Currency Unions and Trade: How Large is the Treatment Effect?*

Torsten Persson
Institute for International Economic Studies†

May 30, 2001

1 INTRODUCTION

It has been conspicuously hard to empirically identify clear-cut effects of fixed exchange rates and other monetary regimes on the real economy.[1] This paucity of empirical findings has lead many academic economists as well as other observers to the view that radical monetary reform – such as the adoption of a common currency – might have limited economic consequences. For example, it probably explains the common view that the EMU is "first and foremost a political rather than an economic project".

A recent Economic Policy paper by Andy Rose challenges the conventional wisdom (Rose, 2000). He uses evidence from existing currency unions in the world economy to estimate the effects of a common currency on trade. According to his regression estimates, a currency union expands bilateral trade between two average member countries by a dazzling 200 percent or more. Given the novelty of the finding and the importance of the issues, these results have received considerable attention.

Critics voiced a number of concerns about Rose's methodology, questioning the accuracy of his findings. Provoked by the critique, Rose (2000) conducted a large battery of robustness checks, however, showing that the central result holds up to the points raised by the critics. Recent work by Rose and van

*Paper presented at the *Economic Policy* panel meeting in Stockholm on April 6-7, 2001. I would like to thank members of the panel, as well as Robert Barro, Harry Flam, Patrick Honohan, Hide Ichimura, Per Pettersson, Andy Rose, David Strömberg and Lars Svensson for their comments. I am also grateful to Alessandra Bonfiglioli for research assistance and Christina Lönnblad for editorial assistance. The research was supported by a grant from FRIE.

†Stockholm University
E-mail: torsten.persson@iies.su.se
Web: http://www.iies.su.se/~perssont/

[1] Baxter and Stockman (1989) is one of the first systematic studies of the consequences for the real economy of different exchange rate regimes. See Frankel and Rose (1995) for an overview.

Wincoop (2001) goes beyond within-sample estimation to the more difficult question of out-of-sample prediction, investigating the hypothetical trade and welfare effects of new currency unions with alternative country constellations.

In this short paper, I take another look at Rose's within-sample estimates of the "treatment effect of currency unions on trade", borrowing this language and some methodology from the labor literature. I argue that these estimates might be seriously biased if the countries belonging to existing currency unions are non-randomly selected and the relation from measured trading costs to trade is non-linear; I also argue that such complications are likely to exist in reality. I thus suggest an alternative empirical strategy that is more robust to misspecification than Rose's linear regression strategy. More specifically, I use simple, non-parametric matching estimators that allow for systematic selection into currency unions as well as non-linear effects of trading costs on trade. While these techniques were originally developed for medical applications, they are now quickly making their way into the standard tool-box of labor economists. Applying them, I obtain considerably smaller treatment effects of a common currency than the 200 plus percent obtained by Rose: my preferred point estimates range from 13 to 65 percent, but with large enough standard errors that they are not significantly different from zero.

The next section explains the prospective problems with Rose's empirical strategy, Section 3 suggests an alternative matching approach, while Sections 4-6 present my empirical results and discuss why they are so different from those obtained by Rose.

2 ROSE'S EMPIRICAL STRATEGY – WHAT'S THE PROBLEM?

Rose (2000) uses United Nations data on trade among close to 200 countries in five consecutive five-year periods from 1970, yielding more than 33,000 observations of bilateral trade flows. About 330 of these trading pairs share a common currency. (See the original paper for a more detailed discission of the data and their sources.) As there are very few regime changes – countries entering or leaving common currencies – in these data, the treatment effect of a common currency on trade must be identified from the cross-sectional variation.

Specifically, Rose estimates a linear gravity equation, where the log of bilateral trade flows in country pair i, t_i, is regressed on a vector x_i, including about 10 different measures of trading costs in a wide sense. To this equation, he adds an indicator variable c_i which takes a value of 1 if pair i shares a common currency and 0 if it does not. The coefficient γ on c_i measures the treatment effect of a common currency. The default estimate (Table 2, column 6) of γ is 1.21 with a relatively tight confidence band. As $e^{1.21} - 1 \cong 2.35$, this suggests an expansion of trade by 235%. Countries entering into common currencies also achieve bilateral exchange rate stability, which might further expand trade. Rose's point of estimate of this effect is much smaller; evaluating it at the sample mean of exchange rate volatility, adds about another 10% to bilateral trade.

