Endogeneity of Currency Areas and Trade Blocs:

Evidence from a Natural Experiment

I. Introduction

What are the effects of political arrangements on international trade? Empirical research on the gravity model of trade has often found that common institutions increase trade considerably over the predictions from geography and economic size. Politically unified areas trade disproportionately more with each other than with similar regions across political borders (McCallum (1995), Helliwell (1998)). Trade within customs unions is much larger than a gravity model would predict. Famously, the same has been found by Rose (2000, 2001) and Rose and Glick (2002) for currency unions.

The common question behind such regularities is whether they carry policy implications. What would be the effect of a political unification between the U.S. and Canada on trade? Would customs unions like a proposed North Atlantic Free Trade Area increase trade by factor 2 or 3, building on results, e.g., by Frankel and Rose (1998)? And would currency unions, such as the Euro zone, bring about a similar increase in trade?

After initial enthusiasm, the literature on the gravity model has grown increasingly skeptical of such conclusions. On the one hand, a series of papers reestimating the original results produced smaller coefficients, applying slight variations to the methodology and the data. "Honey, I just shrunk the currency union effect on trade" (Nitsch (2002), commenting on Rose (2000)) is emblematic of this strand of research[1]. On the other hand, policy experiments like NAFTA and the Euro appeared to generate far lower trade effects than predicted. While there is a consensus that both had positive effects on trade, there seems to be general disillusionment about the magnitude of these effects relative to the earlier estimates (on the Euro see e.g. Micco et al., 2003, Berger and Nitsch, 2005, as well as Baldwin, 2006; other recent studies on the effects of economic and political institutions on trade include Maurel and Cheikbossian, 1998, and de Groot et al., 2004).

This raises the issue of what may have gone wrong in the earlier estimates and the large coefficients they created. The methodological starting point is the theory of optimal

---

[1] See also Rose's reply, "Honey, the currency union effect on trade hasn't blown up", Rose (2002).

currency areas (OCA), as introduced by Mundell (1961) and McKinnon (1963) and further developed by Frankel and Rose (1997; Frankel and Rose (1998), Dixit (2000), and Alesina and Barro (2002). OCA theory argues that trade integration may give rise to monetary integration, as the degree of economic integration between two countries affects the possible welfare gains of a monetary arrangement. This may cause endogeneity bias: national monetary policy may have been abandoned and a currency arrangement formed precisely *because* trade was already high.

In a study on the Euro, Berger and Nitsch (2005) examine the post-war trends of trade among the members of the later Eurozone, and find no trade effects of introducing the Euro after controlling for a trend in trade integration. While the economic underpinnings for such a trend are not clearly spelled out, this finding is in line with endogenous monetary integration along pre-existing trade patterns. In a recent survey, Baldwin (2006) also stresses the issue of endogeneity ("reverse causality") and calls for more and better data. From an econometric viewpoint, the challenge is to identify the effects of discretionary policy in the possible presence of endogeneity and anticipation effects. A currency or trade bloc may be formed as an endogenous response to trade integration. Alternatively, trade might already increase before the creation of such a bloc because agents anticipate its effect. In the parlance of labor econometrics, the problem is to identify the treatment effects of forming the political arrangement – whether it is a customs union, a currency area, or a change in political borders – on bilateral trade. Modeling such treatment effects of policy requires spelling out counterfactuals and thus implies strong identifying assumptions.

In this paper, we examine the possible endogeneity bias of political arrangements in the context of the gravity equation using identification from a natural experiment. We draw on the treatment effect methodology of labor econometrics, which aims to evaluate the effects of, e.g. participation in labor market training programs on employment outcomes and wages. We find pervasive evidence of endogeneity and selectivity in these arrangements, and show that standard methods in the gravity literature would fail to detect this. Along the time dimension, we examine the most commonly used panel estimators for the gravity model, and find that they fail to detect endogeneity. This is partly due to the particular structure of the gravity model, whose time-invariant structural relationships would be collinear with fixed effects in the units of observation. We argue that the estimators which have been proposed to circumvent this collinearity problem do not identify treatment effects whenever random selection is violated. As a consequence, the observed coefficients on policy dummy variables do not provide unbiased estimators of the effects of policy.

One way out of the impasse that we explore is difference in differences (DD) estimation. This estimator is easy to implement in the gravity equation but is only rarely used. We show that DD estimation crucially introduces a group fixed effect or fixed effect on the treated (henceforth, FET) for the treated group. We also show that country pair fixed effects (CPFE), which are also often used in the literature do not implement the DD estimator. While the latter produces similar (sometimes identical) point estimates for coefficients, it gives rise to a serious problem of overfitting. However, even DD estimation of policy treatments is still seriously restrictive: conditional on the fixed and time effects, the treatment or policy arrangement itself is considered random. Self-selection that also endogenizes the policy treatment itself is not captured by any of the specifications unless anticipation effects and treatment-specific time trends can be controlled for, see Besley and Case (2000). To model self-selection, labor econometrics has developed various matching approaches, see Heckman et al. (1999). Use of this approach to detecting endogeneity and selection problems is gradually emerging in the literature on the gravity model, see Persson (2001), Baier and Bergstrand (2007).

Our approach to self-selection in the gravity model goes further, and explicitly models the endogeneity of the policy treatment itself. To this end, we build on research by Barro and Tenreyro (2003), who suggest analyzing the formation of currency areas around anchor countries. We implement this concept in the context of our natural experiment and specify a matching process that groups countries by their propensity join one or the other trade and currency arrangement prior to the Great Depression. This new analytical tool uses the gravity model to analyze the formation of political arrangements and allows constructing balanced covariates for the analysis of treatment effects when such arrangements arise endogenously.

We implement this approach with data from a historical episode in which new currency areas formed in the aftermath of a cataclysmic, unexpected event. This event is the Great Depression and the subsequent collapse of the Gold Standard. In the aftermath of the Great Depression, the Gold Standard was replaced by several regional currency blocs (see Wolf 2008). There exists a solid body of literature which argues that the Great Depression was indeed unexpected. Hamilton (1987; Hamilton (1992) has argued that deflation after 1929 came unexpectedly. Dominguez, Fair and Shapiro (1988) examined the performance of contemporary business-cycle forecasts through leading indicators, and found that they fail to predict the recession. While the Gold Standard itself was no longer the unquestionable commitment device it had been before 1914, bond spreads among its European members were small in 1928, see Obstfeld and Taylor (2003). All these results imply that an imminent breakdown of the Gold Standard was not expected before 1929 and that anticipation of future currency blocs did not influence agents' behavior.

