Monte Carlo Appraisals of Gravity Model Specifications

Introduction

We have argued that both theory and evidence suggest that history plays a role in shaping the direction of international trade. The standard gravity-model formulation, which neglects the role of historical factors, suffers from omitted-variable bias...The implication is that we will never run another gravity equation that excludes lagged trade flows. If our paper is successful (and widely read), neither will other investigators.

Barry Eichengreen and Douglas A. Irwin (1996), pp. 55-56.

...the empirical application of the gravity model is still rather basic. As demonstrated by Cheng and Wall (1999), though providing a high R^2, the standard estimation method tends to underestimate trade between high-volume traders, and overestimate it between low-volume traders. They attribute this to heterogeneity bias...Their fixed-effects method, which I will use in this study...allows(s) for trading-pair heterogeneity and (is) statistically superior to the standard model.

Howard Wall (1999), pp. 35, 40

...I find a large positive effect of a currency union on international trade, and a small negative effect of exchange rate volatility, even after controlling for a host of features...These effects are statistically significant and imply that two countries that share the same currency trade three times as much as they would with different currencies. Currency unions like EMU may thus lead to a large increase in international trade, with all that entails.

Andrew K. Rose (2000), p. 7.

It has become increasingly common to analyze data on international trade flows with the help of gravity models. These models are simple in structure, fit the data well, and are in principle consistent with a wide range of theoretical underpinnings (Deardorff, 1998). The gravity model has been characterized as "a very simple model that explains the size of international trade between countries with a remarkably consistent (and thus, for economics, unusual) history of success as an empirical tool." (Rose, 2000, 14) Gravity models have provided "some of the clearest and most robust empirical findings in economics." (Leamer and Levinsohn, 1995)

Each of this paper's introductory quotations is a strong statement from a widely cited paper. While the first two are about issues of econometric hygiene (specification) and the third addresses an issue of public policy, all three claim that some omission of variables has important implications for the interpretation of results. Eichengreen and Irwin (1998) raise the specter that any regressions which ignore the effect of history by omitting a lagged dependent variable may be misspecified, with the result that policy conclusions drawn from such regressions may be misleading. Cheng and Wall (1999) state that the omission of fixed effects may lead to a pattern of over- and under-estimates of trade flows. Rose (2000) found a high effect of currency unions on trade, enough to imply a tripling of trade flows under a common currency. Rose and Stanley (2005), in a review of 34 recent studies, offer a more conservative consensus estimate of between a 30 to 90 percent increase in trade. Even this more modest result has important practical implications, as currency unions represent a policy option that is in some places being pursued with a great deal of public discussion.

This paper argues that many previous findings about gravity models suffer from a common problem which has not previously been addressed: The data may be plentiful and rich to a fault. Because there are annual (or higher-frequency) observations on many pairs of bilateral trading partners, sometimes for many product categories, it is not uncommon for gravity modelers to operate with thousands or tens of thousands of degrees of freedom in identifying a very few parameters.[1] This makes it relatively easy to obtain statistical verification (i.e. high t-statistics for individual variables or F-statistics for sets of variables) whenever the model specification is elaborated in any way. In turn, this easy verification creates incentives to elaborate the model further.[2]

Given their early general success, gravity models have been required to answer ever more specific policy questions. For example, to assess the effects of removing an economic sanction, a fair degree of precision in the estimate of a residual or an out-of-sample projection may be required. To estimate the effects of international borders, non-tariff barriers, or new trade agreements, precision in the estimate of a dummy variable may be wanted. Identification issues also became important for policy purposes. For example, because countries that join trade agreements are usually near each other, it is important but difficult to separate the effect of distance from the effect of the trade agreement.

Thus, a variety of refined gravity-model specifications have been proposed. Moving from cross-sectional data to time-series panels has allowed the use of a lagged dependent variable, country fixed effects for exporters and/or importers, log-first-differences of variables, and estimations of time-varying regression parameters. The proliferation of specifications has unfortunately not been met by a rapid sifting of the possibilities; there is a great deal of controversy

---

[1]For a panel of bilateral trade between n countries with observations for t time periods, there are $t(n^2\text{-}n)$ available observations. By the standards of gravity modeling, (n=35, t=10) is a relatively modest panel, and contains 11,900 observations. Feasible panels on the order of (n=100, t=30), which yield 297,000 observations, have been available in electronic form from a variety of sources (IMF Direction-of-Trade Statistics, UN COMTRADE, Statistics Canada) for about a decade now.

[2] It is ironic that the original appeal of the gravity model lay in the fact that it explained a high degree of variation in the data with a relatively small number of parameters. The original double-log gravity models of Tinbergen (1962) and Pöyhönen (1963) in fact required only three variables to explain trade flows, which were further simplified into only two regressors by multiplying together exporter's GDP and importer's GDP for a simple activity variable. Economic distance (usually measured as shipping distance) was the second regressor. This structure could only have been simplified further by dispensing with regression altogether, constraining the coefficients to equal those implied by the physical-science model of Newtonian gravity (1 for activity, -2 for distance). Actually letting the regression pick the coefficients was sufficient to cause the amazing goodness-of-fit, with two regressors often explaining 50-70 percent of the variation in large datasets. This is what attracted researchers to the methodology in the first place.

over the "correct" specification of the gravity model. There is also debate about just which things the model is able to discern, and where its limitations lie. Given the very large volume of data available to gravity modelers, many different specifications of the gravity equation can be made to "work" in the sense that they generate statistical tests that tend to corroborate the specification's validity.3

While some researchers have confidence in gravity models, others suggest their strong statistical results may be the result of poor measurement or incorrect specification of the model. In an application of Leamer's extreme bounds analysis, Ghosh and Yamarik (2004) raise doubts that much can be known with confidence from the application of gravity models, given the proliferation of proposed specifications. (The authors mention as many as eighteen candidate variables that have been used in the literature, without exhausting them all.)4 This literature has by no means reached a consensus, and many versions of the model currently coexist.

