1 Introduction

Monetary union has been one of the most controversial and paradoxical elements of European integration. It is paradoxical because, despite many predictions that a common currency would never be created and, if it did happen, it was bound to collapse, monetary union has happened and two decades later the common currency still exists.[1] In contrast to the doomsday scenario, the road to monetary union can be presented as a drawn-out odyssey from initial steps to establishment of a single currency over the last three decades of the twentieth century. The reality is more complex than a simple story of overcoming obstacles to achieve the happy ending and many pessimists still believe that the story is a tragedy.[2]

The aim of this Element is to provide an analysis of the road to monetary union with explanations of why the pessimists were confounded by the commitment to establish a common currency and why that currency has survived a tumultuous decade in the 2010s. It will end with a cautious review of the euro's prospects in the 2020s.

The analysis combines economics, politics and history. The underlying theme is that economics is the long-run driver; as the EU moves towards deeper integration and more common policies, the costs of monetary disunion increase. However, the process has not been steady. Politics often determined the timing of the next step (e.g. French pressure on the German Chancellor to think of Europe as Germany reunified in 1990), but it has not determined the direction of change. This view is in opposition to the characterization of the euro by Thomas Mayer (2012, 1): 'a highly ambitious political project pursued with an occasional reckless negligence of economics'. The historical background has influenced timing (e.g. the collapse of the Bretton Woods system and the first oil shock ensured the rapid destruction of the Snake in the 1970s), while the calmer global economic situation for almost a decade after the introduction of the euro in 1999 helped the establishment of the new currency. Some personalities stand out in more traditional accounts of European economic and monetary union; the policymakers and influencers are mentioned in my account, but the impersonal economic forces were stronger and, for example, heads of the Commission who were influential along the road to monetary union (e.g. Roy Jenkins in 1977/8 or

[1] Feldstein (1992) was probably the most influential. Jonung and Drea (2009) provide an exhaustive survey of US economists' predictions.

[2] For example, the publisher's publicity flyer to the updated 2020 paperback edition tells us that Ashoka Mody's book *Euro Tragedy: A Drama in Nine Acts* 'makes clear that the euro's structural flaws will continue to haunt the continent'. The title of the 2016 book by Nobel Laureate Joseph Stiglitz, *The Euro: How a Common Currency Threatens the Future of Europe*, describes the content and conclusion.

Jacques Delors in 1988/9) were pushing the accelerator rather than changing direction.

* * *

The next section reviews three branches of international economics that are relevant to analyzing currency union. Optimum currency area theory has been the most commonly applied theoretical framework. It is useful in highlighting economic costs and benefits of adopting a common currency, but less successful in explaining when a specific currency union is formed or dissolved. A more empirical strand of the literature measures the impact of a common currency on bilateral trade. Finally, the phenomenon of global value chains in which trade occurs at more fragmented levels of specialization has reflected the increased importance of trade facilitation, of which a common currency may be part. The significance of the last two topics is that the benefits of currency union may change over time.

The following five sections focus on different stages along the road to European monetary union. The first step, introduced in the 1970s, was an abject failure which quickly broke down as the large economies withdrew their currencies from the process in a classic example of the Impossible Trinity: a fixed exchange rate, capital mobility and an independent monetary policy cannot co-exist. Governments abandoned the fixed exchange rate system intended as a preliminary to monetary union rather than lose monetary policy independence.

Within a few years, however, European leaders attempted to revive the process, and with more success. This cannot be explained in an optimum currency area framework. Better explanations are that the European leaders recognized the benefits of constant exchange rates in reducing the costs of private and public transactions and were converging on agreed-upon monetary policies that prioritized low inflation. On several occasions, countries had to decide between fixed exchange rates and monetary policy independence, most notably following the 1990 reunification of Germany and the subsequent rise in German interest rates. Section 5 describes how the 1992 currency crisis provided the final push to change fixed exchange rates into a common currency for most of the European Union member countries.

The monetary union was successful in its first decade as inflation remained low in the 2000s and more countries adopted the euro (Section 6). However, low interest rate differentials, as lenders believed without exchange rate risk all eurozone public debt was equally safe, encouraged excessive borrowing by some countries. In 2009, Greece announced that it was unable to maintain its debt-servicing obligations.