In response to doubtful or disbelieving readers of the paper – not least the members of the Economic Policy panel – Rose subjects his basic estimates to a large number of careful robustness checks, investigating their sensitivity to the sample, the specification, and the method of estimation. None of these checks substantially change the magnitude of the treatment effect, however.

From these results it thus appears that the effect of a common currency on trade is much larger than previously thought. If this is true, the trade implications are indeed spectacular for countries contemplating unilateral dollarization, euroization or the formation of a new common currency.

I want to argue that Rose's empirical strategy could substantially bias his estimates. Let me first explain the general problem. Rose runs a variety of linear regressions of t_i on x_i and c_i. Suppose now that the effect of either x_i or c_i on t_i depends on the level of x_i. In addition to this non-linearity, suppose that c_i is non-randomly selected such that it is systematically correlated with x_i. Under these conditions, a linear estimate of γ can be an unreliable estimate of the treatment effect for two reasons. Existing currency unions (observations with $c_i = 1$) might have a large effect on trade due to their specific characteristics (values of x_i), which do not generalize to other country pairs with different characteristics (different values of x_i).[2] Or, omitted non-linear terms in the relation between some component of x_i and t_i might be picked up by c_i (as x_i and c_i are correlated).[3] In both cases, the estimate of the average treatment effect is subject to selection bias. Note, however, that the prospective problem here is neither selection on unobservables (omitted variables), nor non-random selection of missing observations – problems that Rose does attempt to address with instrumental-variable and Heckit estimators, respectively. Rather, it is an instance of selection on observables (i.e., of c_i on x_i).[4]

Is the prospective specification bias more than a remote theoretical possibility in a long laundry list of imaginable statistical difficulties? I would like to argue that we have good – theoretical as well as empirical – reasons to believe that it might strike with particular force against Rose's data and specification.

Consider first the relation between trading costs and trade. The work by McCallum (1996) and others on the home-bias puzzle in trade strongly suggests that low trading costs can have massive effects on trade. Obstfeld and Rogoff's (2000) analysis of major puzzles in international macroeconomics also

[2] In this case – referred to as heterogeneous treatment effects in the labor literature – the estimate of the so-called *average effect of treatment on the treated* does not coincide with the *average treatment effect* (the expected effect on bilateral trade of a common currency on a randomly drawn country pair).

[3] In this case both the average treatment effect and the average effect of treatment on the treated will be estimated with bias.

[4] Labor economists have compared different sources of bias when estimating treatment effects with non-experimental data, using the results obtained by controlled experiments as a benchmark. For example, in their study of the JTPA (a major U.S. job training program), Heckman, Ichimura and Todd (1997) find that an unbalanced distribution of observables among treated and non-treated is a considerably more important source of bias than the conventional selection problem.

suggests a non-linear relation – on both theoretical and empirical grounds – as does the model of currency-union formation in the recent work of Alesina and Barro (2000); see further below. Furthermore, we do not measure trading costs directly but only some of their proximate determinants (x_i in the notation above). Why on earth would the relation between these proxies and true trading costs be linear? For instance, sharing a common colonial history (and thus similar institutions) together with geographical proximity, or a common language together with a regional trade agreement, might reduce trading costs between two countries by more than the summed partial effects of these features.

What about non-random selection? In one of the few **positive** models of the adoption of common currencies, Alesina and Barro (2000) suggest that the benefits of a common currency – as well as the trade effects thereof – are likely to be particularly high if trading costs are low for other reasons. In terms of our notation, this corresponds to a systematic link between x_i and c_i. A first cursory look also suggests that such a correlation is present in Rose's data set.[5] Indeed, the correlation with c_i is 0.2 or higher (in absolute value) for six out of the nine variables in x_i (excluding exchange rate variability, which is zero when $c_i = 1$, by definition) (see Rose, 2000, Table A4) – 0.2 is a high number with this large amount of data. Another way of expressing the non-random selection, is to compare the means of the variables in x_i across the country pairs with $c_i = 1$ and $c_i = 0$. As Table 1 shows, pairs with a common currency are smaller, poorer, and geographically closer; they more often share a common language, common borders, a common free trade area, a common colonizer, and a common country, and more often involve a previous colonial relationship. Formal tests resoundingly reject equality of means across groups for every variable in the table. We may also note that common-currency country pairs have a slightly smaller bilateral trade.

