxation, In: Piero Sraffa (Ed.). *The Works and Correspondence of David Ricardo: Vol. I*, Indianapolis: Liberty Fund.

Robinson, J (1973), The Need for a Reconsideration of the Theory of International Trade, In: Joan Robinson (Ed.). *Collected Economic Papers, Volume 4*, Oxford: Basil Blackwell, pp. 14-24.

Robinson, J (1977), What are the Questions?, *Journal of Economic Literature*, 15 (4), pp. 1318-1339.

Rodrik, D (2001), *The Global Governance of Trade: As if Development Really Mattered*, New York: UNDP.

Rose, A K (1990), Exchange Rates and the Trade Balance: Some Evidence from Developing Countries, *Economics Letters*, 34 (3), pp. 271-275.

Rose, A K (1991), The Role of Exchange Rates in a Popular Model of International Trade: Does the 'Marshall-Lerner' Condition Hold?, *Journal of International Economics*, 30 (3-4), pp. 301-316.

Ruffin, R J (2002), David Ricardo's Discovery of Comparative Advantage, *History of Political Economy*, 34 (4), pp. 727-748.

Salvatore, D (2011), *International Economics: Trade and Finance*. 10th ed, Hoboken: Wiley.

Samuelson, P A (1948), International Trade and the Equalisation of Factor Prices, *The Economic Journal*, 58 (230), pp. 163-184.

Samuelson, P A (1972) [1969], The Way of an Economist, In: Robert C. Merton (Ed.). *The Collected Scientific Papers of Paul A. Samuelson, Vol. III*, Cambridge: M.I.T. Press, pp. 675-685.

Sardoni, C (1986), Marx and Keynes on Effective Demand and Unemployment, *History of Political Economy*, 18 (3), pp. 419-441.

Seligman, E R A and J H Hollander (1911), Ricardo and Torrens, *The Economic Journal*, 21 (83), pp. 448-468.

Shaikh, A (1980), The Laws of International Exchange, In: Edward J. Nell (Ed.). *Growth, Profits, and Property: Essays in the Revival of Political Economy*, Cambridge: Cambridge University Press, pp. 204-235.

Shaikh, A (2007), Globalization and the Myths of Free Trade, In: Anwar Shaikh (Ed.). *Globalization and the Myths of Free Trade: History, Theory, and Empirical Evidence*, London: Routledge, pp. 50-68.

Shaikh, A and R Antonopoulos (2013), Explaining Long Term Exchange Rate Behavior in The United States and Japan, In: Jamee K. Moudud, Cyrus Bina and Patrick L. Mason (Eds.), *Alternative Theories of Competition: Challenges to the Orthodoxy*, London: Routledge, pp. 201-228.

Shell, G R (1996), Trade Stakeholders Model and Participation by Nonstate Parties in the World Trade Organization, *University of Pennsylvania Journal of International Law*, 17 (1), pp. 359-381.

Shirvani, H and B Wilbratte (1997), The Relationship Between the Real Exchange Rate and the Trade Balance: An Empirical Reassessment, *International Economic Journal*, 11 (1), pp. 39-50.

Skarstein, R (2005), Economic Development by Means of Free Trade?, In: Günther Chaloupek, Arne Heise, Gabriele Matzner-Holzer and Wolfgang Roth (Eds.), *Sisyphus als Optimist: Versuche zur zeitgenössischen politischen Ökonomie*, Hamburg: VSA-Verlag, pp. 352-366.

Skarstein, R (2007), Free Trade: A Dead End for Underdeveloped Economies, *Review of Political Economy*, 19 (3), pp. 347-367.

Smith, A (1904) [1776], *An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. 2*, London: Methuen.

Sraffa, P (1930), An Alleged Correction of Ricardo, *The Quarterly Journal of Economics*, 44 (3), pp. 539-544.

Stadermann, H-J (1996), *Monetäre Theorie der Weltwirtschaft*. 2nd ed, Tübingen: Mohr.

Stalker, P (2000), *Workers Without Frontiers: The Impact of Globalization on International Migration*, Boulder: Lynne Rienner.

Steedman, I (1991) [1979], Introductory Essay, In: Ian Steedman (Ed.). *Fundamental Issues in Trade Theory*, London: Gregg Revivals, pp. 1-14.

Stein, J L (1997), Appendix: International Finance Theory and Empirical Reality, In: Jerome L. Stein and Polly Reynolds Allen (Eds.), *Fundamental Determinants of Exchange Rates*, Oxford: Clarendon Press, pp. 225-255.

Subasat, T (2003), What Does the Heckscher-Ohlin Model Contribute to International Trade Theory? A Critical Assessment, *Review of Radical Political Economics*, 35 (2), pp. 148-165.

The Economist (2009), *So Much Gained, So Much to Lose*, 393 (8656), pp. 13-14.

The Economist (2011), *The Great Mismatch*, Special Report on the Future of Job, 400 (8750), pp. 3-5.

World Economic Review

Turnell, S (2001), 'The Right to Employment: Extending the Core Labour Standards and Trade Debate', *Macquarie Economics Research Papers*, (1/2001), <http://www.econ.mq.edu.au/research/2001/3-2001Turnell.PDF>.

Verdoorn, P J (2002) [1949], Factors that Determine the Growth of Labour Productivity, In: John McCombie, Maurizio Pugno and Bruno Soro (Eds.), *Productivity Growth and Economic Performance: Essays on Verdoorn's Law*, Basingstoke: Palgrave Macmillan, pp. 28-36.

Viner, J (1931), Cost Curves and Supply Curves, *Zeitschrift für Nationalökonomie*, 3 (1), pp. 23-46.

Williams, J H (1929), The Theory of International Trade Reconsidered, *The Economic Journal*, 39 (154), pp. 195-209.

WTO (1994), *Marrakesh Agreement Establishing the World Trade Organization*, https://www.wto.org/english/docs_e/legal_e/04-wto_e.htm accessed 21.03.2012.

WTO (2013), *The Case for Open Trade*, http://www.wto.org/english/thewto_e/whatis_e/tif_e/fact3_e.

The Emergence of Profit and Interest in the Monetary Circuit

Egmont Kakarot-Handtke¹

University of Stuttgart, Institute of Economics and Law, Germany

Abstract

Efficient progress of the monetary theory of production (MTP) is hampered by an unsatisfactory account of how profit and interest emerge in the monetary circuit. As matter of fact, this question puzzled already the classics. It seems evident that it cannot be answered by applying the usual tools. The present paper's purpose is to overcome the deadlock. This is done by setting the circulation approach on general structural axiomatic foundations.