Moreover, none of the currency blocs in our dataset came anywhere near to being a currency union. The blocs that formed in the 1930s were fixed-exchange rate systems, some of them with tight capital controls, quotas and other non-tariff trade barriers. Conceived as a stopgap in the financial chaos following the Great Depression, these arrangements were perceived as transitory, and the costs of exiting were probably small (Eichengreen 2007). For all these reasons, no sizeable trade effects from these arrangements should be expected a priori. In other words, most of the arrangements we study should be mere placebos. If anything, the historical literature has emphasized the protectionist nature of these blocs, and attributed large trade diverting effects to them, see Ellis (1941), Child (1957). Nevertheless, in standard estimates of a gravity model of trade, these historical currency and trade blocs exhibit very high coefficients (see Eichengreen and Irwin (1995)), similar to the ones that are known for currency unions from the work of Rose and others.

Our choice of this particular historical episode is motivated by the favorable characteristics of the data[2]. First, due to the surprising breakdown of the Gold Standard into various trade and currency blocs, the trade data are unlikely to be contaminated by anticipation effects. Second, "pre-treatment" observations are available for the whole sample: the researcher can actually observe trade flows for the Gold Standard years, prior to the beginning of the Great Depression, and prior also to the establishment of the currency and trade blocs. These conditions are rarely given for the commonly studied datasets. In the wide IMF dataset used in the work on currency unions by Rose (2000; Rose (2001), and Glick and Rose (2002), pre-treatment observations are extremely sparse, as most of the currency unions in that dataset were inherited from colonial times. For these former colonies, the observations in this dataset are actually on breakups of currency

[2] More recent research on the pre-1914 gold standard by Flandreau and Maurel (2001), Estevadeordal, Frantz and Taylor (2003), and López-Córdova and Meissner (2003) has confirmed high trade creation coefficients of historical monetary arrangements in standard gravity models of trade. Estevadeordal, Frantz and Taylor (2003) also look into the effects of the collapse of the gold standard on trade, and again find high coefficients. The innovation of the present paper over existing historical research is to search for the treatment effects of such currency arrangements rigorously. And we will show that it is very small.

unions, not on their formation. Fixed effects estimation of the trade effects of currency unions then implicitly assumes that the trade effects of currency union formation and breakups are symmetrical. The postwar European data studied, among many others, by, Eichengreen and Irwin (1997), Berger and Nitsch (2005) Flam and Nordstrom (2003), Micco, Stein and Ordoñez (2003), Mongelli, Dorrucci and Agur (2007), do have pre-treatment observations but are plagued by a double contamination problem. One is trend growth in trade integration, which itself is a likely consequence of the recovery from interwar trade disintegration and World War II. The other is a possible anticipation effect, as every step in monetary integration was announced well in advance. In contrast, the interwar dataset allows us to identify treatment effects along the time axis, as we have pre-treatment observations and as anticipation effects are likely to be absent. Together, these characteristics make this data set a near-ideal natural experiment for studying endogeneity in a panel estimate of the gravity equation.

The remainder of this paper is structured as follows. Section II briefly introduces the data for our natural experiment. Section III presents panel estimates of the gravity model and discusses the identification of treatment effects along the time axis. The issue of self-selection is dealt with in Section IV, where we use matching techniques and binary choice approaches to model the endogeneity of policy treatments. Section V concludes with suggestions for further research.

II. Data and Institutional Setup

Our estimates combine two different data sets. Export and import data were collected by the League of Nations for the benchmark years of 1928, 1935, and 1938 as published in Hilgerdt (1942) and converted to 1936 dollars. In addition to gravity controls like distance, languages, common borders etc., in some estimates we employ GNP data at purchasing power parities for the interwar years from Prados (2000) for 30 countries, including the United States, Canada, and most of Europe. We converted output data for the Soviet Union from Maddison (1995) to fit in with the Prados data set. We also include an index of exchange rate variability as in Eichengreen and Irwin (1995). Per year, this provides 435 country pairs, or 870 observations, recording exports and imports separately. Pooled over the three years, the data set includes 2610 observations.

As stated above, the interwar gold standard was almost universal before the onset of the Great Depression, see Temin (1989), Eichengreen (1992). Among the countries in our sample, only the Soviet Union and Japan were not on gold in 1928. After the Great Depression, all countries in our sample reneged on the gold standard, except for five countries led by France that carried on to 1935/6. As a substitute, regional currency and trade blocs formed. The currency agreements and trade blocs we look at were as follows:

(i) *Gold bloc*: five countries that remained on the gold standard to 1936, namely France, the Netherlands, Belgium (to 1935), Switzerland, and Poland

(ii) *Sterling bloc*: nine countries that left the gold standard in 1931/2 and tied their currencies to the British pound, namely Great Britain, Ireland, Norway, Denmark, Sweden, Finland, Portugal, Australia, and New Zealand.

(iii) *Commonwealth trade bloc*: five countries that formed a protectionist tariff area in the Ottawa preferences of 1932, namely Great Britain, Ireland, Canada, Australia, and New Zealand.

(iv) *Reichsmark bloc*: seven countries formerly on the gold standard that had currency pegs to the Reichsmark around 1937/38, namely Germany, Czechoslovakia, Austria, Hungary, Romania, Bulgaria, and Greece.

(v) *Foreign exchange control bloc*: thirteen countries that maintained foreign exchange agreements with multiple exchange rates with each other and with Germany, namely Austria, Hungary, Romania, Yugoslavia, Turkey, Italy, Spain, Germany, the Netherlands, Denmark, Norway, Sweden, and Finland (see Ellis (1941).

Evidently, there existed major overlaps in bloc membership. This implies that the presence of other, competing arrangements need to be accounted for in order to evaluate the effects of any single bloc properly.

III. Trade and Currency Blocs: Panel Estimates

III.1.Identificaton of Treatment Effects along the Time Axis

This section estimates a standard benchmark gravity equation (1) under various identifying assumptions about the treatment effects of currency areas and trade blocs. Panel estimates of the standard gravity equation are commonly a special case of the following specification:

$$tr_{ij,t} \quad = \quad X_{ij,t}\,\beta + D_{ij,t}\,\delta + Z_{ij}\,\gamma + \chi_i + \chi_j + TIME_t \cdot \eta_t + u_{ij,t} \qquad (1)$$

Here, tr_{it} denotes a measure of exports from country i to j, X_{ij} represents suitably chosen fixed effects and/or a constant, while D is a 0/1 indicator variable indicating the policy treatment. Finally, Z_{ij} stands for the time-invariant country-pair specific gravity controls, such as distance, common language, common borders etc. To this add country-specific controls χ_i for any of the exporting countries i, and χ_j for the importing countries j. Possible time fixed effects, *TIME* may be present, and may include interaction terms with any of the terms in X_{ij}, Z_{ij}, and χ_i, χ_j. Stacking all $n \cdot (n-1)$ country pairs ij and T observation periods, the model becomes:

$$tr = X\beta + \Delta\delta + Z\gamma + (\chi_i + \chi_j + \eta) \cdot TIME + u \qquad (1')$$

where tr is an $n \cdot (n-1) \cdot T \times 1$ vector and where Δ is a vector of the same dimension with $\Delta = D$ if $TIME = t$, and 0 otherwise.