The Greek crisis that dominated much of the decade of the 2010s is the most controversial part of the road to monetary union (Section 7). Should Greece have been allowed into the eurozone? How could Greece's mounting debt have been better monitored? Could the conditional financial assistance to Greece have been better managed? Should Greece have left the eurozone and devalued its currency? These questions are difficult to answer because it is hard to specify the counterfactual situation. Many critics argue that the eurozone was introduced too soon, without institutions that may have forestalled the Greek crisis. Many of these institutions were being introduced as the crisis was playing out, which says something about policymaking in crisis periods.

The final section draws conclusions.

2 Theories

A difficulty for evaluating monetary integration has been the absence of a suitable theoretical model. Since the 1960s, optimum currency area (OCA) theory has been a common approach and it remains the foundation for many textbook analyses of European monetary integration. The OCA literature is helpful in highlighting the trade-offs associated with changing currency domains, but it is less useful in explaining the size of actual currency zones or the magnitude of the costs and benefits from currency unification.

A more recent literature has addressed the impact of currency union on trade. While the econometric evidence confirms the presumed positive sign of the relationship, there has been debate over the size of the impact. The very large effect estimated in the pioneering paper by Rose (2000) attracted considerable attention. Rose's results have been questioned, but, whatever the magnitude, this literature redirected debate into a more balanced comparison of the microeconomic benefits and macroeconomic costs of currency union.

The benefits and costs may be changing over time. An increasing share of trade is being conducted along global value chains (GVCs) in which fragmented production is coordinated by a lead firm. This phenomenon is related to increased interest by trade economists and policymakers in the costs of doing international trade; as trade costs fall, increasing specialization by tasks and location of production stages in different countries becomes feasible. There is much more to trade costs than having to exchange currencies, but as the hard and soft infrastructure of international trade is improved, the relative burden of crossing currency boundaries is increasing.

2.1 Optimum Currency Areas

Optimal currency area theory dates from Mundell (1961). The optimum area is where, at the margin, the microeconomic benefits from a larger currency area are equal to the marginal costs of reduced macroeconomic policy effectiveness. The microeconomic benefits from a larger currency area in terms of reduced transactions costs, elimination of currency risk on international trade and foreign investment, greater transparency and ability to compare prices over a larger area are obvious but difficult to measure. The macroeconomic costs of a larger currency area concern lack of flexibility as monetary policy must be one size fits all, which is problematic if members of the currency union face asymmetric shocks.

In practice, the OCA literature has been dominated by macroeconomists arguing about determinants of the effectiveness of macroeconomic policy.[3] Both the costs and benefits of currency union are difficult to measure but 'almost all the interesting stuff comes from looking at factors that might mitigate the costs arising from the loss of monetary flexibility that comes with adopting someone else's currency' (Krugman, 2013, 441). If factors of production are mobile, then asymmetric shocks within the currency area can be met by labour or capital relocating from the depressed area to the booming area (Mundell, 1961). If fiscal transfers are possible within the currency area, then asymmetric shocks can be mitigated by transfers (Kenen, 1969). Both of these arguments may explain why the USA is closer to an optimum currency area than the European Union. For McKinnon (1963), openness is a key determinant of the extent to which an independent monetary policy is possible; small open economies have the least effective exchange rate policies, because the changes in relative prices following a devaluation will quickly pass through the domestic economy.[4] Each of these arguments is plausible but taken together they are hard to operationalize to explain existing or future currency areas beyond the obvious unions involving microstates.[5]

[3] Krugman (1993) was a prominent dissident who emphasized the potentially large microeconomic benefits.

[4] Kenen (1969) lengthened the list of criteria which might be relevant. For Alesina and Barro (2002), the trade-off might be mediated by history and by geography, but otherwise their criteria are similar to those in the survey by Tower and Willett (1976) or in the textbook treatment of de Grauwe (2020).

[5] McKinnon's openness argument explains currency unions involving microstates like Andorra, Monaco or San Marino, and small countries like Lichtenstein (that uses the Swiss franc), Timor-Leste (that uses the US dollar) or Brunei Darussalam (that uses the Singapore dollar), which would have little macroeconomic policy independence with a national currency. An independent currency would also impose large microeconomic costs as international trade would involve significant foreign exchange transactions costs for microstates' currencies.