Method and Data

Monte Carlo simulations were devised for just such agnostic moments. They allow the researcher to create hypothetical but reasonable data that emerge from a *known* data-generating process, and then to evaluate which models most nearly and frequently draw correct conclusions about the underlying structure of the data. Monte Carlo techniques have been profitably employed in examining the relationship between trade and investment (Keller (1998)), trade patterns and technology flows (Keller (2000)), and have been used fruitfully to explore what underlies current-account fluctuations (Nason and Rogers (2006).[5] Surprisingly, this approach has yet to be used to help sort out the merits of alternative gravity-model specifications. We find that some common gravity-model findings should be taken with a large grain of salt. They may simply reflect correlations between variables included in the estimated model and those included in the "right" specification, that is, included in the data generating process. We use Monte Carlo simulations to evaluate a number of common gravity-model specifications:

A. We begin with panel data on trade flows and trade-related, country-pair-specific attributes provided by Andrew Rose (Rose, 2000; data available at http://faculty.haas.berkeley.edu/arose/RecRes.htm[6]). The

[3] One might ask if there is therefore an optimal size of series when applying gravity models. Perhaps some data should be discarded to reduce the likelihood of drawing unfounded conclusions. But the optimal amount of data must always be "as much as is available." More is always better, because it is impossible to state in advance what criteria might be used to exclude data without biasing results. This is why we take the approach we do: knowing that more data is always better, but that it also has the potential to allow spurious results, we try to identify procedures that minimize the likelihood of such spurious results.

[4] This particular literature points out a limit of Leamer's approach. In effect, the whole literature has already taken Leamer's approach, trying a number of different specifications and finding that they all can be statistically justified. Following Leamer, at this point we would need to say that there is just no way of knowing the truth in this literature. Our paper is, in effect, an option to pursue when the Leamer approach leads to a standoff. We construct multiple virtual realities in which we know what the truth is, then ask which model specification most consistently reports the truth we already know.

[5] For a more general discussion of the use of Monte-Carlo methods see Kleijnen (2004)

[6] Though our estimations of Rose's model (Scenarios Eleven and Twelve below) are very close to those he reports, we are unable to exactly replicate his results. Apparently this is because he includes "year controls" in his pooled regressions; their coefficients are not reported. A few minor

data comprise 33,903 bilateral trade observations for the years 1970, 1975, 1980, 1985, and 1990, for the 186 countries and other entities for which the United Nations Statistical Office reports trade data. Rose estimates that these data cover 98% of all trade (Rose, 11).[7]

B. We estimate the specifications of the gravity model that we judge to be either most common or most at issue, which we number throughout the paper according to the following scheme:
   1. A simple gravity model (which we call "the Newtonian model"): logged two-way trade as a function of logged product of GDPs, logged product of per-capita GDPs, logged distance, the standard deviation of changes in the relative exchange rate, and dummy variables for common languages, contiguous borders, having a common colonizer, being part of the same country (as with overseas departments of France), and for cases in which one trade partner colonized the other. This is perhaps the most common extant gravity model.
   2. A trade-pair fixed-effects process, removing the time-invariant variables. This is similar to the model preferred by Howard Wall, whom we quote in the introduction.
   3. A lagged dependent variable model, which simply adds a lagged dependent variable to the original Newtonian specification. With this model we aim to assess Eichengreen and Irwin's conclusion that they (and we) should never again run a gravity model that does not include a lagged dependent variable.
   4. A model supplementing the Newtonian model with a dummy variable indicating if the trade partners are members of the same free-trade association. (No lagged dependent variable is included.) By comparing these results to a simulation that includes a lagged dependent variable, we can test Eichengreen and Irwin's proposition that FTA variables only appear to be significant because they proxy an omitted lagged dependent variable.
   5. A model with the FTA dummy and a common-currency-union dummy. (No lagged dependent variable is included.) This model is used by Rose to reach his provocative conclusion about the large trade-creating effects of currency unions.

To summarize the models, we have the following chart:

Model Number	Model name	Independent variables
01	Newtonian	Ln(GDPi*GDPj), Ln((GDPi/Popi)*(GDPj/Popj)), Ln(Dist), SD(XΔ), common language, common colonizer, same country, contiguous border, colonial relationship

---

data-set errors were reported by reviewers after publication, all involving the language-variable coding of several observations for Switzerland and Belgium. Since our aim is Monte Carlo simulations, it is a matter of indifference whether the original data or the corrected data are employed; we have used the original data for consistency with the published literature.

[7] Trade data come from the World Trade Database. Population and real-GDP data are from the Penn World Table 5.6, filled in with World Bank World Development Indicator data where there were gaps. Data on location/distance, official language, colonial background, and related indicators came from the CIA's web site. The FTA variable was constructed from the WTO's web site.

02	Fixed Effects (FE)	Newtonian, plus trade-pair fixed effects, minus last five Newtonian variables
03	Lagged Dependent Variable (LDV)	Newtonian, plus lagged logged two-way trade
04	Free Trade Association (FTA)	Newtonian, plus FTA dummy
05	Common Currency Union (CU)	Newtonian, plus FTA dummy and common CU dummy

Appendix One details the definition of each variable. For each of these five models, we estimate the parameters of the model with an appropriate panel-data regression program.

C. Our aim is to conduct a series of pair-wise comparisons among these five models. In each comparison, one model serves as the "true," maintained data-generating process, and the other relies upon that process for its dependent variable. We proceed as follows: Using the regression results for the maintained "true" model, we generate a full panel of predicted values for the left-hand-side variable in that model. We perturb each predicted observation with a zero-mean, normally distributed random number generator (with standard error set equal to the estimated root mean square error from the original regression) to form a Monte Carlo data set. For each pair-wise comparison, we repeat this perturbation process 250 times to generate 250 Monte Carlo data sets.[8] These 250-observation panels would be reasonable observations of trade flows if the maintained "true" specification of the gravity model accurately explained world trade.

D. We then use each 250-observation batch of Monte Carlo datasets to estimate parameters for a competing gravity model specification. If the competing model identifies a statistically significant explanatory variable that does not appear in the maintained true model, we have a "false positive;" the competing model has yielded an incorrect (though statistically justified) explanation of trade flows. If the competing model fails to ascribe statistical significance to a variable that *is* significant in the maintained true model, we have a "false negative."

This approach allows us to evaluate some of the questions raised in the literature. (For example, do we detect a lagged-dependent-variable coefficient when it's not actually present in the data-generating process, or do some specifications fail to detect it when it is present? Can the effects of FTAs be distinguished from those of distance?) Our aim is to offer provisional conclusions about the relative strengths and weaknesses of the various specifications of gravity models, and if possible to suggest which specifications may most frequently lead to correct judgments about the actual (unknown) data-generating process. In the Conclusions section we will gather together these findings to offer some general comments about the ability of gravity models to draw inferences

---

[8] Beginning with simulations involving 1250 iterations, we found that gradually reducing the number of datasets to 120 had virtually no effect on any of the results. We use 250 iterations to stay on the safe side.

about trade flows. We also offer some best-practice advice about modeling trade flows with gravity models.

Results

We present our results as a series of twelve scenarios, each pairing a particular data-generating process with a competing (mistaken) specification of the gravity model.[9] Following our model-numbering scheme in the last section of the paper, our naming convention in presenting the results is to refer to *Scenario 0x0y* as the scenario in which *specification 0x* is the true data-generating process and *specification 0y* is the Monte Carlo estimation model. For example, *Scenario 0102* is the case in which the data are generated by the Newtonian model, but we use those data to estimate a fixed effects model.

