Thirlwall, A. P.

Working Paper
Balance of payments constrained growth models: History and overview

School of Economics Discussion Papers, No. 11,11

Provided in Cooperation with:
University of Kent, School of Economics

Suggested Citation: Thirlwall, A. P. (2011) : Balance of payments constrained growth models: History and overview, School of Economics Discussion Papers, No. 11,11, University of Kent, School of Economics, Canterbury

This Version is available at:
http://hdl.handle.net/10419/50582

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Balance of Payments Constrained Growth Models: History and Overview

A. P. Thirlwall

May 2011

KDPE 1111
Balance of Payments Constrained Growth Models: History and Overview

A.P. Thirlwall
University of Kent

May 2011

Abstract

This paper surveys balance of payments constrained growth models from Thirlwall’s original contribution in 1979 to the latest tests of the model using cointegration techniques. Historical antecedents of the model are explored (e.g. the Harrod trade multiplier; dual gap analysis; Prebisch’s centre-periphery model), and various extensions of the model are outlined including: capital flows; interest payments on debt, and generalisation of the model to include many countries and many goods. All the empirical literature, using time series, panel and cross section data, is documented with discussion of the tests employed. The basic model that long run GDP growth can be approximated by the ratio of export growth to the income elasticity of demand for imports is remarkably robust. The relevance of the model is shown for the current discussion of global imbalances in the world economy.

Keywords:
Balance of Payments Constrained Growth; Harrod Trade Multiplier; North-South Models; Global Imbalances

JEL Classification: F02; F32; F43.

1 Paper prepared for Workshop on ‘Thirlwall’s Law and Balance of Payments Constrained Growth’, University of Coimbra, 24th-25th June 2011. The author is grateful to Gilberto Lima, Alberto Bagnai and John McCombie for helpful comments on an early draft of the paper.
Introduction

It is now over thirty years since my paper ‘The Balance of Payments Constraint as an Explanation of International Growth Rate Differences’ was first published in \textit{Banca Nazionale del Lavoro Quarterly Review}, March 1979. I am very grateful to the editor, Alessandro Roncaglia, for accepting it because it has generated a lot of interest, which surprises me in a way, because the basic idea is so simple, and isn’t new. In this introduction I’m going to describe the background to the formulation of the model (and its extensions), and to discuss the historical antecedents of the view that the balance of payments of a country matters for its long-run growth performance. I shall go back to: mercantilism; the attack on mercantilism; Keynes’s defence of mercantilism; Harrod’s foreign trade multiplier; Prebisch’s centre-periphery model; Chenery’s dual-gap model, and then go on to criticise orthodox trade theory for its neglect of the balance of payments consequences of trade, and orthodox growth theory for its neglect of the balance of payments.

Later in the paper I will outline the balance of payments constrained growth model and discuss the important recent extensions that have been made to the model to make it more realistic, and finally I will give a brief overview of some of the older and more recent empirical studies that have fitted the model to track the growth performance of countries, particularly developing countries. I remain convinced that foreign exchange is a major constraint on the growth performance of many poor countries, and that with improved export performance and a lower income elasticity of demand for imports, they would grow faster. Foreign exchange is a scarce resource; scarcer than saving in many contexts.

The History of an Idea

The 1979 paper grew out of work that I had been doing with Robert Dixon from Melbourne University in the early 1970s on Kaldor’s four-equation regional growth model (Kaldor, 1970) in which output growth is determined by the growth of exports; export growth is determined by income growth outside the region and by competitiveness; competitiveness is partly determined by
productivity growth, and productivity growth is determined by output growth (via Verdoorn’s Law). Kaldor defined a circular, cumulative process of regional growth (and regional growth rate differences) in the spirit of Myrdal’s theory of circular and cumulative causation (Myrdal, 1957) – in Kaldor’s case, a process driven by exports. Dixon and I were interested in formulating the model rigorously, and exploring its dynamic properties (Dixon and Thirlwall, 1975a, 1975b). To test the model requires regional data for countries which are not readily available. So we applied the model to the UK economy and found that it seriously over-predicted the UK historical growth rate. One obvious explanation was that the model contains no balance of payments constraint because imports are not modelled. But suppose the variables and parameters of the model generate a growth of imports which exceeds the growth of exports which is not sustainable? Something has to adjust. In a regional context, the issue doesn’t arise in the same way it does for a country because regions share a common currency and there is no exchange rate to defend, but in the case of countries there are only two forms of adjustment – either exchange rate depreciation (which may or may not work) or income adjustment. We knew that the UK had been plagued with balance of payments crises ever since the Second World War. The question then became how to incorporate imports and the balance of payments into the Kaldor model if the model is to be applied to countries? My colleague, Robert Dixon, returned to Australia, and it took three years for the penny to drop (in the garden, pruning roses, as it happens) that the best way to start the model is to begin with the basic long-run balance of payments equilibrium requirement on current account that the value of exports should equal the value if imports. Then export growth can be modelled (with or without cumulative causation); import growth can be modelled, and since import growth is a function of domestic GDP growth, it is possible to solve for GDP growth consistent with balance of payments equilibrium on current account. I tried the model out on my graduate students, and here is what one of them had to say (Hussain, 2006).

The Professor started to work out the mathematics of his manuscript. The good old blackboard notwithstanding, the identities and equations of the model were animated, left-handedly, in a manner that competes easily with Bill Gates’ Power-Point facilities. The model contained three basic equations representing the growth of imports, the growth of exports, and a dynamic
expression for the overall balance of payments equilibrium. He substituted the first two equations into the third and the model was solved to yield an elaborate expression of the growth rate of real gross domestic product (GDP). When the terms of trade were assumed to be constant the elaborate equation collapsed into an expression containing three symbols: $y = x/\pi$.

‘The rate of growth ($y$) of any country in the long run is equal to the growth rate of the volume of exports ($x$) divided by the income elasticity of demand for imports ($\pi$)’, he explained. Our eyes were fixed on the blackboard, attempting to digest the meaning and internalise the implications of this tri-legged animal. That job was not easy. For the animal distilled volumes of legendary work in economic development, encapsulating all of them in a small-sized anti-underdevelopment pill. The teaching of Engel’s law, which implies that the demand for primary goods increases less than proportionately to increases in global income: the Harrod foreign trade multiplier which put forward the idea that the pace of industrial growth could be explained by the principle of the foreign trade multiplier; that Marshall-Lerner condition which implies that a currency devaluation would not be effective unless the devaluation-induced deterioration in the terms of trade is more than offset by the devaluation-induced reduction in the volume of imports and increase in volume of exports; the Hicks’ super-multiplier which implies that the growth rate of a country is fundamentally governed by the growth of its exports; the Prebisch-Singer hypothesis which asserts that a country’s international trade that depends on primary goods may inhibit rather than promote economic growth; the Verdoorn-Kaldorian notion that faster growth of output causes a faster growth of productivity, implying the existence of substantial economies of scale; Kaldor’s paradox which observed that countries that experienced the greatest decline in their price competitiveness in the post-war period experienced paradoxically an increase in their market share and not a decrease; the literature on export-led growth which asserts that export growth creates a virtuous-circle through the link between output growth and productivity growth – all of these doctrines were somehow put into play and epitomised within this small-sized capsule.

Not only that but the capsule was sealed by the novel and powerful ingredient of the balance of payments constraint: ‘in the long run, no country can grow faster than that rate consistent with balance of payments equilibrium on current account unless it can finance ever-growing deficits which, in general, it cannot’.


The time for class discussion came and all the debate seemed to linger around one basic query: if growth could be explained by a rule which contained two variables only, what was the relevance of many other socio-economic variables that could also influence the growth process? What about the role of policies and economic management? What about the role of capital, labour and technical progress? The answers of the Professor were convincing to some students, but confusing to many others. In an attempt to relieve our baffled faces he concluded the discussion by saying in a pleasant fusion of smile and speech ‘Simple laws make good economics’. And as he was leaving the classroom, his smile turned gradually into a laugh that engulfed his remark: ‘if this rule comes to be known as Thirlwall’s Law, I will retire’. Less than one year after the publication of the manuscript in 1979 the rule was crowned as ‘Thirlwall’s Law’.2

The rule that \( y = \frac{x}{\pi} \) is also now known as the dynamic Harrod trade multiplier (see later).

So there you have the story first-hand. Mohammed Nureldin Hussain later went on to be the senior research economist in the Development Research Division of the African Development Bank and editor of the African Development Review. Tragically he died in 2005 aged 51, and Africa lost one of its best applied economists (for an appreciation of his work, see Thirlwall, 2006).