More challenging than evaluating the costs and benefits of currency union and the arguments over macroeconomic policy effectiveness has been the limited testing of whether OCA theory explains actual currency domains (Pomfret, 2005). Kreinin and Heller (1974) synthesized the various OCA criteria into the single question of whether a country could better deal with external imbalance through devaluation or through adjustment of domestic demand. Their conclusion was that Italy, Sweden and Switzerland were the three OECD countries most likely to abandon their national currencies. Forty-five years later, only one of the three has done so, while ten of the 'less likely' countries have abandoned their national currencies. There has been little else in the way of serious testing of OCA theory.

Since the 1950s, despite the increasing openness of national economies and increasing capital mobility, which are both unambiguous pressures for larger currency areas according to OCA theory, the number of currencies has increased substantially and the geographical size of currency domains has shrunk correspondingly. Clearly, the exogenous increase in the number of countries drove the number and size of currency areas, and the OCA criteria were of little relevance for explaining the pattern.

Frankel and Rose (1998) have argued for a two-way relationship between trade intensity and a common currency, so that there is endogeneity in currency area formation or bilateral trade. In the data used in this and other papers by Rose (e.g. the influential Rose (2000) paper), the countries in currency unions are not from a random draw. Several authors (Persson, 2001; Kenen, 2002; Nitsch, 2004) have shown that the currency union members are smaller and more open than their natural comparators, and that history (usually in the form of colonial background) matters. Glick and Rose (2002) identify temporal correlations between changing currency union status and bilateral trade flows, but usually currency union break-up is associated with other events which disrupt trade; Nitsch (2003) showed that, out of some sixty cases of post-1947 currency union dissolutions in the Glick-Rose dataset, over two-thirds broke up within a decade of the end of a colonial relationship.[6] Thus, even the idea that OCA criteria can become self-fulfilling *ex post* due to feedbacks is unconvincing.

The overwhelming global pattern is one country one currency; that is, currency areas are determined by national boundaries. A few small economies

[6] In tranquil currency union changes, notably Ireland's secession from its currency union with the UK in 1979 and subsequent participation in the process leading to the euro, the impact on bilateral trade is unclear. Thom and Walsh (2002) found that breaking the currency union did not have an adverse impact on Ireland–UK trade, while Fitzsimmons et al. (1999) found that trade between Ulster and Ireland after 1979 was greater than predicted by a standard gravity model, despite the absence of a common currency.

use another country's currency (e.g. Brunei, Timor-Leste, Lichtenstein) and virtually no country has multiple currencies that are all accepted as legal tender. Ministates use a foreign currency because the transactions costs of a national currency would be too high and because they have very limited macroeconomic policy independence even if they had their own national currency. Otherwise, countries want control over their monetary policy agenda.

The benefits of a common currency are not only that it reduces transactions and search costs in the private sector. The benefits also apply to the public sector. Nations do not tolerate multiple currencies because they would make public revenue and expenditure decisions difficult. Once the national budget has been negotiated, the political balance would be upset if each province had its own currency that could change in value.

The arguments about control over the monetary policy agenda and about the content of fiscal policy differ from the emphasis of OCA theory on a trade-off between the microeconomic transactions costs benefits of a wider currency area and the macroeconomic cost of losing control over macroeconomic policy instruments. The first set of policy arguments have become increasingly relevant to the European Union as the member states' views on monetary policy converged in the 1980s and as the EU moves towards closer union with a more complex union budget.

2.2 A Common Currency and Bilateral Trade

Andrew Rose (2000) used a gravity model to compare the trade between countries with a shared currency and trade between countries without a shared currency. The gravity model, which hypothesizes that trade between two economic units depends on their economic size and the distance between them, had been developed by Dutch economic planners in the 1950s. In its simplest form, the gravity model is:

$$T_{i,j} = f(Y_i, Y_j, D_{i,j})$$

where the subscripts i and j refer to a pair of countries, $T_{i,j}$ is the bilateral trade between i and j, Y_i and Y_j represent the economic size of the two countries, and $D_{i,j}$ is the distance between them. The planners liked the model because it forecast future trade flows well, but other economists showed little interest because the results seemed obvious.

Modern use of the gravity model dates from the 1990s and especially the article by McCallum (1995) who analyzed bilateral trade flows among US states and Canadian provinces. McCallum found that the simple gravity model explained trade between two US states or between two Canadian

provinces well, but trade between a state and a province was far smaller than predicted. Including a dummy variable for trade that involved crossing the border improved the statistical results and illustrated the economic importance of the US–Canada border even though it was one of the world's most open.