Identification and estimation of policy treatment effects in panel data specifications of the gravity equation is straightforward, although mis-specifications are frequent in the literature. A convenient approach to identify within-variation is to specify group-specific fixed effects and apply difference-in-differences estimation. Define the fixed effect on the treated (FET) as:

$$FET_{ij} = \begin{cases} 1, & ij \in TR, \quad \text{all } t,s \\ 0, & ij \in C, \quad \text{all } t,s \end{cases}$$

where the sample of $n \cdot (n-1)$ country pairs is split into a treatment group $\{1,\ldots,g\} \equiv TR$ and the control group $\{g+1,\ldots n \cdot (n-1)\} \equiv C$.[3]

However, much empirical work in the gravity literature instead assumes individual country-pair fixed effects (CPFE). CPFE assigns an individual regression constant to each country pair, thus studying variation at the country pair level, not the treatment group level.

Absent other regressors, the coefficients obtained under CPFE and FET are identical. However, two differences do arise. One is that the matrix X_{CPFE} of individual country-pair fixed effects is perfectly collinear with the time invariant, country-pair specific characteristics Z of the gravity equation (such as distance, common border,

[3] Then, the regression constant in equation (1) becomes a vector where $X_{ij,FET}$=[1 FET_{ij}].

language etc.).[4] Hence, to study the effects of policy treatments in a gravity model, group-specific FET that allow for separate identification of Z are to be preferred.

A second and more important consequence of specifying observation-specific CPFE is overfitting. Overfitting in treatment effect regressions through redundant fixed effects causes downward bias in the standard error of the estimated treatment effect. This bias is roughly proportional to the reduction in the residual sum of squares introduced by overfitting.[5] In large panels the spurious efficiency of CPFE over the FET estimator can be serious, leading to false positives on the treatment effect coefficient. Further below, we will discuss the effects of overfitting in the context of our data.

We estimate the gravity equation with best practice techniques, employing time-varying exporter and importer effects as suggested by Anderson and van Wincoop (2003) to capture multilateral trade resistance (MTR). All estimates also include time fixed effects. In the above notation, the gravity model then takes the form:

$$tr \;=\; X\beta + \Delta\delta + Z\gamma + (\chi_i + \chi_j + \eta_t)\cdot TIME + u \qquad (1')$$

On the RHS we vary the specification by assuming different specifications of *X*. These are, in turn, a pooled estimation (PE) specification with a common regression constant $X_{PE} = (1,\ldots,1)'$, the difference in differences (DD) specification with X_{FET}, and the country pair fixed effect (CPFE) specification with X_{CPFE} , as introduced above.

We employ three different estimation methods. Santos Silva and Tenreyro (2006) have shown that under heteroskedastic errors, the standard log specification of the gravity equation is seriously biased and inefficient. As an alternative, they suggested estimating the gravity model in its multiplicative form by Poisson Pseudo Maximum Likelihood (PPML). This also circumvents the problem of dealing with zero observations of the dependent variable. For comparison, we also include the more traditional Scaled OLS (SCOLS) and Tobit estimates. Each of these estimation methods affects the specification of the LHS variable. On the LHS of the gravity equation we thus have, in turn, the PPML estimates with $tr_{PPML} = tr_{ij}$ in levels of trade, the SCOLS specification with $tr_{SCOLS} = \ln(1+tr_{ij})$ and the Tobit estimation with $tr_{TOBIT} = \ln(tr_{ij})$. The variation in results across methods is

---

4 In matrix notation, X_{CPFE}=$[I_1 \ldots I_T]'$ where each identity matrix has dimension $[n(n-1)\times n(n-1)]$.

5 If no other regressors are present, or are orthogonal to the fixed effects and treatment dummies, this relation holds exactly, see Ritschl (2009).

considerable and broadly confirms the conclusions of Santos Silva and Tenreyro (2006). However, these issues are unrelated to the problem of identifying treatment effects on the RHS of the gravity equation, which is what we are primarily interested in.

III.2. Naïve Estimates: Results from Pooled Regressions

Estimates of the gravity equation in Table I for 1928, 1935, and 1938 implement the pooled estimator. The independent variables include the log of distance, dummy variables for the presence of currency and trade arrangements in the 1930s, and the usual additional controls for possible source of pair-specific heterogeneity used in the gravity equation (common border and common language, not shown here; we tried out many more, without major changes in results). In the spirit of Eichengreen and Irwin (1995), we also include an index of exchange rate variability.

(Table I about here)

The estimates in Table I differ in the specification of the endogenous variable and the estimation method. The baseline overall change in trade associated with the 1930s is captured by a time fixed effect. With a coefficient around -.55, this time dummy indicates a trade decline of over 40% compared to 1928. As all countries in our sample except for the USSR and Japan were on gold in 1928, this effect can loosely be interpreted as the overall trade destruction effect of dissolving the Gold Standard, although this is necessarily a bit speculative. Assuming symmetry between trade creation by establishing the Gold Standard and trade destruction by dissolving it, this would imply a Gold Standard trade creation effect in excess of 80%.

The estimates disagree quite substantially on the size of the distance term, with the PPML estimate coming in on the low side, as in Tenreyro and Silva (2006). Exchange rate volatility impacts negatively on trade, which is hardly surprising in the context of the interwar years. The estimates also broadly agree on the trade effects of the currency and trade blocs of the 1930s. With coefficients roughly between .4 and 1.5, the estimates from pooled regression assign trade creation effects of 50% to over 300% in levels to the various trade and currency blocs. The exception is the gold bloc of those countries that stayed on gold to 1935/36. Here, the PPML and Scaled OLS estimates, which exploit the full dataset including the zero observations, find a significant effect, while the pooled and Tobit estimators, informed only by nonzero trade flows, reject such an effect.

The high coefficients on the currency and trade bloc dummies in Table I broadly replicate the findings of Eichengreen and Irwin (1995). We notice that these results are remarkably close to the coefficients reported for currency unions by Rose (2000) or Glick and Rose (2002). This looks suspicious, as none of the blocs forming in the 1930s was even remotely comparable to a currency union. The high trade creation effects for these blocs reported by Eichengreen and Irwin (1995) for 1928, five years before these blocs were formed, adds to the doubts. Hence it seems plausible that the high coefficients in Table I are spurious, and just an artifact created by misspecification of the treatment effect.

As outlined in Section II above, the country-specific MTR present in the equations of Table I are technically not fixed effects that reveal within-group variation in the endogenous variable. The regressions therefore implicitly assume that trade prior to currency area formation was normally distributed around the regression line, and that selection into the later currency blocs was purely random.