**Historical Views on the Balance of Payments and Growth**

Throughout history from Ricardo onwards, and in more modern times, from the birth of neoclassical growth theory (Solow, 1956), the orthodoxy has been that the balance of payments and growth of demand don’t matter for long run economic growth. Supply creates its own demand; the balance of payments is self-adjusting, and economic growth is supply-driven by the growth of factor inputs and technical progress, exogenously given. There has also grown up a divorce between real trade theory and the monetary, or balance of payments, consequences of trade. Trade, based on the law of comparative advantage, is always welfare-enhancing because it is assumed that the balance of payments looks after itself and full employment of resources is always maintained

---

2 The first paper to use the term was Skolka (1980), but in German: ‘Thirlwallschen Gesetz’
whatever the pattern of specialisation dictated by comparative advantage.

Before the orthodoxy developed that the balance of payments doesn’t matter for growth, however, there was a serious discussion of the relationship between trade and growth by the Mercantilists of the 16th and 17th centuries represented by Thomas Mun and Edward Misselden in England and by Antonio Serra in Italy, who argued that countries can become rich through running balance of payments surpluses and accumulating precious metals or foreign exchange which keeps the rate of interest low and encourages investment. This was the emphasis of Mun in his famous book *England’s Treasure by Foreign Trade* published in 1664. What is important is not treasure for its own sake, but its stimulus to investment by keeping the cost of borrowing money low. The doctrine of mercantilism came under attack, however, firstly from the Scottish philosopher David Hume; secondly from Adam Smith, and thirdly from David Ricardo.

Hume (1752) attacked mercantilism in his two essays ‘Of Money’ and ‘Of the Balance of Trade’ in which he argued that an increase in precious metals (gold and silver) would simply drive up the price level and have no real effects. This is the origin of the doctrine of the quantity theory of money; of the idea of money neutrality; of the so-called classical dichotomy, and particularly of the view that the rate of interest is a real, not a monetary, phenomenon. The balance of trade would have no real effects on the rate of interest.

Adam Smith, in his *Wealth of Nations* (1776), continued the attack, also by arguing that the mercantilists confused money and wealth, and that they were anti free trade. Neither of these critiques is fair to Mun, but the critique stuck. Mun was against trade protection because of fear of retaliation. It wasn’t until later that mercantilism tended to take on a protectionist stance for the promotion of infant industries and for the creation of domestic employment, although early on some mercantilists stressed the importance of industry relative to other activities. For example, the Italian mercantilist Serra (1613) identified three advantages of industry: it is more reliable because it is not dependent on the weather; it has a more secure market because industrial goods are not perishable, and (most significantly) Serra recognised that industry experiences increasing returns to scale. Industry can always be multiplied, as he put it, with *proportionately less expense* (*con minor proporzione di*
spesa’). It was not Nicholas Kaldor who first brought to the fore the role of manufacturing industry in the growth process, nor for that matter Adam Smith, but an Italian economist writing in the early 17th century.

But to return to the attack of Hume and Smith, the neutrality of money argument ignores two important considerations. Firstly if the rate of interest is partly a monetary phenomenon, money will have real effects working through variations in investment expenditure and the capital stock. Secondly, if there are unemployed resources, the impact of increases in the money supply will first be on output not on prices. It was, indeed, Keynes's view expressed in his General Theory (1936) that throughout history the propensity to save has been greater than the propensity to invest, and that pervasive uncertainty and the desire for liquidity has in general kept the rate of interest too high. Given the prevailing economic conditions of the 1930s when Keynes was writing, it was no accident that he should have devoted part of a chapter of the General Theory (chapter 23) to a defence of mercantilism as containing important germs of truth. In response to a comment from Roy Harrod on drafts of the General Theory, Keynes replied:

What I want is to do justice to schools of thought which the classicals have treated as imbeciles for the last hundred years and, above all, to show that I am not really being so great an innovator, except as against the classical school, but have important predecessors, and am returning to an age-long tradition of common sense (Moggridge, 1973).

The mercantilists recognised, like Keynes, that the rate of interest is determined by monetary conditions, and that it could be too high to secure full employment, and in relation to the needs of growth. As Keynes put it in the General Theory: ‘mercantilist thought never supposed as later economists did [e.g. Ricardo, and even Alfred Marshall] that there was a self-adjusting tendency by which the rate of interest would be established at the appropriate level [for full employment]’ (ibid. p.341).

Now let us turn to Ricardo. It was Ricardo in his Principles of Political Economy and Taxation (1817) who accepted and developed Say’s law of markets that supply creates its own
demand, and who for the first time expounded the theory of comparative advantage, which laid the early foundations for orthodox trade and growth theory that has prevailed (more or less) ever since. I won’t elaborate here on Ricardian theory, but rather mention two things that I think are wrong with it in the present context. First, Ricardian trade theory is *real* theory relating to the reallocation of real resources through trade which ignores the monetary aspects of trade; that is, the balance between exports and imports as trade takes place. In other words, it ignores the balance of payments effects of trade that arises as a result of trade specialisation, and the feedback effects that the balance of payments can have on the real economy. Secondly, continuous full employment is assumed because supply creates its own demand through variations in the real rate of interest. But in relation to trade, as Keynes put it: ‘free trade assumes that if you throw men out of work in one direction you re-employ them in another. As soon as that link in the chain is broken the whole of the free trade argument breaks down’ (Moggridge, 1973). In other words, the real income gains from specialisation may be offset by the real income losses from unemployment.

To return to the balance of payments, suppose that payments deficits arise in the process of international specialisation and the freeing of trade, and the rate of interest has to be raised to attract foreign capital inflows to finance them. Or suppose deficits cannot be financed and income has to be deflated to reduce imports. The balance of payments consequences of trade may offset the real income gains from trade.

This raises the question of why the orthodoxy ignores the balance of payments? There are several reasons, both old and new, that all relate to the balance of payments as a self-adjusting process, or simply as a mirror image of autonomous capital flows, with no income adjustment implied. In classical times, till the First World War, the mechanism was the gold standard. The balance of payments was supposed to be self-equilibrating because countries in surplus, accumulating gold, would lose competitiveness through rising prices (Hume’s quantity theory of money), and countries in deficit losing gold would gain competitiveness through falling prices. The balance of payments was assumed effectively to look after itself through relative price adjustments without any change in income or output.
After the external gold standard collapsed in 1931, the theory of flexible exchange rates was developed, and it was shown that if the real exchange rate is flexible, and the so-called Marshall-Lerner condition is satisfied (i.e. the sum of the price elasticities of demand for exports and imports is greater than unity), the balance of payments will equilibrate; again, without income adjustment.

In modern theory, balance of payments deficits are assumed to be inherently temporary as the outcome of inter-temporal decisions by private agents concerning consumption. Deficits are the outcome of rational decisions to consume now and pay later. Deficits are merely a form of consumption smoothing, and present no difficulty for countries.

Finally there is the Panglossian view that the current account of the balance of payments is of no consequence at all because it simply reflects the desire of foreigners to invest in a country. Current account deficits should be seen as a sign of economic success, not as a weakness.

It is not difficult to question and criticise each of these ideas that the balance of payments looks after itself, or doesn’t have consequences for long-run growth. As far as the old gold standard mechanism is concerned, monetary historians such as Triffin (1964), Cooper (1982) and McClosky and Zecher (1976) have all shown that instead of the price levels of deficit and surplus countries moving in opposite directions, there was a tendency in the 19th century for the price levels of countries to move together in the same direction. In practice, it wasn’t movements in relative prices that equilibrated the balance of payments but expenditure and output changes associated with interest rate differentials. Interest rates rose in deficit countries which deflated demand and output, and fell in surplus countries stimulating demand. Harrod developed in the early 1930s (Harrod, 1933) the static foreign trade multiplier showing that if the real terms of trade (or real exchange rate) stays constant, it is income changes that bring exports and imports into line with another as a result of an autonomous change in imports or exports (see below). But even in the late 1930s, very few economists were teaching this story. An exception was Barrett Whale at the London School of Economics (see Barrett Whale 1932, 1937).
On the question of flexible exchange rates as an equilibrating device, a distinction first needs to be made between the nominal exchange rate and the real exchange rate. It is easy for countries to adjust the nominal rate, but not so easy to adjust the real rate because competitors may ‘price to market’ or retaliate, and domestic prices may rise with a nominal devaluation. Secondly, the Marshall-Lerner condition then has to be satisfied for the balance of payments to equilibrate. This may not be the case in the short run, or because of the nature of goods exported and imported by a particular country. The international evidence over the last forty years since the breakdown of the Bretton Woods fixed exchange rate system in 1971 suggests that exchange rate changes are not an efficient balance of payments adjustment weapon. Currencies appreciate and depreciate and still massive global imbalances of payments remain.

On the inter-temporal substitution effect, it is wrong to give the impression that inter-temporal shifts in consumption behaviour don’t have real effects, particularly if interest rates have to rise to finance deficits caused by more consumption in the present if countries don’t want their exchange rate to depreciate.