In summary, the problem seems worth taking seriously.

3 AN ALTERNATIVE STRATEGY

How can we better cope with the possibility of selection on observables and, at the same time, avoid making strong assumptions about functional form? An empirical strategy with precisely these properties is the so-called matching approach for estimating treatment effects on the basis of observational (non-experimental) data. With their roots in medical research, matching methods are becoming increasingly popular with labor economists, who use them for estimating, say, the effect of training programs on income or employment. As far as I know, they have never been used in macroeconomics.[6] Blundell and Costa Dias (2000) provide a useful and easily accessible introduction, whereas Angrist and Kreuger (1999), Heckman, Ichimura and Todd (1998), and Heck-

[5] All empirical estimates in the present paper are based on Rose's original data set, which is made available, in a very user-friendly way, at: http://haas.berkeley.edu/~arose/RecRes.htm

[6] Persson, Tabellini and Trebbi (2000) apply matching methods to estimate the effect of alternative electoral rules on corruption.

man, Lalonde and Smith (1999) contain more technical discussions of matching and its relation to other methods. These papers give references to the relevant statistical literature.

The basic idea in matching is to mimic a controlled (randomized) experiment. In the present context, we thus want to design a control group among the country pairs not receiving the treatment of a common currency ($c_i = 0$) with characteristics matching those that do receive treatment ($c_i = 1$). Clearly, this requires an observable vector of covariates, y_i. To obtain an unbiased measure of the treatment effect, we can appeal to a property known as "conditional (mean) independence", or "selection on observables (only)". Assume that, conditional on the vector y_i, the expected value of trade t_i in the absence of a common currency is the same for treated and untreated country pairs. Under this assumption, it is legitimate to use a fabricated control group for estimating an unobservable counterfactual: the bilateral trade we would (hypothetically) observe in the absence of treatment in those country pairs (actually) treated with a common currency. Now, the plausibility of this identifying assumption clearly depends on what variables enter into y_i. Given that we want to investigate the results in Rose's paper, it is natural to choose $\mathsf{y}_i = \mathsf{x}_i$, i.e., to match on all the variables, except c_i, that enter on the right-hand side in his default specification of the gravity equation.[7] In this way, we directly address the source of the selection problem discussed in the previous section by removing the different composition of x_i in the country pairs with and without common currencies.

With 10 variables in x_i, and the differences across the two groups of country pairs noted in Section 2, direct matching would be too data-hungry a method, however. Finding close matches for 252 treated observations would be impossible, even with 26,336 prospective controls (the numbers refer to the country pairs for which there are no missing data). Some data reduction is required.

Luckily, there exists a simple, yet powerful, way of reducing the dimensionality of the matching problem. Under the conditional independence assumption, we can also match on a function of x_i and, in particular, on the so-called propensity score, a result due to Rosenbaum and Rubin (1983, 1984). In the present context, this score is the (estimated) probability of being treated with a common currency ($c_i = 1$) as a function of trading costs x_i. A more balanced control group is thus found by matching on the propensity score, which we label $p(\mathsf{x}_i)$. Such matching can be done in a number of ways. Each way corresponds to a different estimate of the treatment effect, namely the (average) effect on t_i of having $c_i = 1$ rather than $c_i = 0$. Moreover, these estimates can be obtained non-parametrically, i.e., without any assumption of a particular functional form: we can just compute the mean of t_i across the treated and control groups.[8]

[7] Note that we are then effectively making the same identifying assumption as Rose. Identifying – as he does – γ from the regression:

$$t_i = \boldsymbol{\beta}\mathsf{x}_i + \gamma c_i + \varepsilon_i \ ,$$

requires ε_i to be uncorrelated with x_i.