III.3. Fixed Effects on the Treated: Difference-in-Differences Estimation

In our panel, the existence of pre-treatment observations for 1928, which are arguably untainted by anticipation effects, provides a clean way to test whether high trade creation effects are spurious. In Table II we include group-specific Fixed Effects on the Treated (FET) alongside each treatment dummy for bloc membership. The coefficient on the FET then measures pre-existing trade integration within each later group in 1928. We again include a time fixed effect, which suggests itself from the downward overall trend in international trade in the 1930s. The group specific FET and the time fixed effect together implement the difference-in-differences (DD) estimator. In this framework, the formal introduction of the currency arrangement is the treatment, and the coefficient on the arrangement captures the Average Treatment Effect on the Treated (ATT).

(Table II about here)

Table II shows two clear results. First, the FET pertaining to each group are large and significant. The second, equally clear-cut effect is that the policy treatment dummies are mostly very small and insignificant. This implies that upon actual formation of the respective trade and currency blocs, the degree of trade integration changed very little. By far the dominant effect is previous regional trade integration, as evidenced by the large fixed effects for 1928.

The path dependency in currency area formation revealed by Table II is quantitatively impressive. Among members of the later Reichsmark zone, trade in 1928 was already between 1.9 and 4.6 times higher than the gravity model would predict. For the Sterling zone, this effect varies between 1.7 and 2.7. And intra-Commonwealth trade was 2.6 to 4.1 times higher than geography and size can explain, even before the protectionist Ottawa agreements of 1932.

Given the historical circumstances of the interwar years, it is surprising that our results are so clear-cut. In 1928, the gold-exchange standard was in full swing (see Wolf 2008), with countries accumulating gold and exchange reserves and world trade recovering swiftly from more than a decade of disruption. In contrast, it has often been stressed that the formation of the various currency blocs in the 1930s was massively trade diverting, because these arrangements were typically accompanied by rises in tariffs, quotas, and capital and exchange controls along those blocs (see Findlay and O'Rourke 2007). However, once we control for pre-existing trade integration, the differential impact of bloc formation in the 1930s, whether driven by currency, tariffs or other factors, becomes minimal and mostly insignificant.

The result would look very different under individual, observation-specific CPFE which are often used in the empirical literature We argued above that overfitting with CPFE in the gravity equation causes any policy treatment effects to be spuriously significant. Table III compares FET and CPFE identifications of the treatment effect under otherwise identical specifications.

(Table III about here)

Two results stand out from Table III. First, the estimated treatment effects under FET and CPFE are far lower than in the pooled estimated of Table I, and quite close to the difference in differences estimates of Table II. This implies that at least in our natural experiment, the characteristics of the gravity model *Z* and the MTR terms do not play a decisive role in identifying the treatment effects. The critical step clearly is controlling for pre-existing trade integration among bloc members. Second, the estimated standard errors under CPFE are far lower than under FET, suggesting that the overfitting problem in the CPFE specification is pervasive. The CPFE results would imply that all arrangements except for the gold and reichsmark blocs had positive, strongly significant trade effects. However, none of the treatment effects is anywhere near significant confidence levels when estimated under the correct FET specification. This holds independently of whether the treatment effects of bloc membership in the 1930s are estimated jointly or separately, whether with PPML or Scaled OLS.[6] The clear-cut lesson from Table III is that while CPFE in the gravity context broadly capture the size of the trade creation effect correctly, they generate strongly downward biased standard errors and hence should be avoided.

Drawing the conclusions from this section together, our results suggest that the formation of trade and currency blocs tends to be path dependent, and that traditional ways to measure their effects in the gravity equation fail to detect this. Such failure to account for pre-existing trade integration may lead to misperceptions of the impact of institutional change on economic relations. We showed that difference in difference estimation is an easy way to deal with this form of endogeneity, and has the advantage of avoiding spurious precision in the standard errors. Whatever the refinements of the specifications and estimation techniques (we tried many more), they only strengthen the main result: currency blocs do not necessarily create trade. Instead, trade often creates currency blocs.

IV. Treatment Effects in Cross-Section Estimates

IV.1. Identification through Selectivity and Matching

Not always is it possible to identify treatment effects from the panel dynamics of the gravity equation. And when it is, strong identifying assumptions need to be made: Ashenfelter dips (i.e., anticipation effects) must either be absent or clearly identifiable using additional information, and common trends must be present among both the treated and the untreated. If non-random selection cannot be identified along the time axis, the coefficient on the policy treatment dummy will be correlated with the residual, and hence be biased. Instrumental variables approaches will hardly solve this issue, as "all instruments will be bad instruments" in the case of currency unions (Baldwin 2006, 2.6.2). Labor econometrics suggests tackling this kind of problem by using matching techniques, which mimic a randomized experiment (see Persson 2001 and Baier and Bergstrand 2007

---

[6] We obtained very similar results (not shown) with Tobit, as well as OLS in logs.

for applications to the gravity equation). The idea of this is to find common characteristics that help explain the selection of country pairs into currency arrangements. These characteristics will at the same time under- and over-explain the formation of currency arrangements. The false positives, or type I errors, of assigning non-members into the treatment can be exploited to build a counterfactual control group.

Let D again be a 0/1 indicator indicating the presence or absence of the treatment. In the panel approach, we had to assume that after conditioning on characteristics, suitable fixed effects and time effects $\widetilde{X}$, any correlation between the residuals from the gravity equation and the treatment itself was removed, i.e.:

$$u_{ij} \perp D \mid \widetilde{X}, D = 0,1 \tag{8}$$

An intuitive way to proceed if this assumption is violated is the method of matching. The method assumes that one has access to a set of conditioning variables Z such that

$$(ij, t \in TR) \perp D \big| P(Z) \tag{9}$$

Here, $P(Z) = PR\ (D = 1 \big| Z)$ denotes the probability or propensity score of participation in a currency arrangement conditional on Z. Estimation of this propensity score is the first step in implementation. Based on a list Z of pairwise characteristics, one estimates the probability that a certain pair of countries takes part in a currency arrangement, and uses this to construct a counterfactual control group. In a second step, estimation of the gravity equation (4) is repeated, comparing treated and control group observations with similar propensity scores to each other. If the propensity to select into the treatment is captured by the list of pairwise characteristics, the method of matching delivers an unbiased estimator of (4).

To estimate the propensity score, Persson (2001) and similarly Barro and Tenreyro (2003) suggest using the geographical characteristics of the gravity model itself. The idea of this is to exploit the non-linearity of the propensity score estimators. However, there is a problem with this: the method of Persson (2001) potentially misses the most important driving force of self-selection into treatment, the level of bilateral trade integration among the future members of an arrangement. Using only the gravity variables to construct the counterfactual reference group implies that by definition, trade integration unexplained in a gravity model will not be taken into account. Since the work of Hamilton and Winters (1992), Frankel and Wei (1993), and Baldwin (1994), the gravity model is the standard tool for assessing the degree of economic integration between countries.