JEL B41, B59, E19, E40

Key words: new framework of concepts, structure-centric, axiom set, monetary circuit, quantity of money, transaction money, profit, distributed profit, rate of interest, profit ratio equalization

"The existence of monetary profits at the macroeconomic (aggregate) level has always been a conundrum for theoreticians of the monetary circuit. If money is created from bank credit, how can we explain profits if firms borrow just enough to cover wages that are simply spent on consumption goods and returned to firms to extinguish their initial debt? Indeed, not only are firms unable to create profits, they also cannot raise sufficient funds to cover the payment of interest." (Rochon, 2005, p. 125), see also (Godley & Lavoie, 2007, p. 3), (Messori & Zazzaro, 2005, pp. 111-112), (Paraguez & Seccareccia, 2000, pp. 109-110), (Smithin, 1994, p. 176)

It is the purpose of the present paper to solve these conundrums. This is done by setting the circulation approach on a comprehensive axiomatic foundation. The general thesis says that human behavior does not yield to the axiomatic method (this rules out the standard approach), yet the axiomatization of the money economy's fundamental structure is feasible. The general case for structural axiomatization has been made elsewhere (2011a), (2011b), thus we can immediately take up circuit theory as specific application.

The formal ground is prepared in Section 1. The analytical point of departure, Schumpeter's 'reasonably small number of equations connecting a reasonably small number of variables', is given with the structural axiom set which represents the pure consumption economy. In Sections 0 and 0 the relations between the household and the business sector's respective stock of money, the quantity of money, and the average stock of transaction money are defined. In Sections 0 and 0 the connection between profit, distributed profit, retained profit and saving is established. This yields the general complementary relation between retained profit and saving–dissaving. In Section 0 the self-reproducing process of profit origination and distribution is constituted. In the final part, Sections 0 to 0 the transaction and banking unit of the central bank are introduced. This enables the determination of all prices and the loan interest rate under the conditions of, at first, zero profit, and then under positive overall profits. Section 0 concludes.

¹ Affiliation: University of Stuttgart, Institute of Economics and Law, Keplerstrasse 17, D-70174 Stuttgart. Correspondence address: AXEC-Project, Egmont Kakarot-Handtke, Hohenzollernstraße 11, D- 80801 München, Germany, e-mail: handtke@axec.de

1. Axioms and definitions

The first three structural axioms relate to income, production, and expenditures in a period of arbitrary length. For the remainder of this inquiry the period length is conveniently assumed to be the calendar year. Simplicity demands that we have at first one world economy, one firm, and one product.

Total income of the household sector Y in period t is the sum of wage income, i.e. the product of wage rate W and working hours L , and distributed profit, i.e. the product of dividend D and the number of share N .

$$Y = WL + DN \quad |t \quad (1)$$

Output of the business sector O is the product of productivity R and working hours.

$$O = RL \quad |t \quad (2)$$

Consumption expenditures C of the household sector is the product of price P and quantity bought X .

$$C = PX \quad |t \quad (3)$$

The axioms represent the pure consumption economy, that is, no investment expenditures, no foreign trade, and no taxes or any other state activity.

Definitions are supplemented by connecting variables on the right-hand side of the identity sign that have already been introduced by the axioms (Boylan & O'Gorman, 2007, p. 431). With (4) wage income Y_w and distributed profit income Y_D is defined:

$$Y_w \equiv WL \quad Y_D \equiv DN \quad |t. \quad (4)$$

Definitions add no new content to the set of axioms but determine the logical context of concepts. New variables are introduced with new axioms.

The economic meaning is rather obvious for the set of structural axioms. What deserves mention is that total income in (1) is the sum of wage income and *distributed profit* and not of wage income and profit. Profit and distributed profit are quite different things that have to be thoroughly kept apart.

"A theory consists of a number of assumptions which logically function as axioms. Through specification and by introducing initial conditions, we may deduce predictions from them. If the predictions prove to be valid we may also say that the assumptions are realistic."
(Klant, 1994, p. 75)

2. Money and credit

The dichotomization of the real and the monetary sphere was a central point of Keynes's methodological critique of conventional economics:

"The division of economics between the theory of value and distribution on the one hand and the theory of money on the other hand is, I think, a false division." (Keynes, 1973, p. 293)

The first task, then, is to demonstrate how money follows consistently from the axiom set. If income is higher than consumption expenditures the household sector's stock of money increases. The change in period t is defined as:

$$\Delta \bar{M}_H \equiv^m Y - C \quad | t. \quad (5)$$

The identity sign's superscript m indicates that the definition refers to the monetary sphere.

The stock of money \bar{M}_H at the end \bar{t} of an arbitrary number of periods is defined as the numerical integral of the previous changes of the stock plus the initial endowment:

$$\bar{M}_H \equiv \sum_{t=1}^{\bar{t}} \Delta \bar{M}_{Ht} + \bar{M}_{H0} \quad | \bar{t}. \quad (6)$$

The changes in the stock of money as seen from the business sector are symmetrical to those of the household sector:

$$\Delta \bar{M}_B \equiv^m C - Y \quad | t. \quad (7)$$

The business sector's stock of money at the end of an arbitrary number of periods is accordingly given by:

$$\bar{M}_B \equiv \sum_{t=1}^{\bar{t}} \Delta \bar{M}_{Bt} + \bar{M}_{B0} \quad | \bar{t}. \quad (8)$$

In order to reduce the monetary phenomena to the essentials it is supposed that all financial transactions are carried out by the central bank. The stock of money then takes the form of current deposits or current overdrafts (cf. (Wicksell, 1936, p. 70), (Renversez, 1996), (Lavoie, 2003, pp. 506-509)). Initial endowments can be set to zero. Then, if the household sector owns current deposits according to (6) the current overdrafts of the business sector are of equal amount according to (8), and vice versa. Each sector's *stock* of money is either positive or negative. Money and credit are at first symmetrical. From the central bank's perspective the quantity of money at the end of an arbitrary number of periods is then given by the absolute value either from (6) or (8):

$$\bar{M} \equiv \left| \sum_{t=1}^{\bar{t}} \Delta \bar{M}_{Ht;Bt} \right| \quad \text{with} \quad \bar{M}_{H0;B0} = 0 \quad | \bar{t}. \quad (9)$$

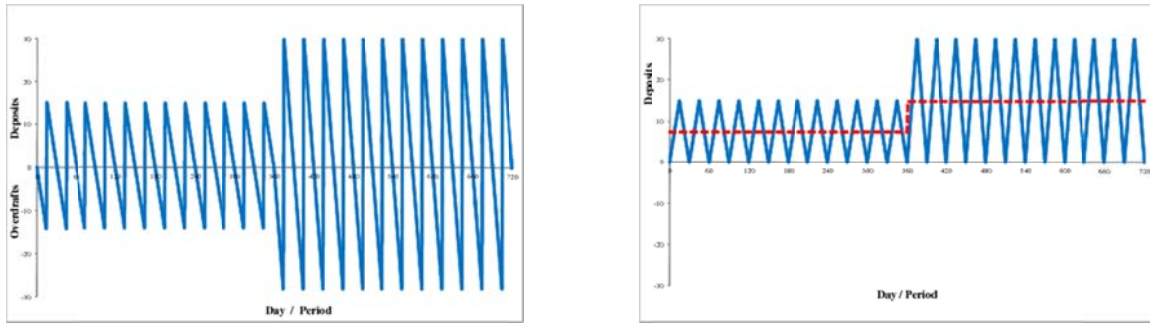
The *quantity* of money is always ≥ 0 and follows directly from the axioms. It is assumed at first that the central bank plays an *accommodative* role and simply supports the autonomous market transactions between the household and the business sector. For the time being, money is the *dependent* variable.