[8] Deheija and Wahba (1999) apply propensity score estimation to estimate the treatment effect of a training program on the same data set as in the well-known study by Lalonde

Our task ahead thus involves two steps of estimation: first of the propensity score, then of the treatment effect. The following two sections present the results from this two-step exercise.

4 THE PROPENSITY SCORE

I estimate the propensity score by running a linear logit regression of the treatment indicator c_i on nine observable variables. Among the 10 variables entering Rose's default specification, I leave out the measure of exchange rate volatility (see further below). But the remaining nine variables in x_i – output, output per capita, distance, contiguity, common language, common free trade area, same nation, same colonizer, and colonial relationship – are all included. The results from the logit are reported in Table 2. Consistent with non-random selection, the trading costs are strong predictors of a common currency: eight out of nine variables are significant at the 1% level or better. Most of the signs are in line with expectations, namely lower trading costs raise the probability of a common currency between the countries in the pair. An exception is that a free-trade area renders a common currency less likely, thereby suggesting that free trade areas and common currencies might be substitute arrangements for promoting bilateral trade between countries with low inherent trading costs. The fit of the logit appears decent, with a pseudo-R^2 close to 0.5.

It is important to point out, however, that the objective here is **not** to build a statistical (let alone an economic or political) model explaining currency-union membership in the best possible way. For my purposes, it is not a problem if some variables systematically influencing memberships in currency unions, but not trade, are missing in the estimation of the propensity score.[9] In fact, a close to perfect fit of the logit would be destructive, as matching on $p(\mathsf{x}_i)$ requires that we have both treated and untreated country pairs at similar levels of $p(\mathsf{x}_i)$. This is why I exclude Rose's measure of exchange rate volatility: including it would allow me to explain $c_i = 1$ with probability one. From the viewpoint of identifying the treatment effect, additional variables should be added to the logit if they systematically influence bilateral trade. But then, these additional variables should have been included in Rose's default specification of the gravity equation in the first place. If such omitted variables of the trade relation are also correlated with the common-currency indicator, then I certainly have a problem with selection on **un**observables, but so has Rose.[10]

When estimating the treatment effect, I want to make sure that the treatment ($c_i = 1$) and prospective control ($c_i = 0$) groups are comparable; i.e., that

(1986). They show that matching on the propensity score substantially reduces bias due to a non-experimentally defined control group obtained from a set of individuals with very different characteristics than the treated group.

[9] Variables correlated with c_i but not with t_i would, naturally, be good instruments for c_i, if we were to address the alternative problem of omitted variables (selection on unobservables), mentioned at the end of Section 2.

[10] In terms of the regression in Footnote 7, such omitted variables would introduce a correlation between ε_i and c_i, biasing the estimates of γ.

they share a common support for $p(\mathsf{x_i})$ (or $\mathsf{x_i}$). Among the 26,336 pairs without a common currency, 7,499 had an estimated propensity score lower than the lowest score among the treated pairs. These country pairs were thus discarded as non-comparable to any common-currency pair. There was no need for discarding pairs at the upper end of the estimated propensity scores.

Before proceeding to the matching, I also want to verify that conditioning on $p(\mathsf{x_i})$ produces similar results as conditioning on the vector $\mathsf{x_i}$. That is, I ask whether the distribution of $\mathsf{x_i}$ is similar across the treatment and prospective control groups, conditional on the propensity score. To answer that question, I first rank the 18,837 observations on the common support according to their estimated propensity scores. Following the procedure in Dehejia and Wahba (1999), I then group the observations into strata: the first stratum includes observations with the lowest estimated probability of having a common currency $0 < p(\cdot) < 0.1$, the second those where $0.1 < p(\cdot) < 0.25$, whereas the third, fourth and fifth strata each have a width of 0.25.

Based on this grouping, I test for equality of means between the treated and non-treated observations for each of the nine variables in $\mathsf{x_i}$ within each stratum; recall that the hypothesis of common means was rejected for all nine variables in the full sample. In the two uppermost strata, I reject common means (at the 1 % level) only for one out of nine variables, and in the third and fourth strata from the top, I reject for two and four variables, respectively. Observations with a probability below 0.1 of a common currency are more problematic: I reject equal means for all variables except one. But when I reject, the means are typically much closer than in the full sample. While the propensity score estimation does not work perfectly, it still allows me to design control groups whose distribution of $\mathsf{x_i}$ is much more balanced relative to the treated group than the full sample.