Our results are also summarized in Appendix One in five tables—one table for each maintained data-generating process. Each table reports all of the Monte Carlo simulations that relied upon the same data-generating process. The tables report the resulting mean parameter estimates, the standard error of the 250 estimates of each parameter, and the means of the F-statistic, R-squared, and standard error of regression. We also report probability values for standard tests of significance of the parameter estimates, along with probability values for tests for differences between true data-generating parameters and the Monte Carlo estimates of these parameters.[10]

Scenario One: Trade is generated by a Newtonian process, but one instead estimates a trade-pair fixed-effects model. We estimate this in two ways: Scenario 0102 compares the full Newtonian model with a model that instead estimates country fixed-effects. Scenario 0102*Lite* compares the fixed-effects model to a Newtonian data-generating model that does not include time-invariant variables (distance, contiguous border, common language, common country, common colonizer, colonizer-colony relationship). These time-invariant variables drop out of fixed-effects estimations, since the fixed-effect coefficient measures the influence of everything in the situation that does not change over time, and by comparing the fixed-effects results to both versions of the Newtonian model we can obtain a fair comparison of the two models while also exploring the possibility of excluded-variables misspecification bias in the smaller Newtonian model. The results are found in Table One Column Three, and Table One(Lite) Column Three.

All variables are highly significant in both the full Newtonian data-generating model and the FE estimations. The FE model's slope estimates accurately measure the true parameters, though the FE constant is of course significantly different than in the full Newtonian model. For example, the log product of the real GDPs (lrgdp) has a coefficient of .785 in the data-generating-process, and a coefficient of .805 in the FE model. Both are highly significant. The numbers along the right-hand margin of column three test the null hypothesis

[9] The Stata programs by which our results were generated are available from the authors.

[10] In the Monte Carlo simulation result columns of our tables, one hopes for P-values near one, not zero, indicating that a mis-specified model has nonetheless correctly estimated the data-generating-process parameter, since the null hypothesis underlying these P-values holds that the Monte Carlo simulation is telling the truth about the data-generating process.

that the coefficient in the FE model equals .785 The good news here (and for the other variables in the FE model) is that the p-value is quite high, .79 in this case, and we fail to reject the null.

Note also the low R-squared in the simulation (8.9 %), despite its respectable F-statistic. Finally, notice that there is very little difference between the "regular" and "Lite" estimates for Model 1, so missing-variable misspecification does not seem to be a problem for the smaller Newtonian model. In sum, the fixed-effects specification cannot directly measure the effects of the categorical variables11, but appears to perform reasonably well otherwise.

Scenario Two: Trade is generated by a country-pair fixed-effects process, but one instead estimates the Newtonian model (Scenarios 0201 and 0201*Lite*, reported in Table Two, Columns Two, Three and Four).

Again, all slopes in the FE data-generating regression and the Newtonian estimations are highly significant. Regarding the data-generating regression (Column Two), note again the very low R-squared in the FE models (4.1 %), despite good F-statistics.[12]

In Scenario 0201 (Column Three), the Newtonian model is far afield of the underlying FE model's coefficients. When a FE data-generating process is estimated by a standard Newtonian model, the results are apparently unreliable. For example, in the data-generating process the coefficient on exchange rate volatility ("sdd," a reference to the standard deviation of the dispersion of the exchange rate) is -0.012, and the Newtonian model yields an estimate wide from the mark at -0.043. A researcher might be inclined to accept its sdd coefficient because of its statistical significance, though the p-value (of 0.0) in the right hand margin of column 3 reveals that the estimate is statistically significantly different from the "true" data-generating-process value. Indeed, all of the p-values in the right hand margins of columns three and four are approximately 0, and the

---

[11] These are also known as "dummy" or "indicator" variables.

[12] Note also that in the data-generating regression results (Column Two) the coefficient on GDP in the FE regressions has a negative (i.e., the wrong) sign. We suspected that this was a complicated artifact of including both Real GDP and Per-Capita Real GDP in the same fixed-effects estimation, which we do throughout the paper in order to be consistent with Rose's original use of the data. It appears our suspicions are correct. When per-capita real GDP is dropped from the data-generating process, real GDP receives a positive coefficient. We removed per-capita real GDP from all of the data-generating processes and Monte Carlo simulations in this paper, and found that none of the results important for our conclusions were affected; there were no changes in the interpretations of the MC simulations. For completeness, we note the following effects when real per-capita GDP is dropped from the regressions and simulations:

Table 1: In the data-generating process, *contiguous border* and *common colonizer* shrink; *contiguous border* also becomes insignificant.

Table 3: The Newtonian and Newtonian-with-FTA MCs capture the *contiguous border* and *common colonizer* effects correctly now. The false-positive finding of the coefficient on FTA (column 5) is twice as large now, going from 0.70 to 1.4.

Table 4: The LDV model (column 4) does a worse job of estimating the *distance* coefficient than before; there's now a significant difference from the data-generating process coefficient. In the data generating process, *contiguous border* and *common colonizer* shrink; *contiguous border* also becomes insignificant.

Table 5: In the data generating process, *contiguous border* and *common colonizer* shrink, and both become insignificant; the coefficient on FTA doubles from estimations that included GDP per capita.

difference between the coefficient estimates and the coefficients in the data generating process are often substantial.

One might infer (from Tables One, One (Lite), and Two) that fixed effects should always be included in gravity models. The FE model appears to be capable of approximating the true data-generating process (except for its categorical variables) when fixed effects are not present, and the Newtonian OLS model is incapable of approximating the true data-generating process when fixed effects are present. This conclusion, however, depends upon whether the categorical variables are of independent interest; they cannot be obtained in the FE specification.

Another striking aspect of the Newtonian and LDV estimations in Scenarios 0201 and 0203 (Table Two, Columns Three and Five) is that each of them contains one or more "false positives." Distance and common colonizer, for example, are absent from the fixed-effects model's data-generating process, but both the Newtonian and the LDV models find a high degree of false statistical significance. As we'll see below, this is not the last simulation to suggest that adding variables to otherwise robust specifications can yield false positives.

Finally, one additional big-picture comment that could be drawn from these first scenarios raises questions about the value of traditional t-tests of significance. There is little relationship in these simulations between the traditional t-statistic and the ability of a model to precisely measure the actual population value of the parameter.