Transaction money

"In different ways, advocates of MTP [monetary theory of production] reject the simultaneous logic of general equilibrium analysis. They consider ... the need for analyzing the successive phases of the economic process." (Fontana & Realfonzo, 2005, p. 9)

By sequencing the initially given period length of one year into months the idealized transaction pattern that is displayed in Figure 1 results (cf. (Newlyn, 1971), (Schmitt, 1996, p. 134)). It is assumed that the monthly income $\frac{Y}{12}$ is paid out at mid-month. In the first half of the month the daily spending of $\frac{Y}{360}$ increases the current overdrafts of the households. At mid-month the households change to the positive side and have current deposits of $\frac{Y}{24}$ at their disposal. This amount reduces continuously towards the end of the month. This pattern is exactly repeated over the rest of the year. At the end of each subperiod, and therefore also at the end of the year, both the stock of money and the quantity of money is zero. Money is present and absent depending on the time frame of observation.

Figure 1: Household sector's transaction pattern for different nominal incomes in two periods



(a) Transaction pattern over two periods

(b) Average stock of transaction money \hat{M}_T

In period₂ the wage rate, the dividend and the price is doubled. Since no cash balances are carried forward from one period to the next, there results *no* real balance effect provided the doubling takes place exactly at the beginning of period₂.

From the perspective of the central bank it is a matter of indifference whether the household or the business sector owns current deposits. Therefore, the pattern of Figure 1a translates into the average amount of current deposits in Figure 1b. This average stock of transaction money depends on income according to the transaction equation

$$\hat{M}_T \equiv \kappa Y \quad |t \tag{10}$$

which resembles Pigou's Cambridge equation; the underlying theory, though, is thereby not adopted. For the regular transaction pattern that is here assumed as a idealization the index is $\frac{1}{48}$. Different transaction patterns are characterized by different numerical values of the transaction pattern index.

For formal convenience the expenditure ratio ρ_E and the sales ratio ρ_X is defined as:

$$\rho_E \equiv \frac{C}{Y} \quad \rho_X \equiv \frac{X}{O} \quad |t. \tag{11}$$

An expenditure ratio $\rho_E = 1$ indicates that consumption expenditures are equal to income, or, in other words, that the household sector's budget is balanced. A value of $\rho_X = 1$ of the sales ratio means that the quantities produced and sold are equal in period t or, in other words, that the product market is cleared. Taking (10) and (11) together one gets the explicit transaction equation for the limiting case of market clearing *and* budget balancing:

$$(i) \hat{M}_T \equiv \kappa \frac{\rho_X}{\rho_E} RLP \quad (ii) \frac{\hat{M}_T}{P} = \kappa O \quad \text{if } \rho_X = 1, \rho_E = 1 \quad |t. \tag{12}$$

We are now in the position to substantiate the notion of accommodation as a money-growth formula. According to (i) the central bank enables the average stock of transaction money to expand or contract with the development of productivity, employment, and price. In other words, the real average stock of transaction money, which is a statistical artifact and no physical stock, is proportional to output (ii) if the transaction index is given and if the ratios ρ_E and ρ_X are unity. Under these *initial* conditions money is endogenous (Desai, 1989, p. 150) and neutral (Patinkin, 1989) in the structural axiomatic context. Money emerges from

autonomous market transactions and has three aspects: stock of money (\bar{M}_H, \bar{M}_B), quantity of money (here $\bar{M} = 0$ at period start and end because of $\rho_E = 1$, cf. (Graziani, 1996, p. 143)) and average stock of transaction money (here $\hat{M}_T > 0$).

4. Profit

The business sector's financial profit in period t is defined with (13) as the difference between the sales revenues – for the economy as a whole identical with consumption expenditures C – and costs – here identical with wage income Y_W .²

$$\Delta \bar{Q}_{fi} \equiv C - Y_W \quad |t. \quad (13)$$

In explicit form, after the substitution of (3) and (4), this definition is identical with that of the theory of the firm:

$$\Delta \bar{Q}_{fi} \equiv PX - WL \quad |t. \quad (14)$$

Using the first axiom (1) and the definitions (4) one gets:

$$\Delta \bar{Q}_{fi} \equiv C - Y + Y_D \quad |t. \quad (15)$$

The three definitions are formally equivalent. If distributed profit Y_D in (15) is set to zero, then profit or loss of the business sector is determined solely by expenditures and income. For the business sector as a whole to make a profit consumption expenditures C have in the simplest case to be greater than wage income Y_W . So that profit comes into existence in the pure consumption economy the household sector must run a deficit at least in one period. This in turn makes the inclusion of the financial sector mandatory. A theory that does not include at least one bank that supports the concomitant credit expansion (6) cannot capture the essential features of the market economy (Keynes, 1973, p. 85). Mention should be made that, for quite different reasons, neither neoclassicals nor Keynesians ever came to grips with profit (Desai, 2008, p. 10), (Tómasson & Bezemer, 2010, pp. 1-4). There is no difference on this point with circuitists, or, for that matter, with heterodox attempts (e.g. (Correa, 2012), (Keen, 2010), (Bruun & Heyn-Johnsen, 2009), (Binswanger, 1996)).

5. Retained profit and saving

Profits can either be distributed or retained. If nothing is distributed, then profit adds entirely to the financial wealth of the firm. Retained profit $\Delta \bar{Q}_{re}$ is defined for the business sector as a whole as the difference between profit and distributed profit in period t :

$$\Delta \bar{Q}_{re} \equiv \Delta \bar{Q}_{fi} - Y_D \quad |t. \quad (16)$$

² Profits from changes in the value of nonfinancial assets are neglected here, i.e. the condition of market clearing $O=X$ holds throughout. For details about changes of inventory see (2011c, p. 5). Changes in the value of other nonfinancial assets are treated at length in (2012b).

Using (15) and (7) it follows:

$$\Delta \bar{Q}_{re} \equiv C - Y \equiv^m \Delta \bar{M}_B \quad | t. \quad (17)$$

Retained profit $\Delta \bar{Q}_{re}$ is the residual $C - Y$ as it appears at the firm that represents the business sector.

The same residual appears at the central bank as a change of the business sector's stock of money $\Delta \bar{M}_B$.