IV.2. Endogenous policies and exogenous anchors

The deeper issue at stake is the potential endogeneity of policy itself. Besley and Card (2000) have argued that standard econometric methods to assess self selection might fail if the treatment itself is endogenous and therefore contaminated by selectivity issues. Viewed from this perspective, the task of explaining treatment effects in the gravity model may have to go beyond the model itself, and include deep parameters that are exogenous to the choice of the policy treatment. The gravity variables permit estimation of potential trade levels between countries. Deviations between observed and predicted trade may therefore be interpreted as a proxy for economic integration beyond the gravity model itself.

To fix ideas, we borrow the idea of Alesina and Barro (2002) that currency arrangements often form around anchor countries. The concept of client-anchor relationships in currency arrangements seems well adapted to the political rivalry among Europe's powers after World War I. In line with the results from the panel estimates of the previous section, we would expect currency arrangements in the inter-war period to follow these political fault lines. In fact, this provides a simple and straightforward way to test the OCA hypothesis. If the hypothesis holds, trade integration with an anchor country in 1928 should help to predict future membership in the currency arrangements, even after controlling for the gravity variables as specified in Section II. Crucially, trade in 1928 can safely be regarded as exogenous, because the currency arrangements in question were all formed during or after the Great Depression, and as the Great Depression was an unforeseen event.

In the following we implement the anchor country concept in two different ways. The first is to include trade with anchor countries in 1928 in a probit model of selection into the trade and currency blocs of the 1930s. The second is to use trade with the anchor countries in 1928 as an instrument in gravity equations for the 1930s.

To obtain the propensity scores for entering a currency arrangement, we estimate separate binary choice models for all currency blocs under consideration. The dependent variable in each regression is the respective membership dummy for currency arrangements, FET_{ij}^{m}, defined above in section II. For each country pair *ij*, it takes the value of one if both countries *i* and *j* will be members of the same currency arrangement. The regressors are the observed bilateral trade flows of 1928, the trade with each of the three anchor countries in 1928, and the gravity controls for potential trade. We estimate the following probit model:

$$\text{Prob}\,(FET_{ij}^{m} \mid X_{ij}\beta) = 1 - \Phi(-X_{ij}^{'}\beta) \tag{10}$$

where *Xij* denotes the matrix of trade flow variables and gravity controls for 1928, and where Φ is the cumulative distribution function of the normal distribution (Table IV).

(Table IV about here)

The coefficients on trade with the anchor countries all have the expected signs. Interestingly, low trade integration with an anchor country in 1928 powerfully selects against joining a currency arrangement anchored by that country. Countries that traded intensely with Britain in 1928 but rather less with Germany or France, were highly likely to be members of the Sterling Bloc in the 1930s. But also, countries that traded intensively with Germany in 1928, and less so with France and Britain, were likely to be members of the Reichsmark or the Exchange-Control Bloc in the 1930s. Similarly, future membership in the Gold Bloc is very strongly predicted by low trade integration with Germany or Britain in 1928. We also notice that for the Reichsmark and Exchange Control blocs, overall trade enters negatively. Countries that later joined either of the two blocs thus were on the whole less integrated into the international economy already in 1928. This confirms that trade dependence on Germany, which was instrumental in Germany's aggressive trade policies of the 1930s, was a pre-existing condition, see Ritschl (2001).[7]

In a next step, the propensity scores generated by the probit models in Table IV are used to construct control groups for each currency arrangement. We evaluate the effects of actual membership in the 1930s against all observations with a similar propensity score.

[7] There is historically valuable information also in the false positives generated by the Probit model. To pick one example, the model for the gold bloc wrongly assigns most of Czechoslovakian and Austrian trade to the gold bloc. This is wrong for 1935 and 1938 where we have observations, yet both countries departed from the gold bloc only in 1934, and mimicked it closely until the German occupation. See Eichengreen (1992) for the details.

Table V reports the results for a cutoff level of 0.5. We also lowered the cutoff to two standard errors of the residual (typically around .25), without any qualitative change in results.

(Table V about here)

Table V examines trade creation in the gravity model, just as Table II above. However, attention is now limited to the 1930s, and to the control groups created through the binary choice models of Table IV. Relative to the respective control group, trade creation among the actual currency bloc members is insignificant or even negative. The only exception to this result is the exchange control bloc around Germany. Here, the estimate appears to confirm the evidence from the PPML estimator in Table II above. Results were robust to changes in specifications and estimation methods.[8]

The results of this section are also in line with the results of Section II. First, the difference in differences estimator in the panel (Table II) and the selectivity corrections in the cross section (Table V) are in agreement on the absence of treatment effects. These results confirm our claim that the treatment effects we obtained by conventional gravity estimates in Table I were spurious. Second, the group specific fixed effects (or fixed effects on the treated) from Table II and the probit model from Table IV agree that most of the action in the historical dataset comes from pre-existing trade integration with core countries. Not only do currency areas create trade, but trade also creates currency areas. This is what the theory of Optimum Currency Areas (OCA) has always predicted.

V. Concluding Remarks

Endogeneity of currency areas, trade blocs, and similar political arrangements is often only incompletely addressed in the gravity literature. In this paper, we studied a natural experiment to argue that standard estimates of the trade creation effects of such arrangements are often misspecified and yield spurious results. Borrowing from labor econometrics, we proposed easy ways to implement proper treatment effect methods in the

---

[8] One might worry that some of the various currency arrangements were mutually exclusive. For example, one could not simultaneously join the Gold-Bloc and become a member of the Reichsmark-Bloc, nor was it possible to stay on Gold and join the Sterling-Bloc. To address this issue, we re-estimated the propensity score to join any one currency arrangement with a 4 choice multinomial logit model for the adoption of one of the three currency-blocs of the 1930s. This left our results qualitatively unchanged.

gravity equation. Under fixed effect identifications, this usually implies difference in differences estimation. We presented a simple implementation of this estimator in the gravity equation, which employs group specific fixed effects. Implemented properly, this estimator avoids the issue of overfitting, that is spuriously low standard errors that come with individual country pair fixed effects.

For policy endogeneity that cannot be eliminated by fixed effects, we suggested applying the concept of anchor countries of Alesina and Barro (2002) to the gravity model. We found that pre-existing trade with such anchor countries performs well, both in binary matching models and in the conventional gravity equation. In the binary matching context, we employed trade with anchor countries as a predictor of currency bloc formation. This generalizes previous approaches by Persson (2001) and Baier and Bergstrand (2007) to the endogeneity of policy itself, addressing an identification issue raised by Besley and Case (2000). In the gravity equation, we used pre-existing trade with anchor countries as an instrument for trade creation by currency blocs. This suggests itself as a promising, less data-hungry alternative to the binary matching approach.