I have also performed the same tests based on estimates of the propensity score from a linear probit. The probit produced a relatively similar pattern of point estimates as the logit, but yielded a less balanced distribution between the treated group and the prospective controls.[11]

5 THE TREATMENT EFFECT

In this section, I estimate the treatment effect of a common currency with two different non-parametric estimators. Detailed formulas for these estimators and their (approximated) standard errors are given in Persson, Tabellini and Trebbi (2000, Statistical Appendix).[12] Here, I just describe their properties and report the results.

[11] Results from the probit, as well as the equality-of-means tests are available on my homepage: http://www.iies.su.se/~perssont/.

[12] The standard errors do not take into account the correlation across observations produced by individual countries entering into different country pairs. Neither do they take into account the uncertainty in the estimated propensity score. To address these issues, it would be preferable to calculate the standard errors by a bootstrap procedure.

The first estimator is based on **stratification** and balances the treated and non-treated observations group-wise, within the five strata defined in the previous section. Specifically, it compares **each** common-currency ($c_i = 1$) observation with **all** the non-treated ($c_i = 0$) observations in the same stratum for $p(\mathbf{x}_i)$. Having computed the average difference in bilateral trade t_i within each stratum, the stratification estimator forms a weighted average of these differences, weighing each stratum by its number of treated observations.

The overlap between control and treatment observations varies across the five strata. As expected, we get more treated observations relative to control observations as the estimated propensity score increases. But some overlap of treatment and controls is present in every stratum. Thus, the lowermost bin ($p < 0.1$) includes 75 treated vs. 18,483 controls, the middle bin ($0.25 < p < 0.5$), has 49 treated vs. 120 controls, whereas the uppermost bin ($p > 0.75$) contains 39 treated vs. 4 controls. A small overlap does not bias the estimate of the treatment effect, as long as the two groups are homogeneous enough in terms of the covariates in $\mathbf{x}_i$. But a low number of controls relative to treatments in the higher strata raises the standard error of the estimate.

The next estimator is based on **nearest matching**. Instead of utilizing the full set of controls on the common support (as in stratification), I discard all but the nearest controls and use some controls more than once. Each treated ($c_i = 1$) observation is now matched with only one non-treated ($c_i = 0$) observation, namely the nearest match in terms of propensity score. This way, I obtain 252 pairs, based on 252 treated observations and 181 controls. The nearest matching estimator is just the average difference in bilateral trade t_i across these pairs of treated and control countries.

The rationale for this estimator is to reduce the bias due to different covariates, by finding the nearest match in the control group for every treated observation. If a certain control is the nearest match for more than one treated observation, it is used more than once. While multiple use of certain controls is desirable in terms of reducing bias, it gives less precise estimates, i.e., it increases the standard error. Nearest matching often produces quite intuitive "twins". Thus, many twins with high propensity scores involve two Caribbean country pairs, one where both belong to the East Caribbean Currency Area, another where they do not. For example, among the controls, Antigua and Barbados were about as likely ($p = 0.748$) to have a common currency in 1970 as Grenada and St. Vincent in 1985 ($p = 0.703$). Similarly, many twins with intermediate propensity scores involve West African country pairs, one pair fully in the CFA-zone, the other not: Cameron and Mali 1975 ($p = 0.178$), e.g., are paired with Ivory Coast and Senegal 1970 ($p = 0.178$). An example at the lowermost end is the matching of US and El Salvador 1985 ($p = 0.00052$) with the common-currency pair of USA and Panama 1970 ($p = 0.00055$).[13] Naturally, there are also some non-intuitive matches. But a regression analysis on the whole sample implicitly includes a much greater number of non-intuitive

[13] According to the estimated logit, the recent dollarization of El Salvador was thus as unexpected as the pre-existing dollarization of Panama,

comparisons.