The *two aspects* are kept apart by the notation. It follows immediately that the development of the business sector's stock of money, which may carry a positive or negative sign, is given by (8).

Financial saving is given by (18) as the difference of income and consumption expenditures. This definition is identical with Keynes's (1973, p. 63), only the notation is different.

$$\Delta \bar{S}_{fi} \equiv Y - C \quad | t \quad (18)$$

In combination with (5) this yields the straightforward relation:

$$\Delta \bar{S}_{fi} \equiv Y - C \equiv^m \Delta \bar{M}_H \quad | t. \quad (19)$$

Financial saving $\Delta \bar{S}_{fi}$ is the residual $Y - C$ as it appears at the household sector; the same residual appears at the central bank as a change of the household sector's stock of money $\Delta \bar{M}_H$.

The *two aspects* are kept apart by the notation. It follows immediately that the development of the household sector's stock of money, which may carry a positive or negative sign, is given by (6). Equations (19) respectively (17) determine the changes of the quantity of money as given by (9).

Financial saving (19) and retained profit (17) always move in opposite directions, i.e. $\Delta \bar{Q}_{re} \equiv -\Delta \bar{S}_{fi}$.

Let us call this the complementarity corollary because it follows directly from the definitions themselves. The corollary asserts that the complementary notion to saving is *not* investment but negative retained profit. Positive retained profit is the complementary of dissaving. This entails that the plans of households and firms are only mutually compatible if both retained profit and financial saving are zero. This rarely happens in the real world. Therefore, a behavioral equilibrium in the sense of Arrow and Hahn (1991, p. 16), although formally possible, plays no role in the structural axiomatic context.

In the general case, profit or loss depends on consumer spending and profit distribution. If distributed profit is set to zero, then we face, according to (15), three logical alternatives: $C < Y_w$, $C = Y_w$ or $C > Y_w$.

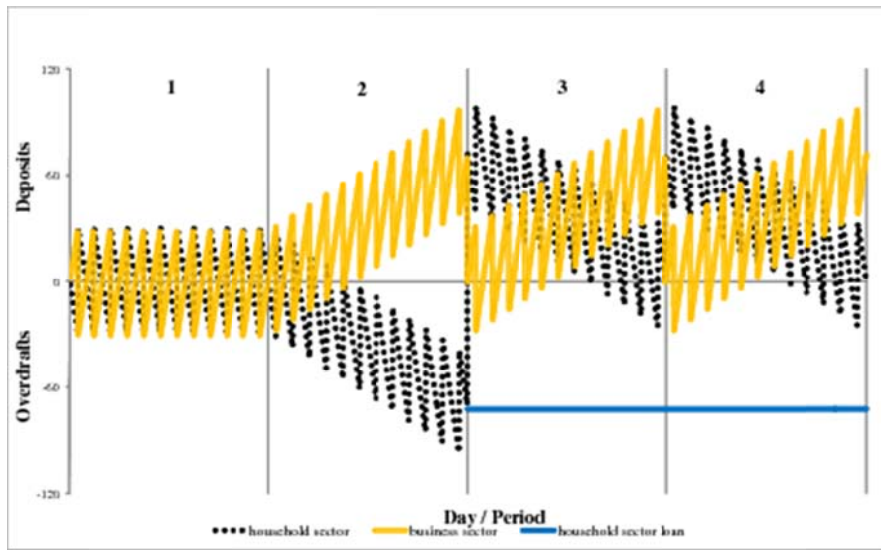
The first alternative entails a loss for the business sector as a whole, the second entails zero profit, and only the third leads to profit which in turn is the indispensable condition for a reproducible economy. Hence the real question is not about the existence of a zero-profit equilibrium, but how the market economy can, and in fact does, avoid this predicament over a longer time span. What is needed for a start is the deficit spending of the household sector at least in one period. When the purchase of long lived consumption goods, e.g. houses, is correctly subsumed under consumption expenditures there arises no problem with regard to collateral for the banking industry and a sound credit expansion may – in principle – proceed for an indefinite time in the pure consumption economy. It needs hardly emphasis that the process of profit origination looks different in the investment economy (for details see (2011d)). The underlying mechanism, though, is essentially the same.

In the pure consumption economy one has labor input as the sole factor of production and wage income as the corresponding factor remuneration. Since the factor capital is nonexistent, profit cannot be assigned to it in functional terms. From this follows as far-reaching methodological consequence: to treat profit as factor income is a category mistake (for a proof see (2012a)).

6. Profit and profit distribution

If, with distributed profit at first set to zero, consumption expenditures get ahead of wage income, i.e. $\rho_E > 1$, the household and business sector's transaction patterns diverge in period₂ as shown in Figure 2. The household sector's current overdrafts increase until period end and, as a perfect mirror image, the business sector's current deposits increase, too.

Figure 2: Dissaving leads in period₂ to an increase of the household sector's current overdrafts and the business sector's current deposits; at the beginning of period₃ the household sector takes up a loan and profits are fully distributed



It is assumed that the household sector consolidates the overdrafts and takes up a one-period loan at the banking unit of the central bank exactly at the beginning of period₃. This reduces overdrafts to zero. The household sector switches from short term liabilities, in fact the shortest possible term, to longer term liabilities.

The business sector posts a profit at the end of period₂ according to (15). It is assumed that this profit is fully distributed at the beginning of period₃. This reduces the business sector's current deposits to zero and at the same time increases the household sector's deposits by the same amount. It therefore holds that distributed profit in period₃ is exactly equal to profit in period₂:

$$Y_{D3} = \Delta \bar{Q}_{fi2}. \tag{20}$$

In period₃ the households no longer dissave but spend their distributed profits. Total consumption expenditures are equal to total income, i.e. $\rho_E = 1$, as they were in period₁. From this follows the profit in period₃ as:

$$\Delta \bar{Q}_{fi3} \equiv \underbrace{C_3 - Y_3}_0 + Y_{D3} \Rightarrow \Delta \bar{Q}_{fi3} = \Delta \bar{Q}_{fi2}. \tag{21}$$

Profit in period₃ is exactly equal to the profit of the previous period. From (16) in turn follows that retained profit is zero. This pattern is repeated in period₄ and it is evident that this configuration is reproducible for an indefinite time span provided that profits are fully distributed and fully spent and the one-period loans are

prolonged in each successive period. The transaction pattern index κ in (12), assumes different numerical values in period₂ and period₃. Subsequently it remains constant. This entails an increase of the average stock of transaction money beginning with period₂. The quantity of money (9) is, after profit distribution, zero at the beginning of period₃ and then at the beginning of the following periods.