Lastly, on the view that deficits are a sign of success, an important distinction needs to be made between types of capital inflows. If the capital flows are autonomous, such as foreign direct investment, the argument is plausible, but if they are ‘accomodating’ in the form of loans from the banking system or the sale of securities to foreign governments and international organisations, the probable need to raise interest rates will again have real effects by reducing investment and output domestically.

Challenges to the Orthodoxy

In the post-war years, there have been very few serious challenges to the orthodoxy that the balance of payments doesn’t matter for growth. The three most important ones have been Harrod’s 1933 static foreign trade multiplier; Prebisch’s 1950 centre-periphery model, and Chenery’s 1960 dual-gap model. I will say a few words about each of these.
Harrod’s foreign trade multiplier of $1/m$, where $m$ is the marginal propensity to import, predates Keynes’s investment multiplier of $1/s$, where $s$ is the marginal propensity to save. Harrod derived his multiplier on the assumptions that (i) income is generated by the production of consumption goods ($C$) and exports ($X$), so $Y = C + X$; (ii) all income is spent on consumption goods and imports ($M$), so $Y = C + M$, and (iii) the real terms of trade are constant. Therefore $X = M$ (balanced trade). If $M = \bar{M} + mY$, where $M$ is autonomous imports and $m$ is the marginal propensity to import, we have:

$$Y = \frac{(X - \bar{M})}{m} \quad (1)$$

and, therefore,

$$\frac{\partial Y}{\partial (X - \bar{M})} = \frac{1}{m} \quad (2)$$

Any change in $X$ or $\bar{M}$ always brings the balance of trade back into equilibrium, but through changes in income not through a change in relative prices. Clearly the balance of payments matters for income determination, and by extension (see later) for growth.

The assumptions of no saving and investment, and no government spending and taxation, are clearly unrealistic but these assumptions can be relaxed and the same result obtained if, first, all saving is done for investment or if all investment is assumed to generate its own saving, and governments run balanced budgets, or, second, if any surplus/deficit in the private sector is exactly offset by a corresponding deficit/surplus in the public sector. If ‘leakages’ exceed ‘injections’ exports will exceed imports and there is no balance of payments constraint, but if ‘injections’ exceed ‘leakages’ there will be a payments deficit and the question then is how long the deficit can persist without corrective action having to be taken. If relative price changes (including exchange rate changes) are ineffective, output would have to be depressed through government contraction of demand. We would be back in a Harrod trade multiplier world, with the level (and growth) of income fundamentally determined by the level (and growth) of export demand in relation to the propensity to import. One of the surprising things about Harrod’s analysis, however, is that he never managed to develop the growth implications of his model and to integrate a ‘balance of payments constrained
growth rate’ with his famous growth and cycle model (Harrod, 1939) of the relation between the actual, warranted and natural rates of growth (see Thirlwall, 2001, for a discussion of this issue).

Raul Prebisch (1950, 1959) was the first economist in the post-war era to seriously question the doctrine of the mutual profitability of free trade between developed and developing countries. The traditional approach of the measurement of the gains from trade is from the classical standpoint of real resource augmentation from specialisation which trade permits. By contrast, Prebisch concentrated attention on the monetary, or balance of payments, aspects of trade, arguing that the real resource gains from specialisation may be offset by the underutilisation of resources if foreign exchange is the dominant constraint on output. The losers are the less developed countries which tend to specialise in diminishing returns activities with a low income elasticity of demand in world markets i.e. land-based primary products, and the gainers are the developed countries specialising in increasing returns activities with a higher income elasticity of demand i.e. processed manufactured goods. Prebisch illustrated the nature of the problem with a simple numerical example. Assume two countries: a less developed country (LDC) exporting solely primary commodities with an average income elasticity of demand of 0.8 ($\varepsilon_{ldc} = 0.8$) and a developed country (DC) exporting solely manufactured goods with an average income elasticity of demand of 1.3 ($\varepsilon_{dc} = 1.3$). The export elasticity of the LDC is the import elasticity of the DC ($\pi_{dc} = 0.8$), and the export elasticity of the DC is the import elasticity of the LDC ($\pi_{ldc} = 1.3$). If both countries grow at the same rate, the situation is clearly not sustainable. For example, at a growth rate of 5 percent, the growth of imports of the LDC would be $5 \times 1.3 = 6.5$ percent, while the growth of exports would only be $5 \times 0.8 = 4$ percent. The LDC would be in perpetual deficit while the DC would be in perpetual surplus. Balance of payments equilibrium of the LDC requires that the growth of output be constrained so that imports grow no faster than exports. The constrained growth rate would be equal to:

$$g_{LDC} = \frac{x_{LDC}}{\pi_{LDC}} = \frac{\varepsilon_{LDC}}{\pi_{LDC}} \times g_{DC} = \frac{5 \times 0.8}{1.3} = 3.1 \text{ per cent} \quad (3)$$
Equilibrium balance of payments in both countries implies 3.1 percent growth in the LDC compared with 5 percent in the DC. The relative growth rates of the two countries is given by rearranging the above equation (3) to give;

\[
\frac{g_{LDC}}{g_{DC}} = \frac{\epsilon_{LDC}}{\pi_{LDC}} = \frac{0.8}{1.3} = 0.6
\]

(4)

On the assumptions here, the LDC is constrained to grow at only 60 percent of the growth of the DC.

Prebisch’s equation (4), which is the basis for his classic centre-periphery model, can be shown to be the dynamic analogue of the static Harrod foreign trade multiplier result discussed earlier\(^2\), and the true forerunner of my balance of payments constrained growth model developed much later. It is also the basis of Krugman’s 45-degree rule that one country’s growth rate relative to another’s will be equi-proportional to the ratio of its income elasticities of demand for exports and imports if the real exchange rate is constant (Krugman, 1989). In Krugman’s model, however, the direction of causation is implausibly reversed from growth rate differences to differences in income elasticities, but this is another story (see Thirlwall, 1991).

Now let us turn to the concept of dual-gap analysis. There is clearly a link between the Prebisch model above and models of dual-gap analysis applied to developing countries, made famous by Hollis Chenery and his associates in the 1960s (see, for example, Chenery and Bruno, 1962). The essence of dual-gap analysis is to show that growth may be constrained either by domestic saving or by foreign exchange, and that the role of foreign borrowing in the development process is to relieve whichever is the dominant constraint. Chenery’s view, like Prebisch’s, was that for most developing countries, at least in the intermediate stage of economic development, the dominant constraint is likely to be a shortage of foreign exchange associated with balance of payments deficits, so that growth would be balance of payments constrained.

\(^2\) The proof is: from Harrod \(\Delta Y = \Delta X/m = \Delta X/(\Delta M/\Delta Y)\). Multiplying the l.h.s. by \(X/Y\) and the r.h.s by \(M/Y\) (since \(X=M\)), we have \(\Delta Y/(X/Y) = (\Delta M/(\Delta M/\Delta Y))(M/Y)\) or \(\Delta Y/Y = (\Delta X/X)/(\Delta M/M)/(\Delta Y/Y)\). Therefore, \(y = x/\pi\), where \(y\) is the growth of income, \(x\) is the growth of exports and \(\pi\) is the income elasticity of demand for imports.
Consider the following model. Growth requires investment goods which may either be provided domestically or purchased from abroad. The domestic provision requires saving; the foreign provision requires foreign exchange. If it is assumed that some of the investment goods for growth can only be provided from abroad, there is always a minimum amount of foreign exchange required to sustain the growth process. In Harrod’s (1939) growth model, the relation between growth and saving is given by the incremental capital-output ratio \( c \), which is the reciprocal of the productivity of capital \( p \), i.e. \( y = s/c \) or \( y = sp \), where \( y \) is the growth rate and \( s \) is the savings ratio. Likewise the growth rate can be expressed as the product of the incremental output-import ratio \( \Delta Y/M = m' \) and the ratio of investment-good imports to income \( M/Y = i \), i.e. \( y = im' \). If there is a lack of substitutability between domestic and foreign resources, growth will be constrained by whatever factor is most limiting – domestic saving or foreign exchange. Suppose, for example, that the growth rate permitted by foreign exchange is less than the growth rate permitted by domestic saving. In this case, growth would be foreign exchange constrained, and if the constraint is not lifted there will be unemployed domestic resources and a proportion of domestic saving will be unused. For instance, suppose that the product of the import ratio \( i \) and the productivity of imports \( m' \) gives a permissible growth rate of 3 percent, and the product of the savings ratio \( s \) and the productivity of capital \( p \) gives a permissible growth rate of 4 percent. Growth is constrained to 3 percent, and for a given \( p \) a proportion of savings available cannot be absorbed. Ways must be found of using unused domestic resources to earn more foreign exchange and/or to raise the productivity of imports. As many developing countries will testify, however, this is easier said than done. It is not easy to sell more on world markets if external conditions are unfavourable and price elasticities are low.