Table 3 reports the treatment effect of a common currency as estimated with these two methods. These should be compared to Rose's point estimate, implying an expansion of trade by 235%.[14] The stratification estimator yields a point estimate of 0.12 – an expansion of trade by a more modest 13%. With the nearest-matching estimator, the point estimate is 0.51, corresponding to a 66% expansion. But the standard errors of these estimates are about double those of Rose's OLS estimate. If we take the nul hypothesis to be a zero effect on trade, the nearest-matching estimate only reaches borderline statistical significance, while the stratification estimate clearly does not. On the other hand, if the nul hypothesis is Rose's estimate, both estimates are significantly smaller.

Should we be concerned about the wider confidence bands? Not necessarily. As already mentioned, the idea behind the non-parametric estimators is precisely to trade off reduced bias due to specification error against less efficiency. In particular, more precise estimates might have to come at the price of more restrictive functional-form assumptions.

Note that the two estimators rely on samples of very different size. In both cases, the number of treatments are 252, which coincides with the number entering in Rose's regression estimates. While the nearest matching method exploits a mere 181 controls, the stratification method relies on a control group which is a hundred times larger. Thus, it is reassuring that the two methods produce relatively close estimates.[15]

6 WHERE'S THE RABBIT?

Why are my results so different from those of Rose (2000)? And do the differences indeed reflect the statistical problems suggested in Section 2: biased estimations due to non-linearities in the relation from measured trading costs to trade and non-random selection into common currencies?

One way of approaching these questions is obviously to extend the regression analysis, searching for direct evidence of important non-linear terms, correlated with common currencies. This is easier said than done, however, because one could imagine an infinite number of non-linear formulations over the ten variables appearing in the basic specification of the gravity equation.

Nevertheless, I have done some experimentation, in order to illustrate how a bias might arise. I first add some quadratic terms to Rose's default specification as estimated on the full (22,948 observation) sample. Two parsimonious spec-

[14] Note that my non-parametric estimators cannot, by definition, give a separate estimate of the effect on trade from the decline in exchange rate volatility. When comparing my results with Rose's estimates, one should thus really add to the latter the effect of eliminated volatility, which (recall Section 1) is on the order of 10%.

[15] As mentioned in Section 4, I have also estimated the propensity score with a probit. Even though these estimates appeared to produce a less balanced control group, I used them to obtain alternative estimates of the treatment effect. The nearest matching and stratification estimates are – 27% and – 18%, respectively, both insignificantly different from a zero (or a modestly positive) treatment effect.

ifications and the associated treatment effects are reported in Table 4 (where again Rose's results have been added for ease of comparison). In column (1), I add interaction terms between, on the one hand, a common language and a free trade area (expected to raise bilateral trade) and, on the other hand, a common colonial history and output (expected to flatten the relation between output and trade). Coefficients on both these terms have the expected sign and are strongly significant. The estimated treatment effect drops, from 235% to about 150%, but still preserves the same order of magnitude.

In column (2), I instead add squared output to the default specification. This produces a more substantial drop in the treatment effect, which is now about 97%, though still highly significant.[16] What is going on here? Figure 1 provides an illustration. The figure shows the partial relation between bilateral trade and output in the full sample. More precisely, it plots output against the residuals from a linear regression of t_i on all variables in x_i except output and c_i. Observations with $c_i = 0$ are indicated by black dots, and those with $c_i = 1$ by ovals. The dashed line shows the estimated linear relation between bilateral trade and output, when I only add output to the underlying regression. Because most of the circles lie above the dashed line, adding also the common-currency indicator to the regression produces the large estimate of the treatment effect found by Rose.

But as the figure illustrates, the output-trade relation appears quite non-linear; in particular, trade is larger than suggested by the estimated linear relation both at low and high levels of output. This is verified by the solid line, showing the estimated trade-output relation when output and output squared are added to the regression underlying the figure.[17] Because we find a larger number of common-currency observations below the solid line than below the dashed line, we get a lower estimate of the treatment effect in column (2). Moreover, as Table 1 showed and Figure 1 illustrates, $c_i = 1$ observations are more concentrated at low levels of output than $c_i = 0$ observations. This correlation between output and the common currency indicator makes the latter pick up the omitted nonlinearity in Rose´s default specification, biasing upwards his estimate of the treatment effect.