7. The transaction unit

The business sector consists of a consumption goods producing firm A and the central bank as the second firm B . To begin with, the central bank handles only the money transactions. Total employment is given by:

$$L \equiv L_A + L_B \quad |t. \quad (22)$$

To focus exclusively on the monetary phenomena variations of total employment are excluded. Total income consists according to (1) of wage income and distributed profit. To simplify the analysis the wage rates for all firms are set equal. Distributed profits are at first zero:

$$Y = \underbrace{W_A}_{\bar{w}} L_A + \underbrace{W_B}_{\bar{w}} L_B + \underbrace{(D_A N_A + D_B N_B)}_{Y_D=0} \quad |t. \quad (23)$$

The household sector apportions its consumption expenditures between the purchase of consumption goods and the purchase of transaction services. With X_B the number of transactions per period that are carried out by the central bank on behalf of the households is denoted:

$$C = P_A X_A + P_B X_B \quad |t. \quad (24)$$

Consumption expenditures are equal to income, i.e. $\rho_E = 1$. The households neither save nor dissave. Overall financial profit (14) is differentiated for the two firms:

$$\begin{aligned} \Delta \bar{Q}_{fiA} &\equiv P_A X_A - W L_A \\ \Delta \bar{Q}_{fiB} &\equiv P_B X_B - W L_B \end{aligned} \quad |t. \quad (25)$$

Under the condition of market clearing, i.e. $\rho_X = 1$, this can be rewritten as:

$$\begin{aligned} \Delta \bar{Q}_{fiA} &= P_A R_A L_A \left(1 - \frac{W}{P_A R_A} \right) \quad \text{if } \rho_{XA} = 1 \\ \Delta \bar{Q}_{fiB} &= P_B R_B L_B \left(1 - \frac{W}{P_B R_B} \right) \quad \text{if } \rho_{XB} = 1 \end{aligned} \quad |t. \quad (26)$$

Overall profits are zero because of $C = Y$ and $Y_D = 0$. The zero profit condition for a single firm reads $\frac{W}{PR} = 1$. Under this conditions follows from (26) that absolute prices are equal to unit wage costs, i.e. $P_A = \frac{W}{R_A}$ respectively $P_B = \frac{W}{R_B}$. In sum: both markets are cleared, the household sector's budget is balanced and profits are zero for both the consumption goods producing firm and the transaction unit of the central bank. Money transactions consume resources, the less so the higher the productivity of the transaction unit is. Alone for this reason money cannot be neutral. The price the households pay for each transaction P_B follows from (26) and the zero profit condition.

8. The banking unit

The inclusion of the banking unit entails that the given resources of the business sector L have first to be reallocated:

$$L \equiv L_A + L_B + L_C \quad |t. \quad (27)$$

As a consequence total income is then given by:

$$Y = \underbrace{W_A}_{W} L_A + \underbrace{W_B}_{W} L_B + \underbrace{W_C}_{W} L_C + \underbrace{(D_A N_A + D_B N_B + D_C N_C)}_{Y_D=0} \quad |t. \quad (28)$$

The interest payments of the household sector to the banking unit have to be subsumed under consumption expenditures:

$$\begin{aligned} C &= P_A X_A + P_B X_B + I_C \bar{A}_C \quad |t. \\ C &= C_A + C_B + C_C \end{aligned} \quad (29)$$

The quantity bought from the banking unit X_C can here be replaced by the amount of the loan \bar{A}_C (for the consistent derivation of the rate of interest from the differentiated axiom set see (2011b, pp. 12-14)).

The reallocation of labor input is neutral with regard to the price of the consumption good. When labor input L_C is taken away from firm A output falls. At the same time consumption expenditures are redirected away from purchases of consumption goods to purchases of the services of the banking unit, i.e. C_A goes down and C_C goes up. This leaves the price of the consumption good unaffected under the given conditions. The household sector buys less consumption goods and more banking services. According to this demand shift the unaltered total labor input is reallocated.

Profit for each firm is zero, i.e. $\frac{W}{PR} = 1$:

$$\begin{aligned} \Delta \bar{Q}_{fA} &= P_A R_A L_A \left(1 - \frac{W}{P_A R_A} \right) \quad \text{if } \rho_{XA} = 1 \\ \Delta \bar{Q}_{fB} &= P_B R_B L_B \left(1 - \frac{W}{P_B R_B} \right) \quad \text{if } \rho_{XB} = 1 \end{aligned} \quad |t. \quad (30)$$

$$\Delta \bar{Q}_{fC} = I_C \bar{A}_C \left(1 - \frac{W}{I_C \frac{\bar{A}_C}{L_C}} \right) \quad \text{if } \rho_{XC} = 1$$

The zero profit conditions and the market clearing condition define the commodity price, the transaction price and the rate of interest. All are equal to the respective unit wage costs. The inclusion of the banking unit and the appearance of interest on the one-period loan results in a reallocation of demand and resources. The loan interest rate is, at first, alone determined by the production conditions of the banking unit. The same holds for the price of the consumption good P_A and the price of a monetary transaction P_B . All firms recoup their costs. Interest payments of the households on the one-period-loan are equal to wage income in the banking unit. All relative prices are objectively determined by the respective productivities. The case for business loans is analogous (for details see (2011c, pp. 2-7)).

9. Equal profit ratios

In order to eliminate all subjective elements and to determine all prices analytically for the general case of positive overall profit an additional assumption is required. The most suitable condition is profit ratio equalization. The overall profit ratio ρ_Q follows from (15) as:

$$\rho_Q \equiv \frac{\Delta \bar{Q}_{fi}}{WL} \Rightarrow \rho_Q \equiv \rho_E (1 + \rho_D) - 1 \quad |t. \quad (31)$$

The profit ratio ρ_Q for the business sector as a whole is positive if the expenditure ratio ρ_E is > 1 or the distributed profit ratio ρ_D is > 0 , or both. The distributed profit ratio is defined as:

$$\rho_D \equiv \frac{Y_D}{Y_W} \quad |t. \quad (32)$$

The profit ratio for each firm is then given by:

$$\rho_{QA} \equiv \frac{P_A X_A}{W_A L_A} - 1 \quad \rho_{QB} \equiv \frac{P_B X_B}{W_B L_B} - 1 \quad \rho_{QC} \equiv \frac{I_C \bar{A}_C}{W_C L_C} - 1 \quad |t. \quad (33)$$

Under the condition of equal profit ratios $\rho_{QA} = \rho_{QB} = \rho_{QC} = \rho_Q$ follows for the market clearing prices and the rate of interest:

$$\begin{aligned} P_A &= \frac{W}{R_A} \rho_E (1 + \rho_D) \quad \text{if} \quad \rho_{XA} = 1 \\ P_B &= \frac{W}{R_B} \rho_E (1 + \rho_D) \quad \text{if} \quad \rho_{XB} = 1 \quad |t. \\ I_C &= \frac{W}{\frac{\bar{A}_C}{L_C}} \rho_E (1 + \rho_D) \quad \text{if} \quad \rho_{XC} = 1 \end{aligned} \quad (34)$$

If the overall expenditure ratio ρ_E is unity and the distributed profit ratio ρ_D is zero then prices and the interest rate are equal to unit wage cost in each firm as in (30). In the general case, prices and the interest rate depend also on the expenditure ratio and the distributed profit ratio. An expenditure ratio ρ_E of unity and a distributed profit ratio $\rho_D > 0$ yields the reproducible configuration of Figure 2 which entails both interest and profit. Relative prices are the same as in the zero profit case.