The correspondence between the dynamic Harrod trade multiplier result of \( y = x/\pi \), and the Chenery model of \( y = im' \) is immediately apparent. If balance of payments equilibrium is a requirement so that an increase in imports for growth requires an increase in exports, an increase in \( x \) will raise \( i \) and the foreign exchange constraint is relaxed. Orthodox economists criticised the model because of its rigid assumptions relating to the lack of substitutability between types of imports and between foreign and domestic resources. Yet we continue to witness in the world so many developing
countries with serious balance of payments difficulties, desperate for foreign exchange, which could
grow faster if the foreign exchange constraint was relaxed. The critics do not have the evidence on
their side.

Garcia-Molina and Ruiz-Tavera (2009-10) have recently attempted to combine the balance of
payments constrained growth model (to be outlined in detail below) with the Chenery and Bruno two-
gap model in what they call a ‘unified dynamic gap model’. Their novel approach is to make dynamic
the investment-savings gap equation of $S = I + XP_d - MP_{fd}$, where $S$ is savings, $I$ is investment, $X$ is
export volume, $P_d$ is the price of exports, $M$ is imports and $P_{fd}$ is the price of imports in domestic
currency, and then substituting expressions for the growth of exports and imports (from equations (7 )
and (8 ) below). In this savings-constrained growth equation, output growth is then also a function of
export growth and the income elasticity of demand for imports, as in the conventional external gap
model, made dynamic.

The interaction of the two gaps leads to the conclusion that changes in variables may have
different effects depending on the starting point of the economy – savings-constrained or foreign
exchange-constrained. For example, a rise in the income elasticity of demand for imports worsens a
foreign exchange-constrained economy, but in the case of a savings-constrained economy, higher
imports relaxes the savings constraint on growth.

**The 1979 Balance of Payments Constrained Growth Model**

My 1979 model started from the proposition that no country can grow faster than that rate
consistent with balance of payments equilibrium on current account unless it can finance ever-
growing deficits, which in general it cannot. There is a limit to the deficit/GDP ratio, and international
debt/GDP ratio, beyond which financial markets get nervous. The simplest way to model growth
within a balance of payments constrained framework, therefore, is to start with the balance of
payments equilibrium condition; specify export and import demand functions, and since import
growth is a function of domestic income growth, solve the model for the growth rate consistent with
long-run balance of payments equilibrium. This gives equations (5) to (8) below.
Current account equilibrium is given by:

\[ P_d X = P_f M E \]  \hspace{1cm} (5)\]

where \( X \) is exports; \( M \) is imports; \( P_d \) is the domestic price of exports; \( P_f \) is the foreign price of imports, and \( E \) is the exchange rate measured as the domestic price of foreign currency.

Export and import demand functions are specified as multiplicative with constant elasticities giving:

\[ X = a \left( \frac{P_d}{P_f} \right)^\eta Z^\varepsilon, \quad \eta < 0, \quad \varepsilon > 0 \]  \hspace{1cm} (6)\]

\[ M = b \left( \frac{P_f E}{P_d} \right)^\Psi Y^\pi, \quad \Psi < 0, \quad \pi > 0 \]  \hspace{1cm} (7)\]

where \( \eta \) is the price elasticity of demand for exports; \( \varepsilon \) is the income elasticity of demand for exports; \( \Psi \) is the price elasticity of demand for imports; \( \pi \) is the income elasticity of demand for imports; \( Z \) is world income, and \( Y \) is domestic income.

Taking logarithms of equations (6) and (7), differentiating with respect to time, substituting the growth of exports and imports into equation (5) in growth rate form, and solving for the growth of income, gives:

\[ y_B = \frac{\left( 1 + \eta + \Psi \right) (P_d - P_f - E) + \varepsilon (Z)}{\pi} \]  \hspace{1cm} (8)\]

where lower-case letters stand for the growth rates of variables.

Equation (8) expresses a number of familiar economic propositions:

(i) An improvement in the real terms of trade (or real exchange rate) by itself, \( (P_d - P_f - E) > 0 \), will raise the growth of income consistent with balance of payments equilibrium.

(ii) If the sum of the price elasticities of demand for exports and imports is greater than -1, however, an improvement in the real terms of trade (or a deterioration in competiveness), \( (P_d - P_f - E) > 0 \), will worsen the growth rate consistent with balance of payments equilibrium.
(iii) A depreciation of the exchange rate, $e > 0$, will improve the growth rate if $(\eta + \psi) > -1$.

This is the Marshall-Lerner condition for a successful devaluation. Note, however, that a once-for-all depreciation (or devaluation) will not put a country on a permanently higher growth path. For this to happen, the depreciation would either have to be continuous, or affect the parameters of the model favourably.

(iv) One country’s growth rate is dependent on other countries’ growth rates ($z$), but how fast one country grows relative to others depends crucially on the income elasticity of demand for exports, $\varepsilon$. This depends on the structure of production and exports, as we saw earlier in discussing the Prebisch centre-periphery model.

(v) A country’s growth rate consistent with balance of payments equilibrium is inversely related to its appetite for imports, $\pi$. This also a function of the structure of production and imports.

If relative prices in international trade, or real exchange rates, are constant, equation (8) reduces to:

$$y_B^* = \frac{\varepsilon(z)}{\pi}$$

(9)

and, on the same assumption,

$$y_B^{**} = \frac{x}{\pi}$$

(10)

I showed this result to my colleague at the University of Kent, Charles Kennedy (who had been a friend of Roy Harrod in Oxford for many years), and he said to me that this looks like a dynamic version of Harrod’s static foreign trade multiplier. To my shame, I had not been familiar with the Harrod trade multiplier result, but it transpired that I had reinvented the wheel in dynamic form (although, as I have indicated, Prebisch got there first, but he never tested the model empirically).

Perraton (2003) has called equation (9) the ‘strong’ version of Thirlwall’s Law, and equation (10) the ‘weak’ version because if the parameter $\varepsilon$ has not been estimated, using equation (6), then export growth ($x$) must also include the effect of relative price changes as well as the effect of world
income growth which weakens somewhat the argument that the balance of payments is always brought into equilibrium by domestic income changes. The model is best tested, therefore, using the ‘strong’ version if robust estimates can be made of $\varepsilon$.

The model can be neatly illustrated in diagrammatic form, as in Figure 1.

**Figure 1**
The Balance of Payments and Growth

GDP growth is measured on the horizontal axis and export and import growth on the vertical axis. Export growth ($x$) is autonomous, while import growth ($m$) is a function of GDP growth according to the income elasticity of demand for imports ($\pi$), appropriately estimated from equation (7) and controlling for the effect of relative price changes (otherwise $y = x/\pi$ becomes a tautology, as first pointed out by McCombie, 1981). The GDP growth rate consistent with balance of payments equilibrium is defined where the $x$ and $m$ curves cross. The higher the $x$ curve and the flatter the $m$ curve, the higher the equilibrium growth rate will be, and vice versa.

The question is, how well does the simple rule in equations (9) or (10) fit the data? The proof of the pudding is always in the eating! I originally (in 1979) applied equation (10) to a selection of developed countries over the time periods 1951-73 and 1953-76 (using other peoples’ data sets and estimates of $\pi$ – so as not to be accused of ‘cooking the books’) and found a remarkable correspondence between the actual growth experience of countries and the growth rate predicted from
the balance of payments constrained growth model. Table 1 gives the original data and results for the period 1951-73.

Table 1
Calculations of the Growth Rate Consistent with Balance of Payments Equilibrium, 1951-1973

<table>
<thead>
<tr>
<th>Country</th>
<th>Change in GDP %</th>
<th>Change in Exports (x) %</th>
<th>Income Elasticity of Demand for Imports (π)</th>
<th>Balance of Payments Equilibrium Growth Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>5.1</td>
<td>10.7</td>
<td>n.a.</td>
<td>-</td>
</tr>
<tr>
<td>Belgium</td>
<td>4.4</td>
<td>9.4</td>
<td>1.94</td>
<td>4.84</td>
</tr>
<tr>
<td>Canada</td>
<td>4.6</td>
<td>6.9</td>
<td>1.20</td>
<td>5.75</td>
</tr>
<tr>
<td>Denmark</td>
<td>4.2</td>
<td>6.1</td>
<td>1.31</td>
<td>4.65</td>
</tr>
<tr>
<td>France</td>
<td>5.0</td>
<td>8.1</td>
<td>1.62</td>
<td>5.00</td>
</tr>
<tr>
<td>Germany</td>
<td>5.7</td>
<td>10.8</td>
<td>1.89</td>
<td>5.71</td>
</tr>
<tr>
<td>Italy</td>
<td>5.1</td>
<td>11.7</td>
<td>2.25</td>
<td>5.20</td>
</tr>
<tr>
<td>Japan</td>
<td>9.5</td>
<td>15.4</td>
<td>1.23</td>
<td>12.52</td>
</tr>
<tr>
<td>Netherlands</td>
<td>5.0</td>
<td>10.1</td>
<td>1.82</td>
<td>5.55</td>
</tr>
<tr>
<td>Norway</td>
<td>4.2</td>
<td>7.2</td>
<td>1.40</td>
<td>5.14</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2.7</td>
<td>4.1</td>
<td>1.51</td>
<td>2.71</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>3.7</td>
<td>5.1</td>
<td>1.51</td>
<td>3.38</td>
</tr>
</tbody>
</table>


If relative price changes were an efficient balance of payments adjustment mechanism, no necessary relation between the two rates is to be expected. The results seemed to vindicate the assumptions of the model; that indeed relative prices don’t change in the long run, or relative price changes don’t work as an adjustment mechanism. The rank correlation between the actual and predicted growth rates of countries for the 1951-73 period is 0.891 (and for the 1953-76 period, 0.764), and the mean deviation of the actual from the predicted rates over both samples (excluding Japan) is only 0.56 percentage points.