To make our point, we studied a near-ideal natural experiment from history, the formation of various currency blocs in Europe after the Great Depression. Our choice of this particular historical episode is motivated by the favorable characteristics of the data. First, due to the unexpected breakdown of the Gold Standard in the wake of 1929, the trade data are unlikely to be contaminated by anticipation effects. Second, "pre-treatment" observations are available for the whole sample: the researcher can actually observe trade flows for the Gold Standard of the late 1920s, prior to the beginning of the Great Depression, and prior also to the establishment of the currency and trade blocs. Finally, previous studies on interwar trade using standard specifications of the gravity equation have obtained very high trade creation coefficients for the currency and trade blocs in question, similar in magnitude to the evidence produced by Rose (2000, 2001) and Glick and Rose (2002) for postwar data. However, none of the arrangements of the 1930s were currency unions. Instead, the currency blocs of the 1930s were loose fixed exchange rate systems combined with tariff barriers, tight quotas, and capital controls. This suggested that the high trade creation effects found in standard gravity equations were indeed spurious.

Analyzing the evidence from the interwar period with proper regard to identifying treatment effects, we indeed found pervasive evidence of endogeneity and self selection, which standard approaches to the gravity equation did not detect. Under group specific fixed effects in the panel, we obtained almost no treatment effects of bloc formation in the 1930s. We confirmed these findings in the cross section, employing matching models and instrumental variables designed to capture self selection and policy endogeneity. Our results caution against optimism about trade creation effects of political arrangements. It seems that to a large extend such arrangements are endogenous to the pre-existing pattern of trade.

References

Alesina, Alberto and Robert Barro (2002), "Currency Unions," *Quarterly Journal of Economics* CXVII, 409-36.

Baier, Scott and Jeffrey Bergstrand (2007), "Do Free Trade Arrangements Actually Increase Members' International Trade?," *Journal of International Economics* 71, 72-95.

Baldwin, Richard (1994), *Towards an Integrated Europe*, London: CEPR.

Baldwin, Richard (2006), *The Euro's Trade Effects*, Frankfurt: European Central Bank Working Paper 594.

Barro, Robert J. and Silvana Tenreyro (2003), "Economic Effects of Currency Unions," *NBER Working Paper* 9435.

Berger, Helge and Volker Nitsch (2005), "Zooming Out: The Trade Effect of the Euro in Historical Perspective," *mimeo*.

Besley, Timothy and Anne Case (2000), "Unnatural Experiments? Estimating the Incidence of Endogenous Policies," *Economic Journal* 110, 672-694.

Blundell, Richard and Monica Costa Dias (2002), "Alternative Approaches to Empirical Evaluation in Empirical Microeconomics," *Portuguese Economic Journal* 1, 91-115.

De Groot, Henri L. F., Gert-Jan Linders, Piet Rietveld, and Uma Subramanian (2004), "The Institutional Determinants of Bilateral Trade Patterns", *Kyklos* 57 (1), pp. 103-123.

Dixit, Avinash (2000), "A Repeated Model of Monetary Union," *Economic Journal*, 759-780.

Dominguez, Kathryn, Ray Fair and Matthew Shapiro (1988), "Forecasting the Depression: Harvard versus Yale," *American Economic Review* (78), 595-612.

Eichengreen, Barry (1992), *Golden Fetters. The Gold Standard and the Great Depression 1919-1939*, Oxford: Oxford University Press.

Eichengreen, Barry and Douglas Irwin (1995), "Trade Blocs, Currency Blocs, and the Reorientation of Trade in the 1930s," *Journal of International Economics* 38, 1-24.

Eichengreen, Barry and Douglas Irwin (1997), The Role of History in Bilateral Trade Flows, in: Jeffrey A. Frankel, ed. *The Regionalization of the World Economy*, Chicago: University of Chicago Press.

Ellis, Howard (1941), *Exchange Control in Central Europe*, Cambridge.

Estevadeordal, Antoni, Brian Frantz and Alan Taylor (2003), "The Rise and Fall of World Trade, 1870-1939," *Quarterly Journal of Economics* 118, 359-407.

Flam, Harry and Hakan Nordstrom (2003), *Trade Volume Effects of the Euro: Aggregate and Sector Estimates*, Stockholm: Institute for International Economic Studies, mimeo.

Flandreau, Marc and Mathilde Maurel (2001), "Monetary Union, Trade Integration and Business Fluctuations in 19th Century Europe: Just Do It," *CEPR Discussion Paper*, No. 3087.

Frankel, Jeffrey A and Andrew Rose (1997), "Is EMU More Justifiable Ex Post Than Ex Ante?," *European Economic Review* 41, 753-60.

Frankel, Jeffrey A and Andrew Rose (1998), "The Endogeneity of the Optimum Currency Area Criteria," *Economic Journal* 108, 1009-1025.

Frankel, Jeffrey and Shang-Jin Wei (1993), Is There a Currency Bloc in the Pacific?, in: Adrian Blundell-Wignall, ed. *Exchange Rates, International Trade and the Balance of Payments*, Sidney: Reserve Bank of Australia.

Hamilton, Carl and Alan Winters (1992), "Opening Up International Trade with Eastern Europe," *Economic Policy* 14, 77-104.

Hamilton, James D. (1987), "Monetary Factors in the Great Depression," *Journal of Monetary Economics* 19, 145-169.

Hamilton, James D. (1992), "Was the Deflation during the Great Depression Anticipated? Evidence from the Commodity Futures Market," *American Economic Review* 82 (82), 157-177.

Hilgerdt, Folke (1942), *The Network of World Trade*, Geneva: League of Nations.

López-Córdova, J and Christopher Meissner (2003), "Exchange Rate Regimes and International Trade: Evidence from the Classical Gold Standard Era," *American Economic Review* 93, 344-353.

Maddison, Angus (1995), *Monitoring the World Economy, 1820-1990*, Paris: OECD.

Maurel, Mathilde and Guillaume Cheikbossian (1998), "The New Geography of Eastern European Trade", *Kyklos* 51 (1), pp. 45-72.

McKinnon, Ronald (1963), "Optimum Currency Areas," *American Economic Review* 53, 717-725.

Micco, Alejandro, Ernesto Stein and Guillermo Ordoñez (2003), "The Currency Union Effect on Trade: Early Evidence from EMU," *Economic Policy* 37, 315-356.

Mongelli, Francesco Paolo, Ettore Dorrucci and Itai Agur (2007), "What Does European Institutional Integration Tell Us About Trade Integration?," *Integration and Trade* 26, 151-200.

Mundell, Robert (1961), "A Theory of Optimum Currency Areas," *American Economic Review* 51, 657-665.

Nitsch, Volker (2002), "Honey, I Just Shrunk the Currency Union Effect on Trade," *World Economy* 25, 457-474.

Obstfeld, Maurice and Alan Taylor (2003), "Sovereign Risk, Credibility and the Gold Standard: 1870-1913 versus 1925-31," *Economic Journal* 113, 241-274.