Scenario Three: Trade data are generated by the full Newtonian model, but then modeled by a lagged-dependent variable model. (Scenario 0103, in Table 1 Column 4) This and the next scenario allow us to explore whether hysteresis[13] might appear to be significant when it is not present in the data-generating process, and to see if parameters are mismeasured when real hysteresis is ignored. (Note that our LDV tables report both the "raw" or "impact" coefficients and the long-run coefficients[14])

All coefficients are highly significant, except for the LDV coefficient in the simulation; thus there are no false negatives or false positives. All coefficients are also statistically identical to the true data-generating coefficients. A lagged-dependent-variable model applied to a Newtonian world appears to measure coefficients correctly. Even the R-squared statistics are approximately the same.

Scenario Four: Trade is generated by a lagged-dependent variable process, but estimated with the full Newtonian model. (Scenario 0301, reported in Table Three, Column Three)

Consider Column Three of Table Three.[15] In all cases the simulation coefficients are statistically significant, but also significantly different from the

[13] "Hysteresis" refers to the dependence of the state of a system on the history of its state, the lagging of an effect behind its cause.

[14] See Eichengreen and Irwin (1996) for a discussion of the distinction between impact and long-run coefficients in a gravity model.

[15] The long-run effect of an independent variable in the LDV model is calculated by dividing the coefficient by one-minus-the-coefficient on the LDV variable. These results, reported in Appendix Table Three, Column Two, are the coefficients that should be compared to those of the other models, and the t-statistics and probability values in Columns Three through Five report tests of Monte Carlo estimates relative to the long-run data-generating-process coefficients.

long-run effects in the data-generating process (reported in Column Two). All of the p-values reported in the right hand margin of column three are zero to two decimal places. Some of the coefficients in the Newtonian model are far different from the LDV data-generating process. For example, the coefficient on whether the pair is included in the same nation (comctry) is estimated at 1.737 in the Newonian model, but in fact equals 7.786 in the data-generating process. In sum, the Newtonian model's results do not come close to the LDV's data-generating process.

Scenario Five: Trade is generated by a country fixed-effects process, then modeled ignoring fixed effects but including a lagged dependent variable. (Scenario 0203, Table Two, Column Five) This and the next scenario allow us to see if hysteresis can be distinguished from, or substituted for, trade-pair fixed effects.

GDP, per-capita GDP and the standard deviation of the exchange rate are all significant in all the regressions, but the LDV simulation gives estimates significantly—strikingly–different from the actual data-generating process. The simulation goes on to report four false-positives: distance, common colonizer, colony of the trade partner, and LDV. The estimated coefficients are relatively large, and thus potentially significant for policy, especially in the latter three cases. Colonial relationship, for example, is not in the fixed-effects data-generating-process, but still returns a long-run coefficient of 4.8, with a p-value of 0.00 If country fixed effects exist but are incorrectly estimated with a LDV model, the influence of GDP, per-capita GDP, common colonizer, mutual-colony status, and distance are all exaggerated. These results certainly give reason for some caution about routinely including lagged dependent variables in gravity models.

Scenario Six: Trade is generated by a lagged dependent variable process, but modeled as trade-pair fixed effects with no LDV. (Scenario 0302, Table Three, Column Four[16])

The Monte Carlo FE simulation in this case reports a very poor R-squared (less than 4 percent!), along with coefficients that are all statistically significant but also significantly different from the true parameter values. The errors in the GDP and per-capita GDP coefficients are striking. These results encourage caution that must temper our positive results for fixed-effects specifications in Scenarios One and Two. Taken with Scenario Five, these results also suggest an interpretive quandary: it appears that fixed effects are not easily distinguished from hysteresis.[17]

Scenario Seven: Trade is generated by the simple Newtonian model with an added LDV, but estimated with an FTA model (Scenario 0304, Table

[16] In the initial data-generating regression using world trade data, the LDV must bear the influences of the non-time-varying coefficients that are excluded from the regression. Thus six variables bear statistically insignificant coefficients.

[17] Here is another, more positive interpretation: since gravity models routinely have outstanding fits, fixed-effects models among others, perhaps the under-4-percent R-squared finding in column 4 can be taken as evidence that researchers are seldom in the position of estimating an LDV world with an FE model.

Three, Column Five). In this and the next scenario, we consider the Eichengreen/Irwin-inspired literature that asks if apparent FTA effects might actually be caused by hysteresis.

There is evidence that Eichengreen and Irwin may be correct. The dummy variable for FTA membership shows a false positive relationship with the value of trade; though the variable is not in the data-generating process, the coefficient is .705 and is statistically significant in the FTA model.

More broadly, all of the coefficients in the FTA model are statistically significant, but significantly different from the coefficients in the true data generating process, and the differences are striking. The long-run coefficient on sdd (the measure of exchange rate variability) is -.19, but the FTA model returns an estimate of -.026. Once again, it's interesting to note the limits of statistical significance. The -.026 has a p-value of 0.00 in the FE model, despite missing the population parameter (from the LDV data-generating process) by a factor of more than 3. In sum, if a LDV process is modeled by a FTA regression, it appears that the results are not reliable.

Scenario Eight: Trade is generated by a process in which FTAs are significant, but this process is modeled by a lagged-dependent variable regression without an FTA variable. (Scenario 0403, Table Four, Column Four)

In this case, there is no false positive on the LDV variable; the coefficient on the LDV variable is statistically insignificant. In addition, all coefficients are significant in both regressions, the R-squares are similar, and no LDV-equation coefficients are statistically different from those in the data generating process. If an FTA process is modeled by an LDV regression, there appears to be no significant misrepresentation of the data-generating process. In effect, Scenario 0403 confirms Scenario 0103. Inclusion or exclusion of the additional LDV variable is unlikely to cause spurious findings of significant lags, and unlikely to result in mismeasurement of the other coefficients. Similarly, 0401 and 0104 tend to confirm each other – the coefficients on other variables are robust to the inclusion or exclusion of the FTA variable (with the exception of the Distance coefficient in Simulation 0401).

Scenario Nine: Trade is generated by the full Newtonian model, but modeled with an FTA regression. (Scenario 0104, Table One, Column Five) This scenario and the next allow us to consider whether the influence of FTAs can be distinguished from influences of time-invariant factors like distance.[18]

All coefficients are significant in all regressions, and none in the simulation are statistically different from those in the data generating process. In addition, the coefficient on the FTA variable is insignificant (with a p-value of .47). It appears that no harm is done in this case by including an irrelevant FTA variable.

Scenario Ten: FTAs influence the data-generating process, but it is modeled by the full Newtonian model without an FTA variable. (Scenario 0401, Table Four, Column Three)

[18] This provides evidence on whether FTAs cause, or result from, high levels of trade. Put differently, do FTAs generate trade, or are FTAs formed where trade already flourishes due to of proximity?