But neither rank correlation nor mean deviations of actual from predicted values are parametric tests. Since these early days, however, three main parametric tests have been developed. One, first suggested and used by McGregor and Swales (1985), is to run a regression across countries
of \( y \) on \( y_{B*} \) or \( y_{B**} \) and test whether the constant is zero and the regression coefficient doesn’t differ significantly from unity. The problem with this test, however, is that it requires a full set of countries for the whole world in which deficits and surpluses cancel out. If the only countries taken are predominantly deficit countries, the constant would exceed zero, and the slope is likely to differ from unity, and the hypothesis that growth is balance of payments constrained would be rejected because of inappropriate sample choice. Also, one or two outliers (such as Japan in the 1950s and 1960s, or oil producers) running huge payments surpluses are enough to upset the relationship. It only requires a few countries not to be balance of payments constrained for all the rest to be so, yet the existence of outliers would lead to a statistical rejection of the hypothesis.

The second parametric test, and much more satisfactory, is the McCombie (1989) test which calculates the income elasticity of demand for imports (say, \( \pi^* \)) that would make \( y = y_{B*} \) or \( y_{B**} \), and if there is no statistically significant difference between \( \pi^* \) and the estimated \( \pi \), then \( y_{B*} \) or \( y_{B**} \) will be a good predictor of \( y \). When this test has been performed on individual countries and samples of countries (see, for example, Hussain, 1999; Perraton, 2003) the balance of payments constrained growth model performs well in the majority of cases.

A third test, suggested by Alonso (1999) for an individual country, is to use cointegration techniques to estimate \( \pi^* \) taking levels of variables in the export and import demand functions. The level of income consistent with balance of payments equilibrium is then calculated and the actual growth of income is regressed on this ‘equilibrium’ level of income. If the constant is not significantly different from zero and the regression coefficient is not significantly different from unity, this indicates parallel evolution of the two series. Alonso found this to be the case for Spain over the period 1960 to 1994.

**The Harrod Trade Multiplier and the Hicks Super-multiplier**

As a component of demand, exports are unique. They are the only component of demand that can pay for the import content of other components of demand such as consumption, investment, government expenditure and exports themselves. If any of these components of demand increase,
while exports are static, the balance of payments will worsen, and growth becomes demand constrained. In other words, exports allow all components of demand to grow faster than otherwise would be the case. The dynamic Harrod trade multiplier is likely to be picking up these induced ‘multiplier’ effects. McCombie (1985) has shown how the dynamic Harrod trade multiplier can be interpreted as a Hicks super-multiplier. The model is as follows. Let

\[ Y = \frac{(X+E)}{k} \]  
(11)

where \( Y \) is income; \( X \) is exports; \( E \) is other autonomous expenditure, and \( 1/k \) is the ordinary Keynesian multiplier (including the propensity to import). From equation (11) we have:

\[ y = \frac{(a_X x + a_E e)}{k} \]  
(12)

where lower-case letters are the growth rate of variables and \( a_X \) and \( a_E \) are the shares of exports and other autonomous expenditure in income, respectively. If the only increase in autonomous expenditure comes from exports, the growth of income is given by:

\[ y = \frac{(a_X x)}{k} \]  
(13)

This represents the direct impact of export growth on income growth, operating through the traditional multiplier (\( k \)). But the maximum growth rate consistent with balance of payments equilibrium (starting from equilibrium, \( X=M \)) is given by equation (10):

\[ y^* = \frac{x}{\pi} \]  
(14)

This may be written equivalently as:

\[ y^* = \frac{(a_X x)}{m} \]  
(15)

where \( a_X = \frac{X}{Y} (=M/Y) \) and \( m = \frac{\Delta M}{\Delta Y} \).

Since \( k > m \), it can be seen from a comparison of equations (13) and (15) that the balance of payments equilibrium growth rate is higher than income growth determined solely by the growth of exports. The growth of exports allows other components of expenditure to grow faster than otherwise would be the
case. How much faster is given by taking the difference between equations (12) and (15) and solving for $e$ (the growth of other autonomous expenditure) which gives:

$$e = k \left( \frac{1}{m} - \frac{1}{k} \right) \left( \frac{a_X}{a_E} \right) x$$  \hspace{1cm} (16)

If autonomous expenditure growth is less than this, the balance of payments will have a (growing) surplus, and the growth rate will be below the rate consistent with balance of payments equilibrium, and vice versa.

It is big surpluses, and big deficits financed by capital inflows, combined with volatile terms of trade movements, that sometimes upsets the predictions of the basic model (see later).

**Reconciling the Balance of Payments Equilibrium Growth Rate with the Capacity Rate**

The balance of payments equilibrium growth rate is by definition a demand-constrained growth rate. It would only be by chance that it equals the capacity, or supply-constrained, growth rate determined by the availability of factor supplies. Palley (2003) raises the issue of how the long-run equilibrium growth rate of the economy is determined which avoids ever-increasing over-capacity utilisation if $y_B > y_C$ or under-capacity utilisation if $y_B < y_C$. Palley undermines the demand-constrained model by making the income elasticity of demand for imports endogenous to the degree of capacity utilisation, rising with the degree of over-capacity utilisation, which therefore pulls down $y_B$ to the capacity rate\(^4\). Supply growth, as in neoclassical growth theory, determines long run equilibrium growth.

Setterfield (2006), however, offers an alternative adjustment mechanism in which induced productivity growth (sometimes called the Verdoorn effect) is a function of the degree of capacity utilisation so that $y_B > y_C$ raises the potential growth of output towards the balance of payments equilibrium growth rate. Supply adjusts to demand, and demand ‘rules the roost’. Induced increases in

\(^4\) There is some evidence that labour market bottlenecks and demand disequilibrium between markets raises the level of imports. See White and Thirlwall (1974) and Hughes and Thirlwall (1979). For a rigorous critique of the Palley model, see McCombie (2011).
labour supply would have the same effect. There is now a lot of empirical evidence that capacity growth (or Harrod’s natural rate of growth) is endogenous to demand (for a survey of results for OECD countries, Latin America, and Asia, see Dray and Thirlwall, 2010).

Another mechanism of reconciliation might be the interaction between the growth of exports and investment. When export growth is high and $y_{b} > y_{c}$, investment is encouraged putting up $y_{c}$ and vice versa.4

In practice, both mechanisms suggested by Palley and Setterfield are likely to operate, but for a stable equilibrium they must work to reconcile the two growth rates within strict bounds because the degree of capacity utilisation cannot fall below zero or exceed unity.

Nell’s Generalisation of the Model to Many Countries

In the basic balance of payments equilibrium growth model, an individual country exports to, and receives imports from, the rest of the world ($Z$). In practice, however, a country exports to, and imports from, many different destinations. This requires disaggregation of the model to allow for several countries. This leads to the ‘generalisation’ of equation (8) first suggested by Nell (2006) which disaggregates the world income growth variable ($z$), and takes into account the different income elasticities of demand for exports and imports to and from each trading partner ($p$):

$$y = \frac{(1 + \psi + \eta)(p_{d} + p_{f} - e) + \sum_{p} w_{xp} \varepsilon_{p}(y_{p})}{\sum_{i=1}^{n} w_{mp} \pi_{p}}$$

where $y_{p}$ is the growth rate of the trading partner ($p = 1 - - - n$); $w_{xp}$ is the share of exports to country $p$ in total exports; $\varepsilon_{p}$ is the income elasticity of demand for exports to each destination ($p$); $\pi_{p}$ is the income elasticity of demand for imports from each trading partner ($p$), and $w_{mp}$ is the share of imports of each sector in total imports.

5 I owe this point to Gilberto Lima.
When such a model is estimated empirically, it may be found that a country is balance of payments constrained with respect to some countries but not others. Nell himself estimates the disaggregated model for South Africa, the rest of the Southern African Development Community (RSADC) and the OECD, and finds that South Africa is balance of payments constrained with respect to the OECD, while RSADC is balance of payments constrained with respect to South Africa. Such a disaggregated approach highlights the need for different policies with regard to trade relations with partner countries.