The next two columns of Table 4 present further evidence obtained by regression analysis on the matched sample defined by the nearest-neighbor procedure in the previous section. As discussed above, this sample is more balanced between the treated and control groups when it comes to the distribution of covariates. The sample is also more balanced in the number of observations with and without common currencies, meaning that the estimated relations between trade and the components of x_i are not as dominated by the non common-

[16] Rose (2000) indeed adds squared output in one of the regressions (cf. the last column in his Table 6a) and obtains a similar result. In fact, this is the lowest estimate of γ in the whole paper, but Rose does not discuss it further.

[17] The layout of the figure might give the impression that the non-linearity derives only from the $c_i = 1$ observations. This is not the case. When I run Rose's default specification, dropping the c_i dummy but adding squared output, on the full sample, the coefficients on output and squared output (std. errors in brackets) are: -1.09 (.092) and .278 (.0013). If I drop the $c_i = 1$ observations, I obtain the coefficients -0.93 (.082) and 0.254 (.0012).

currency pairs as they are in the full sample.[18]

Column (3) reports the results of running Rose's default specification on the matched sample. Now, the point estimate of the treatment effect is down to 0.52, corresponding to an expansion of trade by 66%. But the standard error is large enough that it is not statistically significant at conventional levels (the p-value is 0.11). Furthermore, the coefficient on output drops from 0.80 to 0.59. As common-currency pairs are smaller, on average, this is consistent with the finding of a flatter relation between output and trade at lower output levels.

In column (4), I once again add squared output to the default specification. The squared output term is still highly significant and the treatment effect is now estimated to 45%, again insignificantly different from zero (p-value 0.25). Interestingly, this parametric estimate is right in between the two non-parametric estimates presented in the previous section.[19]

It should be borne in mind that the results presented in this section are really just illustrations of how a bias might arise. Nevertheless, I believe they support my claim that non-random selection and heterogeneity might explain a large share of the spectacular treatment effect found by Rose.

7 CONCLUSION

I have argued that Rose's (2000) findings of a huge treatment effect of a common currency on bilateral trade are likely to reflect systematic selection into common currencies of country pairs with peculiar characteristics. Using the same data set as Rose, I have provided alternative non-parametric estimates of the treatment effect that should be more robust to selection and non-linearities. These estimates suggest a much more modest expansion of trade: the point estimates are positive but associated with more uncertainty. I have also provided parametric regression estimates which either allow for non-linearities in the specification, or rely on a more balanced sample. The results are similar to the non-parametric estimates and give additional support for my proposed explanation.

My alternative estimates are certainly not the last word. Even if a common currency expands trade only by, say, 40% – a number in the mid-range of my estimates – that is still a very sizeable effect. Additional work on this important issue is badly needed. My findings suggest that such work would benefit from modeling – theoretically and empirically – the selection into common currencies jointly with the effect of those currencies on trade or other variables.

[18] 23 of the 433 observations defined by the matching in the previous section must be discarded for lack of data on exchange rate volatility.

[19] Repeating the same exercise on a matched sample obtained from the probit-generated propensity score, produces estimates very similar to those reported in columns (3) and (4).

REFERENCES

Alesina, A. and R. Barro (2000), "Currency Unions", NBER Working Paper, No. 7927.

Angrist, J. and A. Kreuger (1999), "Empirical Strategies in Labor Economics", in Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics, Vol IIIC*, North-Holland.

Baxter, M. and A. Stockman (1989), "Business Cycles and the Exchange-Rate System", *Journal of Monetary Economics* 23: 377-400.

Blundell, R. and M. Costa Dias (2000), " Evaluation Methods for Non-Experimental Data," mimeo, University College London.

Dehejia, R., and S. Wahba (1999), " Causal Effects in Non-Experimental Studies: Re-evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94: 1053-1062.

Frankel, J and A. Rose (1995), "Empirical Research on Nominal Exchange Rates", Chapter 33 in Grossman, G. and K. Rogoff (eds.), *Handbook of International Economics Vol III*, North-Holland.