Persson, Torsten (2001), "Currency Unions and Trade: How Large Is the Treatment Effect?," *Economic Policy* 33, 435-48.

Prados, Leandro (2000), "International Comparisons of Real Product, 1820-1990: An Alternative Data Set," *Explorations in Economic History* 37, 1-41.

Ritschl, Albrecht (2001), "Nazi Economic Imperialism and the Exploitation of the Small: Evidence from Germany's Secret Foreign Exchange Balances, 1938-40," *Economic History Review* 54, 324-345.

Ritschl, Albrecht (2009), "Spurious Significance of Treatment Effects in Overfitted Fixed Effect Models," *mimeo*, London School of Economics.

Rose, Andrew (2000), "One Money, One Market: Estimating the Effect of Common Currencies on Trade," *Economic Policy* 30, 7-45.

Rose, Andrew (2001), "Currency Unions and Trade: The Effect is Large," *Economic Policy* 33, 449-61.

Rose, Andrew (2002), "Honey, the Currency Union Effect on Trade Hasn't Blown Up," *World Economy* 25, 475-479.

Santos Silva, Joao and Silvana Tenreyro (2006), "The Log of Gravity," *Review of Economics and Statistics* 88, 641-658.

Temin, P. (1989), *Lessons from the Great Depression*, Cambridge, Mass: MIT Press.

Wolf, N. (2008), "Scylla and Charybdis. Explaining Europe's Exit from Gold, 1928-1936". *Explorations in Economic History* 45, 383-401.

TABLE I

POOLED PANEL REGRESSIONS ON TRADE EFFECTS OF INTERWAR CURRENCY BLOCS,

- Gravity Equations with Multilateral Trade Resistance Terms -

	PPML	Scaled OLS	OLS	Tobit
Dependent :	*trade*	ln (1+ *trade*)	ln (*trade*)	ln (*trade*)
TIME FIXED EFFECT 1930s	-0.601	-0.578	-0.531	-0.531
	<-1.649>	<-2.220>	<-1.784>	<-1.719>
LOG (DISTANCE)	-0.328	-0.587	-0.697	-0.752
	<-6.181>	<-12.721>	<-14.943>	<-15.391>
LOG (EXCHANGE RATE VOLATILITY INDEX	-1.933	-2.460	-2.192	-2.780
	<-2.753>	<-4.589>	<-3.406>	<-3.739>
COMMONWEALTH TRADE BLOC	1.487	0.897	1.308	1.529
	<7.780>	<4.356>	<6.680>	<7.545>
STERLING BLOC	0.903	0.408	0.540	0.587
	<6.841>	<5.291>	<5.867>	5.591>
GOLD BLOC	0.262	-0.181	-0.123	-0.236
	<3.176>	<-2.399>	<-1.387>	<-2.413>
EX GOLD BLOC 1938	0.074	0.168	0.176	0.177
	<1.182>	<2.497>	<2.105>	<1.833>
REICHSMARK BLOC	0.681	0.877	1.005	1.114
	<3.539>	<7.719>	<7.952>	<8.273>
EXCHANGE CONTROL BLOC	0.550	0.260	0.248	0.230
	<4.043>	<4.054>	<3.320>	<2.809>
No. of observations	2610	2610	2136	2136
Left-censored observations	-	-		229
Adjusted R-squared	0.840	0.824	0.804	
Log likelihood	-26214.0	-2839.0	-2389.7	-2509.0
SEE	54.953	0.737	0.764	0.764
Mean dependent var	45.002	2.213	2.524	2.524
S.D. dependent var	137.225	1.712	1.674	1.674

OLS and Scaled OLS estimates: White/Arellano HAC *t* statistics in chevrons.

PPML and Tobit estimates: heteroskedsasticity consistent Huber/White *z* statistics in chevrons.

Gravity controls in the equations (not reported): multilateral trade resistance, common border, common language

TABLE II
TREATMENT EFFECTS OF INTERWAR CURRENCY BLOCS, 1928-1938

- Difference-in-differences estimates with multilateral trade resistance terms -

	PPML	Scaled OLS	OLS	Tobit
Dependent:	*trade*	ln(1+*trade*)	ln(*trade*)	ln(*trade*)
TIME FIXED EFFECT 1930s	-0.763	-0.749	-0.725	-0.726
	<-2.118>	<-2.888>	<-2.458>	<-2.378>
LOG(DISTANCE)	-0.381	-0.548	-0.667	-0.719
	<-10.044>	<-12.484>	<-14.411>	<-14.939>
LOG (EXCHANGE RATE VOLATILITY INDEX)	-1.726	-2.552	-2.238	-2.822
	<-2.877>	<-4.699>	<-3.497>	<-3.815>
Treatment Effects:				
COMMONWEALTH TRADE BLOC	0.307	0.183	0.239	0.347
	<1.295>	<0.506>	<0.749>	<1.023>
STERLING BLOC	-0.057	-0.338	-0.057	0.038
	<-0.317>	<-2.330>	<-0.352>	<0.217>
GOLD BLOC	0.150	0.073	0.094	0.061
	<1.216>	<0.500>	<0.585>	<0.360>
EX GOLD BLOC 1938	0.078	0.168	0.176	0.178
	<1.918>	<2.730>	<2.363>	<2.082>
REICHSMARK BLOC	0.052	-0.387	-0.420	-0.385
	<0.212>	<-1.981>	<-1.878>	<-1.664>
EXCHANGE CONTROL BLOC	0.271	0.101	0.015	-0.007
	<1.648>	<0.786>	<0.103>	<-0.046>
Fixed Effects on the Treated:				
FE COMMONWEALTH TRADE BLOC	1.350	0.959	1.319	1.423
	<6.294>	<2.949>	<4.703>	<4.751>
FE STERLING BLOC	0.987	0.753	0.603	0.552
	<7.162>	<6.022>	<4.403>	<3.710>
FE GOLD BLOC	0.194	-0.205	-0.170	-0.245
	<1.872>	<-1.628>	<-1.230>	<-4.751>
FE REICHSMARK BLOC	0.680	1.305	1.459	1.529
	<4.349>	<8.070>	<7.839>	<8.049>
FE EXCHANGE CONTROL BLOC	0.332	0.181	0.252	0.258
	<3.176>	<1.623>	<1.992>	<1.986>
No. of observations	2610	2610	2136	2136
Left-censored observations				229
Adjusted R-squared	0.895	0.826	0.806	
Log likelihood	-22478	-2750	-2310	-2438
SEE	45.741	0.713	0.737	0.737
Mean dependent var	45.002	2.214	2.524	2.524
S.D. dependent var	137.225	1.712	1.674	1.674

OLS and Scaled OLS estimates: White/Arellano HAC *t* statistics in chevrons.

PPML and Tobit estimates: z statistics in chevrons.