**Araujo and Lima’s Multi-Sectoral Model**

Another aspect of the basic model is that it aggregates all exports and imports together. The income elasticities of demand for exports and imports, which ‘drive’ the model, are aggregate elasticities, but in practice, of course, they are weighted averages of sectoral elasticities. Pasinetti’s (1981, 1993) structural economic dynamics (SED) recognises explicitly the role of demand-led structural change in economic growth, but in his model there is no explicit balance of payments constraint on demand.

Araujo and Lima (2007) use Pasinetti’s SED framework to develop a disaggregated multi-sectoral version of the balance of payments constrained growth model (excluding changes in price competitiveness), as in equation (18):

\[ y = \frac{\sum_{i=1}^{n} w_{xi} \varepsilon_i(x)}{\sum_{i=1}^{n} w_{mi}\pi_i} \]  

(18)

where \( \varepsilon_i \) is the income elasticity of demand for exports of industry i (i = 1 - - - n); \( \pi_i \) is the income elasticity of demand for imports of industry i; \( w_{xi} \) is the share of industry i in total exports, and \( w_{mi} \) is the share of industry i in total imports.

What the multi-sectoral model highlights clearly is that even if sectoral elasticities are constant and there is no change in world income growth, a country can grow faster by shifting resources to sectors with higher income elasticities of demand for exports and away from sectors with
a high income elasticity of demand for imports. This is what import substitution and export promotion policies are meant to achieve. Equally, it shows that if there is an increase in world income, a country will benefit more the higher its sectoral income elasticities of demand for exports and the lower its sectoral income elasticities of demand for imports. From a policy point of view, this multi-sectoral specification of the model allows for the identification of key, strategic, growth-promoting tradeable-goods sectors of the economy.

Gouvea and Lima (2010) test this multi-sectoral model for four Latin American countries (Argentina, Brazil, Colombia and Mexico) and four Asian countries (South Korea, Malaysia, Philippines and Singapore) over the period 1962 – 2006, and compare their findings with the simple model. Six sectors are distinguished: primary products; resource-based manufacturing; low technology manufactures; medium technology manufactures; high technology manufactures, and others. Export and import demand functions are estimated for each sector, as well as an aggregate import demand function. In general, technology-intensive sectors have a higher income elasticity of demand for exports, but for imports there is not much of a difference between sectors. The balance of payments constrained growth rate is then estimated from the multi-sectoral model and the aggregate model. For some countries, the multi-sectoral model has a higher predicted error than the aggregate model, but for both groups of countries the mean absolute error is lower for the multi-sectoral model.

More importantly, the authors use the sectoral elasticities to estimate the year by year evolution of the aggregate income elasticities of exports and imports to show how structural change impacts on the balance of payments constrained growth rate. For Latin America, except Mexico, the ratio of the sectorally-weighted income elasticities of demand for exports and imports has hardly changed over the long period, but in Asia it has risen, thereby impacting favourably on the balance of payments constrained growth rate of these countries.

In a broader study of twenty-nine developed and developing countries, using this sectoral approach, Cimoli, Porcile and Rovira (2010) show that the developing countries that succeeded in reducing the income gap between themselves and developed countries were those that transformed
their economic structure towards sectors with a higher income elasticity of demand for exports relative to imports; to sectors with what they call ‘higher Schumpeterian and Keynesian efficiency’. Schumpeterian efficiency refers to products with superior technical characteristics, while Keynesian efficiency refers to the superior demand characteristics of goods.

Capital Flows

Even while the 1979 model was being formulated, I realised that the model might not fit well the developing countries because many of them are allowed to run balance of payments deficits for considerable periods of time financed by various types of capital inflows. This led to extending the model to include capital inflows (Thirlwall and Hussain, 1982) – although without limit (see later). By definition, the overall balance of payments of a country, including current and capital transactions, must balance, so we can write as the starting point:

\[ P_dX + C = P_f M E \]  

which is a simple extension of equation (5) where \( C > 0 \) represents positive capital inflows. Taking rates of change of equation (19), and using expressions for the rate of growth of exports and imports from equations (6) and (7), gives the growth rate consistent with overall balance of payments (ob) of:

\[ y_{ob} = \left[ (1 + \theta \eta + \psi) (p_d - p_f - e) + \theta \varepsilon(z) + (1 - \theta) (c - p_d) \right] / \pi \]  

where \( \theta \) is the share of export receipts in total receipts to pay the import bill; \( c \) is the growth of nominal capital inflows, so that \( c - p_d \) is the growth of real capital inflows.

It can be readily seen from equation (20) that any country’s growth rate can in principle be disaggregated into four component parts:

(i) growth associated with real terms of trade movements: \( (p_d - p_f - e)/\pi \)

(ii) growth associated with terms of trade movements combined with the price elasticities of exports and imports: \( [(1 + \theta \eta + \psi) (p_d - p_f - e)]/\pi \)

(iii) growth related to exogenous changes in income growth abroad: \( \theta \varepsilon(z)/\pi \)
(iv) growth effects of real capital flows: \[\frac{[(1 – \theta)(c – p_d)]}{\pi}\]

If we make the assumption, as before, that relative prices measured in a common currency remain unchanged over the long term, equation (20) reduces to:

\[y_{ob}^* = \frac{\theta x + (1 – \theta)(c - p_d)}{\pi}\] (21)

In other words, the growth rate consistent with the overall balance of payments (or the balance of payments constrained growth rate starting from initial disequilibrium on the current account) is the weighted sum of the growth of exports and real capital flows divided by the income elasticity of demand for imports. If there were no capital flows [i.e. \(\theta = 0\) and \((c - p_d) = 0\)], then equation (21) would collapse to the simple rule in equation (10) that \(y = x/\pi\).

The difference between the actual growth rate and that predicted by equation (21) will be a measure of the pure terms of trade effect on real income growth and of any import volume response from relative price changes relaxing or tightening the balance of payments constraint according to the direction of movement in the terms of trade and whether the import volume response is normal or perverse (i.e. on \(\psi(p_d - p_f - e)\) in equation (20)).

When Hussain and I applied this model to a selection of developing countries over the period of the 1950s to the 1970s we found that for countries which grew faster than predicted by the simple Harrod trade multiplier result the cause was capital inflows, while for countries that grew slower than predicted, the major cause was the (negative) effect of relative price movements. But still the major source of differences in growth performance was the growth of exports. Likewise, when Hussain (1999) fitted this extended model to understand differences in the growth performance of a selection of 29 African countries and 11 Asian countries, the major cause was found to be the export volume effect. Virtually all of the three percentage point growth difference between Asian growth of 6.6 percent per annum and African growth of 3.6 percent is accounted for by the difference in the growth of exports – not differences in the effect of capital flows or terms of trade movements.

**Sustainable Deficits and Debt**
One of the weaknesses of the above model, however, is that it places no limit on the level of current account deficits financed by capital inflows and therefore on a country’s level of indebtedness relative to GDP. McCombie and Thirlwall (1997); Moreno-Brid (1998-99), and Barbosa-Filho (2001) have all addressed this issue. Following Moreno-Brid, rewrite equation (5) as:

\[ P_d X + FP_d = P_t M E \quad (22) \]

where \( F \) is the current deficit in real terms and \( FP_d \) is nominal capital flows (\( C \)) to finance the deficit.

Taking rates of change of equation (22) gives:

\[ \theta (p_d + x) + (1 - \theta) (f + p_d) = m + p_r + e \quad (23) \]

where the notation is as before. Substituting equations (6) and (7) for the growth of exports and imports, respectively, and setting \( f = y \), so that the ratio of the current account deficit to GDP is constant, gives

\[ y_D = \frac{\theta x + (\theta \eta + \eta + 1)(p_d - p_f - e)}{\pi - (1 - \theta)} \quad (24) \]

If the terms of trade are constant, the constrained growth rate consistent with a fixed deficit/GDP ratio is:

\[ y_D^* = \frac{\theta x}{\pi - (1 - \theta)} \quad (25) \]

With no deficit \( \theta = 1 \), and we have the simple rule in equation (10).

The important point to note here is that even large flows of capital to finance current account deficits make little quantitative difference to the predictions of the basic Harrod trade multiplier result. For example, if exports only cover 90 percent of the import bill \( (\theta = 0.9) \) and capital flows cover the rest, with \( x = 10 \) percent and \( \pi = 2 \), the simple rule predicts a growth rate of 5 percent, while the modified model gives a prediction of 4.73 percent. Export growth, not capital flows, is by far the most important variable governing growth performance.