Equations (34) looks like markup pricing formulas. They are nothing of the sort. The prices are determined by the conditions of market-clearing and equal profit ratios. The introduction if the markup assumption would over-determinate the system. Yet the equations go some way in explaining why most economic models of pricing 'derive a reasonably stable markup of price over cost' (Hall, 2011, p. 446).

It is evident that profit ratio equalization is a formal benchmark. Whether profit ratios in fact equalize in the real world is a quite different matter. This, though, is of secondary importance. The indispensable

condition for a viable money economy is that overall profits are greater than zero. In the pure consumption economy this means that the household sector must produce an initial deficit. In a more complex economy the investing business sector (2011d), the foreign trade sector (2011e) or the government sector (2012c) may spark off and maintain an expansionary trend. The basic principle is the same. What the monetary economy needs least is an equilibrium with balanced budgets.

10. Conclusions

Behavioral assumptions, rational or otherwise, are not solid enough to be eligible as first principles of theoretical economics. Hence all endeavors to lay the formal foundation on a new site and at a deeper level actually need no further vindication. The present paper suggests three non-behavioral axioms as groundwork for the circuitist approach. The main results of this paradigmatic application are:

- The quantity of money follows directly from the set of structural axioms.
- Under the initial conditions of market clearing and budget balancing money is endogenous and neutral.
- A positive expenditure-income asymmetry is the ultimate structural originator of profit and therefore the indispensable prerequisite for favorable business conditions. This holds for the elementary consumption economy and the complex investment economy in equal measure.
- In the pure consumption economy total profit of the business sector is greater than zero if the expenditure ratio is > 1 or the distributed profit ratio is > 0 , or both.
- In the pure consumption economy one has labor input as the sole factor of production and wage income as the corresponding factor remuneration. Since the factor capital is nonexistent, profit cannot be assigned to it in functional terms. From this follows as far-reaching methodological consequence: to treat profit as factor income is a category mistake.
- Under the condition of full profit distribution profit remains constant and retained profit is zero. This configuration is reproducible for an indefinite time span.
- Loans are produced like any other commodity. The rate of interest inherits the role of the price.
- Under the condition of profit ratio equalization all prices and the rate of interest on one-period loans to the household sector are objectively determined. Relative prices depend solely on the productivities in the different lines of production.

References

- Arrow, K. J., & Hahn, F. H. (1991). *General Competitive Analysis*. Amsterdam, New York: North-Holland.
- Binswanger, M. (1996). Money Creation, Profits, and Growth: Monetary Aspects of Economic Evolution. In E. Helmstädter, & M. Perlman (Eds.), *Behavioral Norms, Technological Progress, and Economic Dynamics* (pp. 413-437). Ann Arbor: University of Michigan Press.
- Boylan, T. A., & O'Gorman, P. F. (2007). Axiomatization and Formalism in Economics. *Journal of Economic Surveys*, 21(2), pp. 426-446.
- Bruun, C., & Heyn-Johnsen, C. (2009). The Paradox of Monetary Profits: An Obstacle to Understanding Financial and Economic Crisis? *Economics E-Journal, Discussion Papers*, 2009-52.
- Correa, R. (2012). The Determination of Profits. *World Economic Journal, Article Review Online*, Accessed February 6th 2012.
- Desai, M. (1989). Endogenous and Exogenous Money. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Money* (pp. 146-150). London, Basingstoke: Macmillan.
- Desai, M. (2008). Profit and Profit Theory. In S. N. Durlauf, & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics Online* (2nd ed., pp. 1-11). Basingstoke: Palgrave Macmillan.

- Fontana, G., & Realfonzo, R. (2005). Introduction: The Monetary Theory of Production. Tradition and Perspectives. In G. Fontana, & R. Realfonzo (Eds.), *The Monetary Theory of Production* (pp. 1-19). Houndmills, Basingstoke, New York: Palgrave Macmillan.
- Godley, W., & Lavoie, M. (2007). *Monetary Economics. An Integrated Approach to Credit, Money, Income and Wealth*. Houndmills, Basingstoke, New York: Palgrave Macmillan.
- Graziani, A. (1996). Money as Purchasing Power and Money as a Stock of Wealth in Keynesian Economic Thought. In G. Deleplace, & E. J. Nell (Eds.), *Money in Motion* (pp. 139-154). Houndmills, Basingstoke, New York: Macmillan.
- Hall, R. E. (2011). The Long Slump. *American Economic Review*, 101(2), 431-469.
- Kakarot-Handtke, E. (2011a). Keynes's Missing Axioms. *SSRN Working Paper Series*, 1841408, 1-32.
- Kakarot-Handtke, E. (2011b). The Pure Logic of Value, Profit, Interest. *SSRN Working Paper Series*, 1838203, pp. 1-25.
- Kakarot-Handtke, E. (2011c). Reconstructing the Quantity Theory (II). *SSRN Working Paper Series*, 1903663, pp. 1-19.
- Kakarot-Handtke, E. (2011d). Squaring the Investment Cycle. *SSRN Working Paper Series*, 1911796, pp. 1-25.
- Kakarot-Handtke, E. (2011e). Trade, Productivity, Income, and Profit: The Comparative Advantage of Structural Axiomatic Analysis. *SSRN Working Paper Series*, 1970895, pp. 1-23.
- Kakarot-Handtke, E. (2012a). The Common Error of Common Sense: An Essential Rectification of the Accounting Approach. *Levy Economics Institute Working Papers*, 731, pp. 1-28.
- Kakarot-Handtke, E. (2012b). Primary and Secondary Markets. *Levy Economics Institute Working Papers*, 741, pp. 1-27.
- Kakarot-Handtke, E. (2012c). Taxes, Profits, and Employment: A Structural Axiomatic Analysis. *SSRN Working Paper Series*, 1995025, pp. 1-23.
- Keen, S. (2010). Solving the Paradox of Monetary Profits. *Economics E-Journal*, 4(2010-31).
- Keynes, J. M. (1973). *The General Theory of Employment Interest and Money*. London, Basingstoke: Macmillan.
- Klant, J. J. (1994). *The Nature of Economic Thought*. Aldershot, Brookfield: Edward Elgar.
- Lavoie, M. (2003). A Primer on Endogenous Credit-Money. In L.-P. Rochon, & S. Rossi (Eds.), *Modern Theories of Money* (pp. 506-543). Cheltenham, Northampton: Edward Elgar.
- Messori, M., & Zazzaro, A. (2005). Single-Period Analysis: Financial Markets, Firms' Failures and Closure of the Monetary Circuit. In G. Fontana, & R. Realfonzo (Eds.), *The Monetary Theory of Production. Tradition and Perspectives*. (pp. 111-123). Houndmills, Basingstoke, New York: Palgrave Macmillan.
- Newlyn, W. T. (1971). *Theory of Money* (2nd ed.). Oxford: Clarendon Press.
- Paraguez, A., & Seccareccia, M. (2000). The Credit Theory of Money: The Monetary Circuit Approach. In J. Smithin (Ed.), *What is Money?* (pp. 101-123). London, New York: Routledge.
- Patinkin, D. (1989). Neutrality of Money. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Money* (pp. 273-287). London, Basingstoke: Macmillan.
- Reversez, F. (1996). Monetary Circulation and Overdraft Economy. In G. Deleplace, & E. J. Nell (Eds.), *Money in Motion* (pp. 465-488). Houndmills, Basingstoke, London: Macmillan.