All coefficients are significant and not significantly different from the true data-generating process, except that the coefficient on distance is overestimated in the Newtonian model (equaling -1.104 in the FTA model but estimated at -1.149 in the Newtonian). The last several scenarios give some confidence that the FTA variable measures what it claims to measure, and that the influence of FTAs *can* be distinguished from the effects of distance.

Scenario Eleven: Trade is generated by a process influenced by FTAs, but estimated with a model that includes both FTAs and currency unions. (Scenario 0405, Table Four, Column Five) This and the next scenario allow us to consider the potential for measuring independent currency-union effects when they exist in the data-generating process, and rejecting their influence when they do not exist. (e.g., Rose, 2000)[19]

With currency unions absent from the data-generating process, the CU estimations yield a false positive for currency unions' effects, with a relatively large (at .903) and statistically-significant coefficient. None of the other simulation coefficients are significantly different from the data-generating process. If an irrelevant CU variable is included in estimations, it appears that none of the other coefficients become biased, but there is risk of a false positive on the CU variable.

Scenario Twelve: Trade is generated by a process influenced by both FTAs and currency unions, but then modeled by a regression that does not include currency unions. (Scenario 0504, Table Five, Column Three)

All coefficients are significant in the non-CU regression, and none in that regression are significantly different from those in the data-generating process. The unmodeled influence of CUs seems to be assigned primarily to the Common Country and FTA coefficients, as only those coefficients are unlike those in the data-generating process. Even these differences are not statistically significant. Taken with the last scenario, this counsels some caution: If CUs affect trade but remain unmodeled, the resulting coefficients are not biased, but if an irrelevant CU variable is included there is risk of measuring a false positive. There may yet be cause for some agnosticism about the importance of CUs.

By now false positives have been observed in several scenarios. This phenomenon might be expected when sample sizes are large and there is significant multicollinearity among independent variables, and might partially explain the results of Scenarios Eleven and Twelve. In Table Six we explore this

[19] We believe we used the same data, model and programming language as Rose (2000), but were unable to exactly replicate his results. Fortunately exact replication is not necessary for the purposes of our paper. Here we report the names of the model parameters, followed by our parameter estimates (from our Table Five, Column Two, starred when our estimate is 5%-statistically-significantly different from the Rose estimate), followed by Rose's parameter estimates (Rose (2000), Table One, "Pooled" column):

Log-Real GDP:	0.790, 0.80.	Log-Real GDP per capita:	0.584*, 0.66.
Log-Distance:	-1.089, -1.09.	Contiguous Border dummy:	0.594, 0.53.
Common Language dummy:	0.447, 0.40.	Common Country dummy:	1.398, 1.29.
Common Colonizer dummy:	0.511, 0.63.	Colonial Relationship dummy:	2.151, 2.20.
Std. Dev. of Exchange Rate:	-0.043*, -0.017.	FTA dummy:	0.837, 0.99.
CU dummy:	0.903*, 1.21.	N:	22,948, 22,948.
R^2:	0.605, 0.63.	RMSE:	2.093, 2.02.

potential explanation. When the CU variable is treated as a dependent variable and regressed against the remaining independent variables, all of the coefficients have very high statistical significance; seven of the eleven coefficients have double-digit t-statistics.[20] The largest coefficients are associated with the Common Country and FTA variables, consistent with the observation in Scenario Twelve that these two coefficients were the most affected by the exclusion of the CU variable from the simulation regression. Thus a falsely-positive CU effect appears to emerge because of multicollinearity combined with a large number of degrees of freedom.

Conclusions

The gravity model's popularity is based on its strong empirical performance. Indeed its basic architecture is strong—GDP, distance, and other traditional variables perform well throughout our scenarios. But our simulations suggest several cautionary tales that should be explored further using a variety of data sets and methodological approaches.

First, the simulations suggest that there is some wisdom in routinely including either a lagged dependent variable or fixed effects when estimating gravity models, though two cautions must immediately be added: 1. When the data process is driven by a lagged dependent variable rather than fixed effects, a fixed effects estimation appears to go awry, and the converse can be said when FE processes are modeled by LDV estimations. It does not appear to be possible to distinguish fixed effects from hysteresis in traditional gravity modeling—a serious problem, since the investigator never knows in advance which influence is present in the data. Researchers who present the results of either model may owe their readers a presentation of the other for comparison. 2. Fixed effects are sometimes included in gravity equations to account for idiosyncratic country-pair effects on trade that are otherwise not measurable. Unfortunately, this practice also excludes all measurable but time-invariant influences. When fixed effects are present in the data generating process and other categorical variables also influence trade, it appears that traditional gravity modeling may not be a reliable method for measuring the categorical variables' influence, since the simple Newtonian model and the LDV model both returned false positives when fixed effects enter the data generating process.

Our simulations also indicate that FTA effects can successfully be separated from the effects of distance, but again a caution must be sounded: Apparent FTA effects can be due to hysteresis, and apparent hysteresis may in fact reflect an FTA effect. As with FE and LDV models, the researcher faces difficulties in distinguishing LDV and FTA effects with traditional gravity equations.

Our simulations also indicate that apparent currency union effects may be the result of the combined influence of multicollinearity and a generous number of degrees of freedom. It appears that no serious damage is done to other coefficient estimates when CUs do influence trade but the CU variable is deleted from the estimation.

[20] Our principal-components analysis of the same variables confirms that only four eigenvectors have eigenvalues greater than one, and that CU has a relatively large eigenvector coefficient in three of these eigenvectors.

Finally, our simulations raise questions about the value of traditional t-tests in gravity modeling with large data sets, as a number of false positives and false negatives were reported throughout the results. It has become increasingly popular to include fixed effects while omitting other traditional variables so that some new variable of interest can be investigated, under the assumption that the fixed effects "soak up" the effects of the deleted traditional variables. However, as fewer traditional variables are included, it appears to become more likely that any newly-introduced variable will prove (falsely) significant.

As Monte Carlo simulations, there is no guarantee that our results will be supported by other datasets, nor indeed by the addition of variables to the present data set. But we are hopeful that as others produce simulations involving a variety of data sources and variables, researchers will gradually reach a consensus that encourages some approaches and cautions away from others.