4. McCombie and Thirlwall (1997) derive the same result by a more circuitous route.
Interest Rate Payments on Debt

If current account deficits are financed by debt-creating capital flows, the model above needs further modification to include the interest payments on debt. Elliot and Rhodd (1999); Ferreira and Canuto (2001), Moreno-Brid (2003), Vera (2006) and Alleyne and Francis (2008) have all considered this. The easiest way to proceed, following Moreno-Brid, is to take interest payments out of capital flows, and analyse their impact separately by modifying equation (23). We then have:

\[ \theta (p_d + x) - \theta_i (p_d + i) + (1 - \theta + \theta_i) (p_d + f) = m + p_f + e \quad (26) \]

where \( i \) is the rate of growth of real net interest payments abroad (the negative sign implies that the country is a net debtor), and \( \theta_i \) is the share of foreign exchange devoted to interest payments. Again, substituting for \( x \) and \( m \), and setting \( f = y \), gives:

\[ y^*_I = \frac{\theta x - \theta_i i + (\theta \eta + \eta + 1)(p_d - p_f - e)}{\pi - (1 - \theta + \theta_i)} \quad (27) \]

and if the terms of trade are constant:

\[ y^*_I = \frac{\theta x - \theta_i i}{\pi - (1 - \theta + \theta_i)} \quad (28) \]

If there is no debt and no interest payments on debt, equation (28) reduces to equation (25). Even if the growth of interest payments is quite high, say 5 percent per annum, and the debt service ratio is also high (e.g. \( \theta_i = 0.3 \)), it still makes little difference to the predicted growth rate. For example, if \( x = 10 \) percent and \( \pi = 2 \), then equation (28) predicts a sustainable growth rate of 4.68 percent compared with 5 percent for the simple rule. Again, export growth dominates.

North-South Models

At the core of balance of payments constrained growth models are the income elasticities of demand for exports and imports but testing the model for an individual country says nothing about the
process of economic development between countries. These parameters, however, lie at the heart of what Prebisch (1950) originally called centre-periphery models, and which are now often referred to as North-South models, where the income elasticities of demand for exports and imports refer to groupings of countries – industrialised/non-industrialised; developing/developed; least developed countries/other developing countries etc.

The basic Prebisch model of $\frac{y_p}{y_c} = \frac{\pi_c}{\pi_p}$ (where p is the periphery and c is the centre), however, is too simple for a full understanding of centre-periphery, or North-South, relations because of the simplifying assumptions on which it is based, namely balanced trade and a constant terms of trade. Also it doesn’t say what determines the growth process in the two sets of regions, only why there is a balance of payments constraint related to the intrinsic characteristics of goods. Dutt (2002), Vera (2006) and Sasaki (2008-9) have expanded the basic model to better understand the gap between developed and developing countries. Dutt develops an explicit North-South model using Taylor’s (1983) structural assumptions for the two groups of countries regarding saving, consumption and investment, and the different pricing of primary commodities and manufactured goods. The growth rates of North and South, and the evolution of the terms of trade, are derived simultaneously. Long run growth of the world economy is determined by demand in the North, and in long run equilibrium there is uneven development in the sense that Northern capital and output grow at a faster rate than Southern capital and output because the import elasticity of the North is less than that of the South. Despite the sophistication of the model, the core of the balance of payments constrained growth model is preserved. In a follow-up paper, Dutt (2003) gives some empirics, defining the North as OECD countries and the South as non-OECD countries. He finds that the Southern export elasticity with respect to Northern income growth is 1.02, while the South’s import elasticity is 1.16. Using the World Bank’s grouping of industrial countries and developing countries, the difference in the income elasticities is wider: 1.08 and 1.67, respectively. On the Prebisch rule the developing countries (periphery) would grow at only 65 percent of the growth rate of the developed countries (centre).

But Dutt’s model is a one-good model for the South which doesn’t allow for shifts in comparative advantage. Sasaki (2008-9) has developed a North-South model with a continuum of
goods in the South allowing for changes in the pattern of trade, based on technological progress, building on the original work of Cimoli (1988). While the North is assumed to be fully employed, the South is balance of payments constrained in its growth, and whether there is convergence or divergence with the North depends on the size of the technological parameter. The South may catch up with the North by expanding its comparative advantage sectors sufficiently, or fail to catch up despite expanding its comparative advantage sectors. In the former case, it is shown that promoting price competitiveness leads to a virtuous circle of growth in the South, while in the latter case, such a policy exerts a negative effect on growth; a process of immiserising growth.

Vera’s (2006) North-South model takes into account not only terms of trade changes, like Dutt, but also the role played by net financial transfers because of trade imbalances, and by trade and payments interdependence because if countries (or groups of countries) are big, ‘world’ income growth cannot be treated as exogenous as in the simple model. Vera’s model is complex, but basically output changes and the terms of trade between North and South are jointly determined, and with net financial flows included in the model, three different growth regimes can be identified depending on various parameter values: (i) a mutually reinforcing contractionary growth regime if, for example, there is an autonomous rise in net financial transfers from the South to the North because of a rise in interest rates; (ii) a mutually reinforcing expansionary growth regime if there is debt relief, and (iii) a conflicting growth regime in which an increase in financial transfers from South to North reduces growth in the South but raises growth in the North if the North is balance of payments constrained.

**Testing the Model for Countries and Groups of Countries**

Since 1979 there have been a mass of studies applying the model in its various forms to individual countries and groups of countries. Tables 2 and 3 list the most important ones. McCombie and Thirlwall (1994, 1997) give a survey of studies up to 1996, and McCombie and Thirlwall (2004) reprint eighteen studies with an Introductory survey up to 2003. Since then, new individual country studies have appeared for Slovenia (Beko, 2003); India (Razmi, 2005); Brazil (Bertola, Higachi and Porcile, 2002); Ferreira and Canuto, 2003; Jayme, 2003; Carvalho, Lima and Santos, 2008;
Carvalho and Lima, 2009; Britto and McCombie, 2009; Ireland (Garcimartin, Rivas and Sarralde, 2008); Argentina (Alvarez-Ude and Gomez, 2008); China (Jeon, 2009); Pakistan (Felipe, McCombie and Naqvi, 2010) and Portugal (Antunes and Soukiakis, 2011). New studies of groups of countries include Holland, Vieira and Canuto (2004) for Latin America; Kvedaras (2005) for Eastern Europe; Pacheco-Lopez and Thirlwall (2006) for seventeen Latin America; Bagnai (2010) for twenty-two OECD countries, and Gouvea and Lima (2011) for four Asian and four Latin American countries.

The vast majority of studies support the balance of payments constrained growth hypothesis for two basic reasons. The first is that it is shown overwhelmingly that relative price changes or real exchange rate changes are not an efficient balance of payments adjustment mechanism either because the degree of long-run change is small, or the price elasticities of exports and imports are low. It is income that adjusts to maintain balance of payments equilibrium (or a sustainable deficit). Indeed the most initial direct test of the model is to make income changes and relative price changes endogenous to balance of payments disequilibria and to compare the two alternatives by setting up the equations below: $y = \alpha_1 (x-m, p_d - p_f - e)$ and $(p_d - p_f - e) = \alpha_2 (x-m, p_d - p_f - e)$, and testing for the significance of $\alpha_1$ and $\alpha_2$. Alonso and Garcimartin (1998-99) were the first to do this for ten OECD countries and found $\alpha_1$ significant while $\alpha_2$ is not significantly different from zero. Garcimartin, Rivas and Diaz de Sarralde (2008) also find this to be the case for Ireland over the long period 1960-2000.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Journal</th>
<th>Country or Time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atesoglu</td>
<td><em>Journal of Post Keynesian Economics 1993, 1997</em></td>
<td>USA</td>
</tr>
<tr>
<td>Atesoglu</td>
<td><em>Journal of Post Keynesian Economics, 1993-94</em></td>
<td>Canada</td>
</tr>
<tr>
<td>Atesoglu</td>
<td><em>Applied Economics Letters, 1994</em></td>
<td>Germany</td>
</tr>
<tr>
<td>Bertola, Higachi and Porcile</td>
<td><em>Journal of Post Keynesian Economics, 2002</em></td>
<td>Brazil 1890-1973</td>
</tr>
<tr>
<td>Beko</td>
<td><em>Journal of Post Keynesian Economics, 2003</em></td>
<td>Slovenia 1990s</td>
</tr>
<tr>
<td>Britto and McCombie</td>
<td><em>Journal of Post Keynesian Economics, 2009</em></td>
<td>Brazil 1951-2006</td>
</tr>
<tr>
<td>Carvalho, Lima and Santos</td>
<td><em>Revista Economia 2008</em></td>
<td>Brazil 1948-2004</td>
</tr>
<tr>
<td>Carvalho and Lima</td>
<td><em>Economia and Sociedade 2009</em></td>
<td>Brazil 1930-2004</td>
</tr>
<tr>
<td>Felipe, McCombie and Naqvi</td>
<td><em>Oxford Development Studies, 2010</em></td>
<td>Pakistan</td>
</tr>
<tr>
<td>Ferreira and Canuto</td>
<td><em>Momento Economico, 2003</em></td>
<td>Brazil 1949-99</td>
</tr>
<tr>
<td>Heike</td>
<td><em>Journal of Post Keynesian Economics, 1997</em></td>
<td>USA</td>
</tr>
<tr>
<td>Leon-Ledesma</td>
<td><em>Journal of Post Keynesian Economics, 1999</em></td>
<td>Spain</td>
</tr>
<tr>
<td>Moreno-Brid</td>
<td><em>Banca Nazionale del Lavoro Quaterly Review, 1998</em></td>
<td>Mexico</td>
</tr>
<tr>
<td></td>
<td><em>Journal of Post Keynesian Economics, 1998-99</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>International Review of Applied Economics, 1999</em></td>
<td></td>
</tr>
<tr>
<td>Nell</td>
<td><em>International Review of Applied Economics, 2003</em></td>
<td>South Africa</td>
</tr>
<tr>
<td>Serrano Sanz, Sabote and Gadea</td>
<td><em>Journal of International Trade and Economic Development, 1999</em></td>
<td>Spain, 1940-59 and 1960-80</td>
</tr>
</tbody>
</table>
Table 3
Studies for Groups of Countries