McCallum's paper was followed by many other papers using the gravity model to find deviations from the simple specification – when the force of gravity is 'unconstant' (Baldwin and Taglioni, 2006). Rose (2000) was one of the first of these papers. He included a dummy variable that took the value of 1 if two countries shared a common currency and 0 if they did not. The dummy's coefficient was statistically significant and implied that sharing a common currency, on average, tripled bilateral trade between a country-pair.

Rose's paper can be criticized for its econometrics and for its sample selection. Since 2000, gravity modelling has become more sophisticated (Anderson, 2011; Head and Mayer, 2015). Country fixed effects account for individual country factors affecting bilateral trade (e.g. North Korea trades less with other countries than might be suggested by global patterns of the impact of size and distance on bilateral trade flows [Anderson and van Wincoop, 2003]), and a range of estimating methods address the problem of zero observations (e.g. the Pitcairn Islands do not trade with Montserrat and, if the trade data are at a disaggregated level, individual commodities will not be traded between many country pairs). Rose (2000, 11 n. and 41) lists eighty-two countries and territories which used another country's currency or were in a currency union between 1970 and 1990; they are all small, with the most populous being the African CFA countries and Panama and Liberia, and many are quasi-countries (e.g. Isle of Man or Svalbard). On the whole, however, the finding that a common currency has a positive influence on bilateral trade flows appears to be robust, albeit with disagreement over the magnitude of the effect.[7]

In a post-euro assessment, Glick and Rose (2016) estimated that European monetary union boosted participants' bilateral trade by around 50 per cent. Glick (2017) separated the monetary union effect from the trade effect of EU membership and found that the independent effect of monetary union on trade of older EU members was about 40 per cent, and lower for more recent members. In sum, the euro appears to have had a large impact on bilateral trade, although it is difficult to translate this impact into the effect on GDP or on economic welfare more generally.

[7] Nitsch (2002) and Rose (2002) discuss the extent of overestimation in Rose's original results. Frankel (2010) found a 15 per cent increase in trade over seven years of euro use (1999 to 2006), which is far less than Rose's tripling but still substantial.

2.3 The GVC Phenomenon

When the Treaty of Rome was signed in 1957, most international trade consisted of goods that were often manually loaded and unloaded. By the time the euro came into existence, trade logistics had been transformed by the standardized container, door-to-door real-time tracking and other technical change. International production had also been transformed as many goods were manufactured along global value chains (GVCs) in which tasks were carried out in different countries and coordinated by a lead firm. GVCs depended on low transport and border-crossing costs and on reliable just-in-time delivery of components, to avoid the need for costly inventory holdings.

The emergence of the GVCs as a significant feature of world trade is dated by Baldwin (2016) to the 1980s.[8] One feature has been that most of the chains are regional rather than global, and are concentrated in three regions: East Asia, North America and Europe. A catalyst for the European GVCs has been enlargement of the EU to include countries with lower wage rates, that is, the Mediterranean enlargements of the 1980s (Greece, Portugal and Spain)[9] and, especially, the Eastern European enlargements of 2004 and 2007.

Among the Eastern European countries, the Czech Republic, Hungary, Poland and Slovakia participate the most in GVCs (Pomfret and Sourdin, 2018). This reflects many causal factors such as industrial traditions, the relatively high skill levels of the low-wage workers, good infrastructure and connectivity to lead firms in Germany, France, etc. Slovakia stands out and it is the only one of the four countries to be using the euro; Slovakia has the highest per capita car output of any country in the world, and the common currency must facilitate accounting of the components entering the country from elsewhere in the eurozone and of the value of the cars being assembled in Slovakia.

The GVC phenomenon is part of the expansion of international trade into increasingly specialized areas of comparative advantage where margins on each task are competitively squeezed and low costs of international trade are essential for profitability. As in the previous section, a common currency is one element of reduced trade costs. This helps to put some detail into the microeconomic benefits side of the OCA calculus and also, with the reduction of other

[8] On the emergence and characteristics of GVCs, see also Johnson and Noguera (2012; 2017) and UNIDO (2018).

[9] An early high-profile example was the Ford Motors decision in the 1970s, in anticipation of Spanish accession, to build a greenfield factory in Spain to assemble the Ford Fiesta from components sourced from across the European customs union. In the 1970s and 1980s, Malta, with preferential access to European Economic Community markets, became an assembly point for Wrangler jeans, with a quarter of the island's workforce employed in sewing jeans from imported inputs and exporting the jeans to the marketing center in Belgium which was the next stop in the value chain (Grech, 1978).

elements of trade costs and the rise of GVCs, provides a reason why a larger currency domain may be more popular in the twenty-first century than earlier.