Appendix One: Simulation Results

Variable definitions:

lrgdp:	log of product of country real GDPs
lrgdppc:	log of ((product of country real GDPs)/(product of country populations))
ldist:	log of distance between the trade pair
contig:	binary variable, =1 if pair shares a contiguous border
comlang:	binary variable, =1 if pair shares a common language
comctry:	binary variable, =1 if pair is part of the same nation (e.g., France and overseas departments)
comcol:	binary variable, =1 if pair were colonies after 1945 with same colonizer
colonial:	binary variable, =1 if one in pair was colonized by the other
sdd:	standard deviation of change in bilateral exchange rate
FTA:	binary variable, =1 if pair shares a FTA
CU:	binary variable, =1 if pair shares a currency union agreement

Table One: Newtonian Data-Generating Process (DGP)

Column two shows the parameters that generate the data.
Columns three through five show the simulation estimates generated from these data.
Numbers in **bold** in columns three through five show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)* **Newtonian DGP**	*(3)* 0102: Fixed effects estimation		*(4)* 0103: LDV estimation*		*(5)* 0104: FTA estimation	
Variable	*Coefficient (t-stat) Prob{t}*	*Coefficient (t-stat) Prob{t}*	*t(error) Prob{t}*	*Coefficient (t-stat) Prob{t}*	*t(error) Prob{t}*	*Coefficient (t-stat) (Prob{t})*	*t(error) Prob{t}*
lrgdp	**0.785** (135.9) (0.00)	**0.805** (10.5) (0.00)	0.3 **(0.79)**	**0.785** (56.1) (0.00)	0.4 **(0.97)**	**0.785** (135.3) (0.00)	-0.02 **(0.99)**
lrgdppc	**0.596** (55.2) (0.00)	**0.558** (4.8) (0.00)	-0.3 **(0.74)**	**0.594** (4.8) (0.00)	-0.1 **(0.94)**	**0.595** (54.1) (0.00)	-0.07 **(0.94)**
ldist	**-1.148** (-60.4) (0.00)	---	---	**-1.150** (35.9) (0.00)	-0.05 **(0.96)**	**-1.149** (58.3) (0.00)	-0.03 **(0.98)**
contig	**0.587** (6.08) (0.00)	---	---	**0.587** (4.1) (0.00)	-0.00 **(0.99)**	**0.584** (6.1) (0.00)	-0.04 **(0.97)**
comlang	**0.490** (11.7) (0.00)	---	---	**0.490** (7.9) (0.00)	0.01 **(0.98)**	**0.488** (11.6) (0.00)	-0.07 **(0.95)**
comctry	**2.009** (7.6) (0.00)	---	---	**1.948** (3.3) (0.00)	-0.10 **(0.92)**	**1.97** (7.4) (0.00)	-0.15 **(0.88)**
comcol	**0.584** (11.0) (0.00)	---	---	**0.584** (6.7) (0.00)	0.00 **(0.99)**	**0.588** (11.0) (0.00)	0.07 **(0.94)**
colonial	**2.099** (17.9) (0.00)	---	---	**2.095** (12.8) (0.00)	-0.02 **(0.98)**	**2.090** (17.8) (0.00)	-0.08 **(0.94)**
sdd	**-0.043** (-21.2) (0.00)	**-0.043** (14.3) (0.00)	-0.1 **(0.91)**	**-0.043** (15.4) (0.00)	-0.04 **(0.97)**	**-0.043** (21.5) (0.00)	-0.03 **(0.97)**
LDV	---	---	---	**-0.001** (0.08) **(0.47)**	---	---	---
FTA	---	---	---	---	---	**-0.007** (0.07) **(0.47)**	---
constant	**-17.763** (-69.9) (0.00)		-8.0 **(0.00)**	**-17.731** (36.9) (0.00)	-0.1 **(0.95)**	**-17.742** (69.4) (0.00)	0.08 **(0.93)**
$\overline{R^2}$	*0.603*	*0.089*		*0.574*		*0.603*	
F stat	*3870.49*	*526.63*		*1379.85*		*3481.27*	
rmse	*2.098*	*2.099*		*2.098*		*2.098*	
N	*22948*	*250 x 22948*		*250 x 22948*		*250 x 22948*	

*Only the impact coefficients are reported, since the LDV coefficient is statistically zero.

Table One (Lite): Lite-Newtonian Data-Generating Process (DGP)

Column two shows the parameters that generate the data.
Column three shows the simulation estimates generated from these data. Numbers in **bold** in column three show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)* **Lite-Newtonian DGP**	*(3)* 0102Lite: Fixed effects estimation	
Variable	*Coefficient* *(t-stat)* *Prob{t)*	*Coefficient* *(t-stat)* *Prob{t}*	*t(error)* *Prob{t}*
lrgdp	**0.688** (113.0) (0.00)	**0.694** (8.0) (0.00)	0.07 **(0.94)**
lrgdppc	**0.594** (49.7) (0.00)	**0.589** (4.5) (0.00)	-0.04 **(0.97)**
ldist	**---**	---	---
contig	---	---	---
comlang	---	---	---
comctry	---	---	---
comcol	---	---	---
colonial	---	---	---
sdd	**-0.055** (-24.0) (0.00)	**-0.055** (16.2) (0.00)	-0.05 **(0.96)**
LDV	---	---	---
FTA	---	---	---
constant	**-15.739** (-67.2) (0.00)	**-23.711** (18.2) (0.00)	---
$\overline{R^2}$	*0.497*	*0.066*	
F stat	*7557.33*	*384.77*	
rmse	*2.3614*	*2.362*	
N	*22948*	*250 x 22948*	

Table Two: Fixed Effects Data-Generating Process (DGP)