Rochon, L.-P. (2005). The Existence of Monetary Profits within the Monetary Circuit. In G. Fontana, & R. Realonzo (Eds.), *The Monetary Theory of Production. Tradition and Perspectives*. (pp. 125-138). Houndmills, Basingstoke, New York: Palgrave Macmillan.

Schmitt, B. (1996). A New Paradigm for the Determination of Money Prices. In G. Deleplace, & E. J. Nell (Eds.), *Money in Motion* (pp. 104-138). Houndmills, Basingstoke, London: Macmillan.

Smithin, J. (1994). *Controversies in Monetary Economics*. Aldershot, Brookfield: Edward Elgar.

Tómasson, G., & Bezemer, D. J. (2010). What is the Source of Profit and Interest? A Classical Conundrum Reconsidered. *MPRA Paper, 20557*, pp. 1-34.

Wicksell, K. (1936). *Interest and Prices*. London: Macmillan.

Book Review

Why Nations fail - the origins of power, prosperity and poverty

by Daron Acemoglu and James A. Robinson, Crown Business, New York, 2012, 529 pages, ISBN 978-0-307-71921-8

Deniz Kellecioglu¹

International economist, Istanbul Bilgi University, Turkey

Why nations fail is a compelling contribution to the classic question of why some countries are poor and some rich, but it is also marred from several shortcomings and under-representations, which cast doubts on the positive messages of creating better institutions and reducing poverty.

The main thesis of the authors is that 'nations fail because their extractive economic institutions do not create the incentives needed for people to save, invest, and innovate'. These institutions are run by elite groups who are exploiting the resources of the country for their own use, leaving little to the population at large. Furthermore, extractive *political* institutions support the economic institutions by cementing the elite power base. The authors argue that 'extractive economic and political institutions, though their details vary under different circumstances, are *always at the root* of the failure' (emphasis added). Moreover, 'another reason why nations fail is that their *states* fail. This, in turn, is a consequence of decades of rule under extractive institutions' (emphasis added). According to the authors, inclusive institutions, on the other hand, provide equal opportunity to all citizens of the country allowing for a broad and sustained prosperity. Furthermore, inclusive institutions are associated with *virtuous cycles of development*, while *vicious cycles of development* are typical of extractive institutions.

Why nations fail is a well-written book based on an intriguing and persuasive narrative over the past half millennium. In my view, the most valuable contribution of the book involves its anti-elitist, anti-racist, and anti-colonial perspectives, as well as the empowerment stance as the way out of extractive institutions, so to '...force the elite to create more pluralistic institutions'. In relation, it is valuable that elite co-optation and intra-elite clashes are pinned down while peoples' struggles against authoritarianism are acknowledged. The appraisal in the chapter 'theories that don't work' is also valuable, albeit rather limited as we shall see below.

A major concern involves the concept of 'inclusive markets', under the larger umbrella of inclusive institutions. Although the book does not contain defining specifications, it is quite easy to trace out important features of their understanding of inclusive markets: 'rule of law', 'property rights', 'patent systems', 'macroeconomic stability', 'creative destruction' and incentives that encourage innovation. It is quite well-known that such a policy bundle falls under the so called new institutional economics. This school-of-thought shares the bundle with neoclassical economics, which is almost entirely equal to mainstream economics. Both these schools-of-thought have been heavily criticized (cf. Ha-Joon Chang 2011 for a concise account), especially in the aftermath of the global financial crises of 2007-2008. The book's 'inclusive markets' incorporate the notion of equal opportunity, but pay no attention to unequal resources. Unequal points of departure, even under inclusive markets and institutions, are likely to benefit the already resourceful (economically, politically, socially, culturally) groups of the society. These groups will find it easier to take-off under improved institutional settings, while resource-poor groups will continue to face severe difficulties in improving their livelihoods. It is highly likely that this asymmetric scenario will obstruct any meaningful development for the society as a whole, as the resource imbalances will be exacerbated by virtuous and vicious cycles of development. Even a 'freedom for all' institutional setting would almost certainly imply unequal points of departures.

¹ A vote of thanks to Professor Koray Akay, as well as Mkhululi Ncube, Malcolm Spence, and John Sloan for valuable comments on the draft version of this paper. The author declares no conflict of interest. Author correspondence: deniz.kellecioglu@gmail.com

This scenario involves a society, with some members or entities that start with an upper hand, which is likely to produce the very things the authors are against: elitism and authoritarianism. South Africa's post-Apartheid performance is one example of such an outcome. Today, inequality in South Africa is actually worse than under the years preceding the fall of Apartheid, including the inequality between ethnic groups (Leibbrandt, et al. 2010). The implementation of democracy, the removal of sanctions against the apartheid elites and the overall 'inclusive markets' trajectory of the country have done little for the unprivileged half of the population since 1994. It should not be controversial to understand that the already resourceful group, plus a new political class, came to benefit from the improved institutional settings. The book would have benefitted greatly from a discussion on such outcomes that contrasts expected aspirations.

There are also unvalidated omissions of well-established explanatory factors of comparative development. Although it goes without saying that no single work can explain everything, it is the responsibility of authors to address relevant issues accordingly. Such a task may involve two main routes. In the first, the authors clarify and validate their interest in focusing on certain factors. In the second route, authors may dismiss unselected factors through an appropriate line of criticism. Acemoglu and Robinson actually exercise both these routes, but in a restricted manner. As mentioned, they provide an appraisal of geography, culture, the ignorance hypothesis, and later in the book, variants of the modernization theory. In addition, they acknowledge the central roles of colonialism, slavery, racism, critical junctures (such as The Black Death), virtuous and vicious development cycles, and contingency. Their inclusion of the unequal dissemination of industrialism, luck and context-related needs as a driving force are quite unsatisfactory, however. What is worse, they do not account for the clear and present international meddling of the post-colonial era, the cold war in particular, nor colour-coded racism over the time period at hand. The latter factor involves how poverty and wealth of nations today are also related to the unfortunate, but real, process of globalised colour-coded racism. This process originated from the onset of western colonisers and slave traders having lighter skin colours, and the colonised and enslaved people having darker skin colours (Kellecioglu 2010).