3 The Werner Report and the Snake

The Treaty of Rome in 1957 paid little attention to monetary integration. The monetary chaos that followed the peace in 1945 was being resolved and the Bretton Woods system of trade with fixed exchange rates was about to come into full operation as exchange controls were loosened in western Europe. However, the system came under increasing stress during the 1960s as countries adjusted their exchange rates, notably the devaluation of the UK pound in 1967 and the devaluation of the French franc by 12.5 per cent and revaluation of the German mark by 9.3 per cent in 1969. The latter, involving the two largest members of the European Economic Community (EEC), triggered serious thought about economic *and monetary* integration.

3.1 The Werner Report

In post-1945 Europe, foreign exchange controls placed strict limits on access to and use of currencies, which was a serious impediment to trade. The European Payments Union was established in 1950 to facilitate trade by acting as a clearing house to settle trade imbalances among European countries, while currency restrictions were maintained for trade in US dollars. The system gradually became more flexible and the Payments Union terminated in 1958 when the currencies became convertible at fixed exchange rates. The Bretton Woods system based on fixed exchange rates was expected to last indefinitely. However, in 1969, a large devaluation of the French franc and revaluation of the German mark undermined expectations of exchange rate stability and raised questions of how well the customs union could function if exchange rates were not fixed.

At the December 1969, Hague Summit, the EC Council created an ad hoc committee of experts under the chairmanship of Pierre Werner to explore possibilities of stage-by-stage progress towards economic and monetary union. The work of the committee was presented in October 1970 in the Werner Report. The Report starts by setting out the final objective of economic and monetary union, which implies:

> Inside its boundaries the total and irreversible convertibility of currencies, the elimination of margins of fluctuation in exchange rates, the irrevocable fixing of parity rates and the complete liberation of national monetary symbols or the establishment of a sole Community currency. From the technical point of view the choice between these two solutions may seem immaterial, but

> considerations of a psychological and political nature militate in favour of the adoption of a sole currency which would confirm the irreversibility of the venture. (Werner Report, 1970, 10)

The Report recommended a three-stage process to achieve monetary union by the end of the decade. During the first stage, commencing on 1 January 1971 and lasting for three years, fiscal and monetary policies would be coordinated and exchange rate fluctuations would be limited to a narrower range than those permitted in the Bretton Woods system. The second stage would involve financial market integration, removal of restrictions on capital flows between members, and short-term economic and fiscal policy coordination. The final stage would see the irrevocable setting of exchange rates, economic policy convergence and a community-level system of central banks. The expectation was that the three stages would be completed by 1980.

The Report's recommendations were adopted and the first stage began in April 1972. The delay was due to the 'Nixon Shock' in August 1971 when the convertibility of the US dollar into gold was suspended until a new set of exchange rates was agreed at the Smithsonian on 18 December 1971. The Smithsonian Arrangement, lauded by President Richard Nixon as 'the most significant monetary agreement in the history of the world', only lasted until February/March 1973, when the Bretton Woods system based on fixed exchange rates to the dollar broke down. The original participants in the first stage were the six original EEC members: Belgium, France, Germany, Italy, Luxembourg and the Netherlands. Denmark, Ireland, Norway and the United Kingdom joined shortly afterwards.[10]

3.2 The Snake in the Tunnel

Implementation of the Werner Report's recommendations took place against the background of the 1971–3 collapse of the Bretton Woods fixed exchange rate system. Under the December 1971 Smithsonian Agreement, exchange rates were fixed against the US dollar with a plus-or-minus 2.25 per cent margin of fluctuation. The European countries agreed that their bilateral exchange rates could only fluctuate by +/–2.25 per cent against each other. This meant that although the strongest and weakest European currency could in principle have a margin of 4.5 per cent if one were at the top of the US dollar range and the other at the bottom of the US dollar range, their European commitment would constrain them to a narrower band. The movement of the European currencies was conceived as a Snake within a tunnel; the wider tunnel represented the potential range of EEC

[10] Denmark, Ireland and the UK joined the EEC on 1 January 1973. Following rejection of membership in a referendum, Norway decided to remain outside the EEC.