Column two shows the parameters that generate the data.
Columns three through five show the simulation estimates generated from these data.
Numbers in **bold** in columns three through five show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)* **Fixed Effects DGP**	*(3)* 0201: Newtonian estimation		*(4)* 0201Lite: Newtonian estimation, Lite version		*(5)* 0203: LDV estimation		
Variable	*Coefficient* *(t-stat)* *Prob{t}*	*Coefficient* *(t-stat)* *Prob{t}*	*t(error)* *Prob{t}*	*Coefficient* *(t-stat)* *Prob{t}*	*t(error)* *Prob{t}*	*Impact Coefficient* *(t-stat)* *(Prob{t})*	*Long-run Coefficient* *=impact/ (1-.935)*	*t(error,LR)* *Prob{t}*
lrgdp	**-0.438** (-9.4) (0.00)	**0.785** (135.9) (0.00)	211.8 **(0.00)**	**0.688** (114.7) (0.00)	185.0 **(0.00)**	**0.038** (5.3) (0.00)	**0.585**	9.3 **(0.00)**
lrgdppc	**1.224** (17.4) (0.00)	**0.596** (55.2) (0.00)	-58.2 **(0.00)**	**0.593** (49.6) (0.00)	-52.7 **(0.00)**	**0.120** (12.0) (0.00)	**1.846**	4.0 **(0.00)**
ldist	---	**-1.148** (60.4) **(0.00)**	---	**---**	---	**-0.074** (4.5) (0.00)	**-1.138**	---
contig	---	**0.588** (6.1) **(0.00)**	---	---	---	**-0.072** (0.95) (0.17)	**-1.108**	---
comlang	---	**0.493** (24.3) **(0.00)**	---	---	---	**0.007** (0.21) (0.41)	**0.108**	---
comctry	---	**1.857** (7.0) **(0.00)**	---	---	---	**0.271** (0.88) (0.19)	**4.169**	---
comcol	---	0.582 (10.9) (0.00)	---	---	---	**0.226** (4.96) (0.00)	**3.477**	---
colonial	---	**2.109** (17.9) **(0.00)**	---	---	---	**0.312** (3.6) (0.00)	**4.800**	---
sdd	**-0.012** (-6.7) (0.00)	**-0.043** (21.5) (0.00)	-15.2 **(0.00)**	**-0.054** (23.8) (0.00)	-18.6 **(0.00)**	**-0.008** (5.3) (0.00)	**-0.123**	4.8 **(0.00)**
LDV	---	---	---	---	---	**0.935** (132.0) (0.00)	**---**	---
FTA	---	---	---	---	---	---	---	---
constant	**-4.662** (6.64) (0.00)	**-17.757** (69.8) **(0.00)**	-8.0 **(0.00)**	**-23.578** (103.9) (0.0)	---	**-2.077** (8.3) (0.00)	**---**	---
$\overline{R}^2$	*0.041(w/in)*	*0.603*		*0.497*		*0.867*	*0.867*	
F stat	*233.51*	*3869.39*		*7559.76*		*6650.51*	*6650.51*	
rmse	*1.273*	*2.099*		*2.361*		*1.101*	*1.101*	
N/#groups	*22948/6707*	*250 x 22948*		*250 x 22948*		*250 x 22948*	*250 x 22948*	

Table Three: Lagged Dependent Variable Data-Generating Process (DGP)

Column two shows the parameters that generate the data—both impact and long-run coefficients.
Columns three through five show the simulation estimates generated from these data. Numbers in **bold** in columns three through five show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter—the *long run* parameter, wherever a long-run parameter is available. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)* **LDV DGP**		*(3)* 0301: Newtonian estimation		*(4)* 0302: FE estimation		*(5)* 0304: FTA estimation	
Variable	*Impact Coefficient* *(t-stat)* *Prob{t)*	*Long-run coefficient =impact/ (1-.958)*	*Coefficient* *(t-stat)* *Prob{t}*	*t(error,LR)* *Prob{t}*	*Coefficient* *(t-stat)* *Prob{t}*	*t(error,LR)* *Prob{t}*	*Coefficient* *(t-stat)* *(Prob{t})*	*t(error,LR)* *Prob{t}*
lrgdp	**0.036** (5.89) (0.00)	**0.857**	**0.776** (101.81) (0.00)	-10.6 **(0.00)**	**-0.711** (8.3) (0.00)	18.3 **(0.00)**	**0.779** (102.2) (0.00)	-10.2 **(0.00)**
lrgdppc	**0.064** (7.59) (0.00)	**1.524**	**0.647** (45.22) (0.00)	-61.3 **(0.00)**	**1.000** (8.2) (0.00)	-4.3 **(0.00)**	**0.633** (43.9) (0.00)	-61.8 **(0.00)**
ldist	**-0.020** (-1.42) (0.16)	**-0.476**	**-1.066** (45.38) (0.00)	25.1 **(0.00)**	---	---	**-1.024** (50.9) (0.00)	27.2 **(0.00)**
contig	**0.082** (1.29) (0.20)	**1.952**	**0.503** (4.21) (0.00)	-12.1 **(0.00)**	---	---	**0.469** (3.9) (0.00)	-12.3 **(0.00)**
comlang	**-0.040** (-1.43) (0.15)	**-0.952**	**0.461** (8.98) (0.00)	27.5 **(0.00)**	---	--	**0.441** (8.6) (0.00)	27.2 **(0.00)**
comctry	**0.327** (1.3) (0.21)	**7.786**	**1.737** (3.55) (0.00)	-12.4 **(0.00)**	---	---	**1.554** (3.2) (0.00)	-12.8 **(0.00)**
comcol	**0.030** (0.79) (0.43)	**0.714**	**0.534** (7.43) (0.00)	-2.5 **(0.01)**	---	---	**0.496** (6.88) (0.00)	-3.0 **(0.00)**
colonial	**0.024** (.3) (0.75)	**0.571**	**1.770** (13.15) (0.00)	8.9 **(0.00)**	---	---	**1.782** (13.3) (0.00)	9.0 **(0.00)**
sdd	**-0.008** (-6.64) (0.00)	**-0.190**	**-0.026** (11.06) (0.00)	-69.8 **(0.00)**	**-0.015** (7.3) (0.00)	-85.2 **(0.00)**	**-0.026** (11.2) (0.00)	-70.6 **(0.00)**
LDV	**0.958** (159.8) (0.00)		---	---	---	---	---	---
FTA	---		---	---	---	---	**0.705** (6.49) (**0.00**)	---
constant	**-1.805** (-8.5) (0.00)		**-19.109** (55.75) (0.00)	-50.5 **(0.00)**	**18.816** (11.8) (0.00)	12.9 **(0.00)**	**-19.364** (56.2) (0.00)	-51.00 **(0.00)**
$\overline{R^2}$	*0.902*		*0.658*		*0.036*		*0.659*	
F stat	*9419.20*		*2184.43*		*74.61*		*1978.20*	
rmse	*0.932*		*1.7420*		*0.963*		*1.738*	
N	*10246*		*250 x 10246*		*250 x 10246*		*250 x 10246*	

Table Four: FTA Data-Generating Process (DGP)