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Journal</th>
<th>Countries and Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alonso and Garcimartin</td>
<td><em>Journal of Post Keynesian Economics</em>, 1998-99</td>
<td>Ten OECD Countries</td>
</tr>
<tr>
<td>Andersen</td>
<td><em>Applied Economics</em>, 1993</td>
<td>Sixteen OECD Countries</td>
</tr>
<tr>
<td>Ansari, Hashamzadeh and Xi</td>
<td><em>Journal of Post Keynesian Economics</em>, 2000</td>
<td>Four Asian Countries</td>
</tr>
<tr>
<td>Bagnai</td>
<td><em>Applied Economics</em>, 2010</td>
<td>22 OECD Countries</td>
</tr>
<tr>
<td>Bairam and Dempster</td>
<td><em>Applied Economics</em>, 1991</td>
<td>Eleven Asian Countries</td>
</tr>
<tr>
<td>Christopoulos and Tsionas</td>
<td><em>International Economic Journal</em>, 2003</td>
<td>Seven Industrial Countries</td>
</tr>
<tr>
<td>Gouvea and Lima</td>
<td><em>Unpublished</em></td>
<td>Panel of 90 countries 1965-99</td>
</tr>
<tr>
<td>Holland, Vieira and Canuto</td>
<td><em>Investigacion Economica</em>, 2004</td>
<td>Ten Latin American Countries</td>
</tr>
<tr>
<td>Hussain</td>
<td><em>African Development Review</em>, 1999</td>
<td>29 African Countries and 11 East Asian Countries</td>
</tr>
<tr>
<td>Landesmann and Poschl</td>
<td><em>Russian and East European Finance and Trade</em>, 1996</td>
<td>Central and East European Countries</td>
</tr>
<tr>
<td>Lopez and Cruz</td>
<td><em>Journal of Post Keynesian Economics</em>, 2000</td>
<td>Four Latin American Countries</td>
</tr>
<tr>
<td>McCombie</td>
<td><em>Journal of Post Keynesian Economics</em>, 1997</td>
<td>USA, Japan and UK</td>
</tr>
<tr>
<td>Moreno-Brid and Perez</td>
<td><em>Journal of Post Keynesian Economics</em>, 1999</td>
<td>Central America, 1950-96</td>
</tr>
<tr>
<td>Pacheco-Lopez and Thirlwall</td>
<td><em>Journal of Post Keynesian Economics</em>, 2006</td>
<td>17 Latin American Countries</td>
</tr>
<tr>
<td>Perraton</td>
<td><em>International Review of Applied Economics</em>, 2003</td>
<td>34 Developing Countries 1973-95</td>
</tr>
</tbody>
</table>
The second reason why the model fits so well is that even if balance of payments
disequilibrium is allowed, capital inflows make no substantial difference to the prediction of the
current account balance model because there is a limit to the current account deficit to GDP ratio that
countries can sustain (see equation (25)).

One interesting feature of the studies is that through time the econometric methods of
estimation have become much more sophisticated; the tests of the model more rigorous, and various
implicit assumptions embodied in the CES export and import demand functions have been relaxed
e.g. the homogeneity and common elasticity of substitution assumptions (Razmi, 2005). But the
most notable shift of all has been towards the use of cointegration techniques to establish long-run
relationships between levels of variables, and the use of the Alonso (1999) technique of taking a long
series of growth rates consistent with balance of payments equilibrium and testing if this series and
actual growth rates are correlated (as discussed earlier). Britto and McCombie (2009) adopt a similar
approach for Brazil using Johansen’s (1988) cointegration procedure which jointly models several
endogenous variables in a VAR framework. First they carry out unit root tests on all the variables;
then they find the lag order of the VAR system and the rank order to determine the number of
cointegrating vectors, and finally estimate the vector and error correction terms. Having estimated the
long-run elasticity of demand for imports, and the hypothetical value of $\pi$ to make actual growth equal
to the balance of payments constrained growth rate, the Alonso test shows the existence of a
significant relationship between actual growth and predicted growth with a constant term close to
zero and a slope coefficient close to unity.

Final Thoughts

Global payments imbalances are bad for the health of the world economy. They give rise to huge,
volatile and speculative capital flows; they contribute to currency instability and the need for
countries to hold large foreign exchange reserves to intervene in currency markets when necessary,
and they lead to an arbitrary reallocation of resources between surplus and deficit countries, often
from poor countries to rich countries. Today, for example, there is something perverse about poor Chinese transferring resources to Americans ten times richer than themselves.

Global imbalances can cause severe difficulties for individual countries, particularly those in deficit, and they exert deflationary bias on the whole world economy. Clearly, not all countries can be balance of payments constrained in their growth performance but it only requires a few countries not to be constrained for all the rest to be so. There is a limit to which deficit countries are willing to finance deficits, and that limit may constrain growth considerably below the rate that would achieve the full employment of resources. That is the surest sign of balance of payments constrained growth; deficits on current account and unemployed domestic resources. Commentators make the obvious point that not all countries can have export-led growth – some countries have to import – but export-led growth from deficit countries is not a zero-sum game, if surplus countries allow their surpluses to diminish. The world as a whole would be better off.

The world economy need not be in this situation of serious global imbalances if it instituted institutional mechanisms to penalise surplus countries that are reluctant, or unable for some reason, to spend more or reduce their surpluses in some other way (I am dubious about the role of currency appreciation). The economy is made for man, not man for the economy! The IMF could declare, for example, if the decision-making bodies agreed, that it will not tolerate members’ surpluses exceeding a certain percentage of GDP – say 2 percent, which is a sustainable level of deficit for most countries. In the old days of the Bretton Woods system, this magnitude of deficit would have put countries on the margin of fundamental balance of payments disequilibrium. Above surpluses of 2 percent of GDP, countries could be fined at progressively higher rates. The proceeds from fines could be given as aid to the poorest countries in deficit. Indeed, Keynes had a similar plan in mind at the Bretton Woods conference in 1944 in his proposals for an International Clearing Union7 which would have been like a world central bank, issuing its own international money (bancor) which countries would have used for payments to each other. Each country would have had a quota with the Union (like countries do now with the IMF which determines borrowing limits). Keynes’s proposal was then that if a country

---

7 Cmd 6437, April 1943. Reprinted in Thirlwall (1987)
had a credit (or debit) balance in excess of one-quarter of its quota, it would pay a charge of one percent of the excess balance, and another one percent if its credit (or debit) exceeded one-half of its quota. He says: ‘these charges would be valuable and important inducements towards keeping a level balance, and a significant indication that the system looks on excessive credit balances with as critical an eye as on excessive debit balances, each one, indeed, the inevitable concomitant of the other’. Keynes’s proposal for an International Clearing Union was rejected by the Americans at Bretton Woods. Keynes used to joke that his proposal for a bank had become a Fund (the IMF), and his proposal for a fund had been named a Bank (the World Bank).

Keynes’s other proposal for a ‘scarce currency’ clause, which would have given the right to deficit countries to discriminate against the import of goods from surplus countries (expected to be the USA), was accepted, but the clause was never implemented because the US soon became a debtor country.

The idea of a scarce currency clause could, however, be resurrected to be used against surplus countries in the way originally envisaged. Both ideas of trade discrimination against surplus countries (notwithstanding the rules of the WTO which has never shown interest in the balance of payments consequences of trade liberalisation) and the penalisation of surplus countries, are ripe for reconsideration for a more stable international economic order and to reduce deflationary bias in the world economy arising from balance of payments constraints on demand and growth in perpetual deficit countries.
References


Serra, A. (1613) *Breve Trattato della Cause che Possano far Abbondare li Regni d’Oro e Argento dove non Sone Minere con Applicazione al Regno di Napoli* (A Brief Treatise on the Causes which can Make Gold and Silver abound in Kingdoms where there are no Mines).