As a matter of fact, the book is seriously underrepresenting the role of powerful nations and international entities, such as the International Monetary Fund (IMF), the World Trade Organization (WTO), the World Bank, and transnational corporations (TNCs). The type of institutional developments assessed in the book (at the national and macro level) are certainly influenced by international institutional configurations, which is a topic almost never touched upon in the book. One example are TNCs in the natural resource sector throughout Africa. These TNCs, which are often larger than the countries they operate within, exploit their upper hand in international judiciary, information access, and economic leverage - regardless of the institutional development of the country and the business sector. In a similar fashion, although Zimbabwe is much talked-upon in the book, the destructive roles of the British government and the World Bank over the 1990s (cf. Bond and Manyanya 2003) should have complemented the valid criticism of Robert Mugabe and his cronies. In relation, there are nations, categorised as having inclusive institutions that systematically carry out extractive interventions in other nations. To be fair, one such example is mentioned in the book: the intervention of Iraq in the year 2003. Unfortunately, this point is not taken further to be included in the final analysis. Such international disruptions are likely to obstruct or reverse institutional developments. Rent-seeking international stakeholders exploit existing power imbalances that are not dependent on institutional settings alone.

This takes us to yet another under-representation of the book: the precedents of institutions. By and large, the authors argue that it is difficult to trace emanating factors to inclusive institutions, while holding that England's Glorious Revolution of 1688 was probably an essential critical juncture for the onset of modern-day inclusive institutions. Although the two statements are agreeable, it is plausible that the Western institutional development would have been delayed, if emerged at all, had it not been for preceding extractive institutions, which involved outright piracy and cruel colonisation throughout the world from late 15th century onwards (cf. Landes 1998 and Zinn 2005 [1980]). In such a case, the analytical vista must be altered fundamentally, since the causality run stronger from economic development to institutions, rather than the other way around. In a similar fashion, necessity could be an essential factor to institutional drift. This is

indirectly touched upon in the book, as the success of the Glorious Revolution is attributed to 'the rise of the Atlantic trade that enriched and emboldened merchants opposing the Crown.' All in all, although the objective of the authors is to frame comparative developments with an institutional lens, exclusion of possible originators to institutions-building adds to the under-representation of the narrative.

The point about possible extractive precedents to inclusive institutions is also valid at the national level. The authors classify, for instance, the United States as having inclusive institutions: '[u]ltimately the good economic institutions of the United States resulted from the political institutions that gradually emerged after 1619' (emphasis added). Such a statement triggers a controversy about the yardstick of inclusiveness. The USA exhibit widespread poverty, gross inequalities, structural discriminations and other poor institutional outcomes, while accommodating a small number of extremely wealthy elites, which are able to influence political decisions in their favour (cf. Haring and Douglas 2012 for an accessible source to the USA-as-a-plutocracy debate). As things stand, it may be worthwhile to consider classifying the United States as having *bad* institutions, since extractive processes (both domestically and internationally) seem to dominate over inclusive ones. In relation, note that such unfortunate outcomes escalated over the past decades, precisely under policy features (cf. Crippler 2012) advocated by Acemoglu and Robinson.

Finally, the book encompasses a contradictory usage of the ignorance hypothesis. Acemoglu and Robinson argue that '...if ignorance were the problem, well-meaning leaders would quickly learn what types of policies increased their citizens' incomes and welfare, and would gravitate toward those policies.' Instead, their position is one in which '...poor countries are poor because those who have power make choices that create poverty.' Later in the book, however, they conclude that the programs of the donor community and the IMF are '...based on an incorrect understanding of what causes poverty.' By doing so, they implicitly acknowledge the presence of ignorance. In fairness, it is plausible to make a distinction between international and national decision making, similar to away- and home-advantages. It is actually possible to trace out such a logic in the book, albeit seemingly unintentional: the ignorance hypothesis is not attached to decision-makers in poorer countries, but to institutions based in rich countries. Such a distinction is absurd and, more importantly, contradicts their explicit position. For instance, their portrayal of the donor community and the IMF is portrayed as rather naive, represented by the statement: 'Western nations feel guilt and unease...and foreign aid makes them feel something is being done..'. There are other, immensely more important, motivations for providing aid, for instance political and business influence.

There is also another reason to question the authors' complete dismissal of the ignorance hypothesis. Although it is easy to share this dismissal at the aggregate level, where decisions about the overall direction and distribution of investments are made, there are decision-making processes and levels that may easily involve limited knowledge and engagement. At the more concrete (micro) level, it is rather likely that decision makers and their members of staff do not possess the necessary skills to implement appropriate programs, especially in times of crises, aggressive lobbying, understaffing and disinformation campaigns.

In conclusion, the book is good read, displaying a compelling narrative of recent world economic history. Unfortunately, it is only semi-successful in meeting its objective of answering the question: *Why Nations fail*. The analytical framework would have benefitted from an explicit inclusion of the (extractive or inclusive) role of international institutions. It is likely that they play a greater role than domestic ones, especially within disempowered nations. After all, the global imbalances (between nations, institutions, corporations, individuals, etc) are immensely larger today than in the past few centuries (Maddison 2001). In addition, the book would have benefitted tremendously if the radical (anti-elitist) narrative was matched with people-centred policy recommendations. Moreover, there are divisive factors, such as colour-coded racism, that are too important to neglect. Finally, the book would have benefitted from an analysis of the notion that (actual or perceived) necessity is the mother of innovations, including institutions.

References

- Bond, P. and Manyanya, M. (2002) *Zimbabwe's Plunge: exhausted nationalism, neoliberalism and the search for social justice*. Durban: University of KwaZulu-Natal Press.
- Cripler, G. R. (2012) *Capitalizing on crises*. Cambridge: Harvard University Press.
- Haring, N. and Douglas, N. (2012) *Economists and the powerful: Convenient theories, distorted facts, ample rewards*. London: Anthem Press.
- Landes, D. S. (1999) *The wealth and poverty of nations: why some are so rich and some so poor*. New York: W. W. Norton & Company, Inc.
- Leibbrandt, M. et al. (2010) *Trends in South African income distribution and poverty since the fall of apartheid*, Paris: OECD and Cape Town: SALDRU.
- Kellecioglu, D. (2010) Why some countries are poor and some rich - a non-Eurocentric view. *Real-World Economics Review*, 52, pp. 40-53. Available from: <http://paecon.net/PAERreview/issue52/Kellecioglu52.pdf> [Accessed 6 December 2012]
- Maddison, A. (2001) *The World Economy – a millennial perspective*. Paris: OECD.
- Zinn, H. (2005) *A people's history of the United States: 1492 - present*. New York: HarperCollins Publishers.