Gravity controls in the equations (not reported): multilateral trade resistance, common border, common language

TABLE III
SPURIOUS PRECISION OF TREATMENT EFFECTS UNDER GROUP AND COUNTRY PAIR FIXED EFFECTS

	PPML												SCALED OLS	
Dependent :	*trade*												ln (1+ *trade*)	
CONSTANT	4.309		4.356		4.353		4.371		4.444		4.330	1.913	2.471	
	<0.095>		<0.097>		<0.094>		<0.095>		<0.101>		<0.120	<0.288>	<0.077>	
YEARS1930	-1.070	-1.070	-1.091	-1.091	-1.045	-1.046	-1.054	-1.054	-1.081	-1.081	-1.173	-1.165	-0.589	-0.589
	<0.112>	<0.026>	<0.116	<0.026>	<0.113	<0.026>	<0.011	<0.022>	<0.122	<0.028>	<0.142	<0.023>	<0.088	<0.026>
COMMONWEALTH TRADE BLOC	0.288	0.288									0.166	0.122	0.063	0.063
	<0.353>	<0.072>									<0.397	<0.095>	<0.712	<0.284>
STERLING BLOC			0.397	0.397							0.369	0.380	0.196	0.196
			<0.318	<0.053>							<0.306	<0.067>	<0.299	<0.073>
GOLD BLOC					0.017	0.017					0.145	0.138	-0.275	-0.275
					<0.380	<0.077>					<0.390	<0.076>	<0.356	<0.068>
EX GOLD BLOC 1938					0.092	0.092					0.092	0.092	0.162	0.162
					<0.330	<0.046>					<0.330	<0.046>	<0.310	<0.048>
REICHSMARK BLOC							0.202	0.202			0.240	0.194	-0.288	-0.288
							<0.345	<0.219>			<0.354	<0.239>	<0.273	<0.086>
EXCHANGE CONTROL BLOC									0.284	0.284	0.321	0.310	0.101	0.101
									<0.231	<0.050>	<0.237	<0.065>	<0.182	<0.048>
Fixed Effects on the Treated (FET)	X		X		X		X		X		X		X	
Individual Country-Pair Fixed Effects (CPFE)		X		X		X		X		X		X		X
Multilateral trade resistance terms	no	no	no	no	no	no	no	no	no	no	no	no	no	no
No. of observations	2610	2610	2610	2610	2610	2610	2610	2610	2610	2610	2610	2610	2610	2610
Adjusted R-squared	0.061	0.952	0.032	0.957	0.033	0.926	0.030	0.952	0.032	0.953	0.064	0.962	0.074	0.926

Scaled OLS:White/Arellano HAC errors in chevrons

PPML: Huber/White QML errors in chevrons

TABLE IV
PREDICTING CURRENCY BLOC MEMBERSHIP
FROM BINARY PROBIT MODEL OF TRADE IN 1928

	Gold Bloc	Str Bloc	RM Bloc	ExchCtrl Bl
Dependent:	FET^{Gold28}	FET^{Str28}	FET^{RM28}	$FET^{ExCtr28}$
LOG(DIST)	-1.585	0.223	-0.243	-0.851
	<-5.965>	<1.920>	<-3.053>	<-9.295>
LOG(TRADE 1928)	0.058	0.136	-0.348	-0.235
	<0.399>	<1.011>	<-5.034>	<-2.967>
LOG(TRADE 1928)*FR	0.182	-3.951	-0.016	-5.474
	<2.311>	<-14.467>	<-0.286>	<-32.036>
LOG(TRADE 1928)*UK	-2.420	0.817	-0.007	-3.107
	<-16.889>	<7.300>	<-0.134>	<-35.453>
LOG(TRADE 1928)*GE	-2.456	-1.939	0.521	0.229
	<-20.075>	<-11.930>	<6.679>	<4.477>
BORDER	-1.645	0.979	0.121	-0.399
	<-4.066>	<2.399>	<0.500>	<-1.701>
LANGUAGE	0.500	0.017	-0.619	-0.559
	<0.920>	<0.040>	<-0.918>	<-1.393>
McFadden R-squared	0.500	0.608	0.364	0.290
Log likelihood	-45.644	-79.820	-273.009	-244.774
Restr. log likelihood	-91.249	-203.686	-429.324	-344.951
LR statistic (12 d.o.f.)	91.209	247.732	312.630	188.462
Hannan-Quinn criter.	0.196	0.292	0.832	0.753
Obs. w/ depvar = 1	20	59	206	134
Pred. as depvar = 1	11	46	269	65
correct	7	32	168	32
incorrect	4	14	101	33
Obs. w/ depvar = 0	850	656	509	581
Pred. as depvar = 0	846	642	408	548

z statistics in chevrons

Controls (nor reported): common border, common language, German border

Cutoff score: p = .5

TABLE V

TRADE CREATION RELATIVE TO MATCHING CONTROL GROUPS, 1930s

	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML
Dependent :	ln(*trade*)	*trade*	ln(*trade*)	*trade*	ln(*trade*)	*trade*	ln(*trade*)	*trade*
LOG(DIST)	-0.699	-0.231	-0.737	-0.400	-0.665	-0.255	-0.716	-0.222
	<-13.462>	<-3.252>	<-14.250>	<-8.393>	<-14.035>	<-4.882>	<-13.102>	<-3.981>
GOLD BLOC								
CONTROL GROUP	0.516	0.382						
	<0.967>	<0.970>						
TREATMENT EFFECT	-0.351	0.068						
	<-0.653>	<0.165>						
CONTROL GROUP 1938	-0.465	-0.352						
	<-0.614>	<-0.794>						
TREATMENT EFFECT 1938	0.561	0.433						
	<0.729>	<0.889>						
STERLING BLOC								
CONTROL GROUP			0.670	1.321				
			<3.381>	<7.955>				
TREATMENT EFFECT			0.141	0.208				
			<0.637>	<1.214>				
REICHSMARK BLOC								
CONTROL GROUP					-0.775	-0.921		
					<-10.571>	<-5.654>		
TREATMENT EFFECT					-0.051	0.185		
					<-0.593>	<0.801>		
EXCHANGE CONTROL BLOC								
CONTROL GROUP							-0.173	0.256
							<-1.523>	<1.064>
TREATMENT EFFECT							0.479	0.420
							<3.747>	<2.366>
Import, export country effects x time effects	yes	yes	yes	yes	yes	yes	yes	yes
No. of observations	1337	1740	1337	1430	1337	1430	1337	1430
Adjusted R-squared	0.755	0.653	0.768	0.763	0.784	0.660	0.759	0.667
Log likelihood		-15111				-13455		
Mean dependent var	2.403	0.172	2.403	34.136	2.403	34.136	2.403	34.136

OLS estimates: White heteroskedasticity consistent t statistics in chevrons.

PPML estimates: White heteroskedasticity consistent z statistics in chevrons.

Gravity controls in the equations (not reported): common border, common language, German border