Heckman, J., H. Ichimura, and P. Todd (1997), " Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", *Review of Economic Studies* 64: 605-654.

Heckman, J., H. Ichimura, and P. Todd (1998), " Matching as an Econometric Evaluation Estimator", *Review of Economic Studies* 65: 261-294.

Heckman, J., R. Lalonde, and J. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs", in Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics, Vol IIIC*, North-Holland.

Lalonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review* 76: 604-620.

Mc Callum, J. (1996), "National Borders Matter: Canada-U.S. Regional Trade Patterns, *American Economic Review* 85: 615-623.

Obstfeld, M. and K. Rogoff (2000), "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?", NBER Working Paper, No. 7777.

Persson, T., G. Tabellini and F. Trebbi (2000), " Electoral Rules and Corruption", mimeo, Institute for International Economic Studies.

Rose, A. (2000), "One Money, One Market: The Effect of Common Currencies on Trade", *Economic Policy* 30: 7-46.

Rose, A. and E. van Wincoop (2001), "National Money as a Barrier to Trade: The Real Case for Currency Union", forthcoming in *American Economic Review, Papers and Proceedings*.

Rosenbaum, P. and D. Rubin (1983), " The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70: 41-55.

Rosenbaum, P. and D. Rubin (1984), " Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association* 79: 516-524.

Table 1
Distribution of covariates

	Non common-currency pairs					Common-currency pairs				
Variable	Obs	Mean	Std.Dev	Min	Max	Obs	Mean	Std.Dev	Min	Max
Output	26356	34.428	2.679	20.026	43.526	252	28.866	3.964	20.326	37.991
Output/per capita	26356	16.238	1.366	11.728	20.805	252	15.248	1.641	12.280	19.355
Distance	26356	8.201	0.793	3.991	9.422	252	6.469	1.387	2.967	9.258
Contiguity	26356	0.025	0.155	0	1	252	0.111	0.315	0	1
Language	26356	0.136	0.343	0	1	252	0.806	0.397	0	1
Free trade area	26356	0.017	0.131	0	1	252	0.298	0.458	0	1
Same country	26356	0.001	0.034	0	1	252	0.143	0.351	0	1
Same colonizer	26356	0.081	0.273	0	1	252	0.710	0.455	0	1
Colonial relation	26356	0.013	0.115	0	1	252	0.028	0.165	0	1
Trade	26356	9.413	3.307	0.132	19.367	252	8.609	2.996	1.742	16.872

Table 2
Logit estimates of propensity score

Output	- 0.240
	(.033)
Output/per capita	- 0.168
	(.058)
Distance	- 1.016
	(.088)
Continguity	- 0.390
	(.278)
Language	1.743
	(.208)
Free trade area	- 1.431
	(.292)
Same nation	6.246
	(.546)
Same colonizer	1.401
	(.203)
Colonial relation	- 1.817
	(.695)
# Obs.	26,607
Pseudo R^2	0.489

Standard errors in brackets

Table 3
Non-parametric estimates of treatment effect

	Rose	Stratification	Nearest Matching
Currency union	1.221 (.142)	0.123 (.254)	0.506 (.257)
% expansion of trade	235	13	66
# Obs.	22,948	18,837	433
Treated	252	252	252
Controls	22,696	18,585	181

Standard errors in brackets

Table 4
Regression estimates of treatment effect

	Rose	(1)	(2)	(3)	(4)
Currency union	1.221 (.142)	0.927 (.150)	0.685 (.147)	0.519 (.320)	0.370 (.320)
% expansion of trade	235	150	97	66	44
Output	0.803 (.006)	0.819 (.006)	- 0.868 (.086)	0.586 (.075)	- 1.009 (0.624)
Free trade area × Language		0.506 (.158)			
Same colonizer × Output		- 0.138 (.020)			
Output × Output			0.024 (.001)		0.027 (.009)
# Obs.	22,948	22,948	22,948	410	410
R^2	0.629	0.630	0.636	0.525	0.544

Robust standard errors in brackets
Columns (1) and (2) estimated by OLS, columns (3) and (4) by WLS (observations among controls are weighted by the number of times they are used in matching).