Column two shows the parameters that generate the data.
Columns three through five show the simulation estimates generated from these data.
Numbers in **bold** in columns three through five show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)* **FTA DGP**	*(3)* 0401: Newtonian estimation		*(4)* 0403: LDV estimation*		*(5)* 0405: FTA + CU estimation	
Variable	*Coefficient (t-stat) Prob{t}*	*Coefficient (t-stat) Prob{t}*	*t(error) Prob{t}*	*Coefficient (t-stat) Prob{t}*	*t(error) Prob{t}*	*Coefficient (t-stat) (Prob{t})*	*t(error) Prob{t}*
lrgdp	**0.788** (136.4) (0.00)	**0.785** (135.9) (0.00)	-0.5 **(0.63)**	**0.778** (56.0) (0.00)	-0.7 **(0.497)**	**0.790** (136.6) (0.00)	0.46 **(0.64)**
lrgdppc	**0.583** (53.6) (0.00)	**0.595** (55.1) (0.00)	1.1 **(0.26)**	**0.594** (31.1) (0.00)	0.6 **(0.54)**	**0.584** (53.7) (0.00)	0.12 **(0.91)**
ldist	**-1.104** (-56.1) (0.00)	**-1.149** (60.5) (0.00)	-2.4 **(0.02)**	**-1.150** (36.1) (0.00)	-1.5 **(0.14)**	**-1.089** (55.0) (0.00)	0.76 **(0.44)**
contig	**0.584** (6.06) (0.00)	**0.584** (6.05) (0.00)	0.0 **(0.999)**	**0.624** (4.3) (0.00)	0.28 **(0.78)**	**0.594** (6.2) (0.00)	0.11 **(0.91)**
comlang	**0.470** (11.2) (0.00)	**0.487** (11.6) (0.00)	0.4 **(0.67)**	**0.491** (7.9) (0.00)	0.3 **(0.73)**	**0.447** (10.6) (0.00)	-0.54 **(0.59)**
comctry	**1.825** (6.9) (0.00)	**1.967** (7.4) (0.00)	0.5 **(0.59)**	**1.988** (3.4) (0.00)	0.28 **(0.78)**	**1.40** (5.1) (0.00)	-1.56 **(0.12)**
comcol	**0.548** (10.3) (0.00)	**0.588** (11.0) (0.00)	0.7 **(0.46)**	**0.593** (6.8) (0.00)	0.52 **(0.60)**	**0.511** (9.5) (0.00)	-0.70 **(0.48)**
colonial	**2.123** (18.1) (0.00)	**2.090** (17.8) (0.00)	-0.3 **(0.78)**	**2.091** (12.8) (0.00)	-0.2 **(0.84)**	**2.151** (18.3) (0.00)	0.24 **(0.81)**
sdd	**-0.043** (-21.5) (0.00)	**-0.043** (21.6) (0.00)	0.2 **(0.87)**	**-0.043** (15.3) (0.00)	0.1 **(0.92)**	**-0.043** (21.5) (0.00)	0.25 **(0.80)**
LDV	---	---	---	**0.006** (0.44) **(0.33)**	---	---	---
FTA	**0.903** (8.8) (0.00)	---	---	---	---	**0.837** (8.1) (0.00)	-0.65 **(0.52)**
CU	---	---	---	---	---	**0.903** (6.2) (**0.00**)	
constant	**-18.023** (-70.6) (0.00)	**-17.744** (69.9) (0.00)	1.1 **(0.27)**	**-17.535** (36.5) (0.00)	1.0 **(0.31)**	**-18.262** (70.8) (0.00)	-0.93 **(0.35)**
$\overline{R}^2$	*0.604*	*0.603*		*0.574*		*0.605*	
F stat	*3502.88*	*3868.17*		*1380.99*		*3193.05*	
rmse	*2.095*	*2.098*		*2.099*		*2.093*	
N	*22948*	*250 x 22948*		*250 x 22948*		*250 x 22948*	

*Only the impact coefficients are reported, since the LDV coefficient is statistically zero.

Table Five: FTA + CU Data-Generating Process (DGP)

Column two shows the parameters that generate the data.
Column three shows the simulation estimates generated from these data. Numbers in **bold** in column three show each estimated coefficient, and the probability value for a test that this coefficient is identical to the true data-generating parameter. (For estimated coefficients that do not appear in the data-generating process, the emboldened probability value presents a test that the estimated coefficient is significantly non-zero.)

(1)	*(2)*	*(3)*	
	FTA + CU DGP	0504: FTA estimation	
Variable	*Coefficient* *(t-stat)* *Prob{t)*	*Coefficient* *(t-stat)* *(Prob{t})*	*t(error)* *Prob{t}*
lrgdp	**0.790**	**0.788**	-0.46
	(136.6)	(136.5)	**(0.64)**
	(0.00)	(0.00)	
lrgdppc	**0.584**	**0.583**	-0.12
	(53.7)	(53.6)	**(0.91)**
	(0.00)	(0.00)	
ldist	**-1.089**	**-1.104**	-0.77
	(-55.0)	(56.1)	**(0.44)**
	(0.00)	(0.00)	
contig	**0.594**	**0.584**	-0.11
	(6.17)	(6.1)	**(0.91)**
	(0.00)	(0.00)	
comlang	**0.447**	**0.470**	0.55
	(10.6)	(11.2)	**(0.59)**
	(0.00)	(0.00)	
comctry	**1.398**	**1.825**	1.61
	(5.1)	(6.9)	**(0.11)**
	(0.00)	(0.00)	
comcol	**0.511**	**0.548**	0.70
	(9.5)	(10.3)	**(0.48)**
	(0.00)	(0.00)	
colonial	**2.151**	**2.123**	-0.24
	(18.3)	(18.1)	**(0.81)**
	(0.00)	(0.00)	
sdd	**-0.043**	**-0.043**	-0.25
	(-21.2)	(21.5)	**(0.80)**
	(0.00)	(0.00)	
LDV	**---**	**---**	---
FTA	**0.837**	**0.903**	0.65
	(8.1)	(8.8)	**(0.52)**
	(0.00)	(0.00)	
CU	**0.903**	**---**	---
	(6.2)		
	(0.00)		
constant	**-18.262**	**-18.023**	
	(-70.8)	(70.6)	
	(0.00)	(0.00)	
$\overline{R^2}$	*0.605*	*0.604*	
F stat	*3193.05*	*3502.88*	
rmse	*2.093*	*2.095*	
N	*22948*	*250 x 22948*	

Table Six: CU Effects and Multicollinearity

Column two reports coefficients in a regression of the Custom Unions variable against all other independent variables in the largest of the preceding models.

(1)	*(2)* **CU as Dependent Variable**
Variable	*Coefficient (t-stat) Prob{t}*
lrgdp	**-0.003** (-11.41) (0.00)
lrgdppc	**-0.001** (-2.83) (0.005)
ldist	**-0.017** (-18.88) (0.00)
contig	**-0.012** (-2.71) (0.01)
comlang	**0.025** (13.4) (0.00)
comctry	**0.472** (39.6) (0.00)
comcol	**0.042** (17.3) (0.00)
colonial	**-0.031** (-5.83) (0.00)
sdd	**-0.001** (-6.2) (0.00)
FTA	0.074 (16.0) (0.00)
constant	**0.265** (23.0) (0.00)
$\overline{R^2}$	*0.178*
F stat	*497.39*
rmse	*0.095*
N	*22948*

References

Cheng, I.H. and Howard Wall (1999). "Controlling for Heterogeneity in Gravity Models of Trade." *Federal Reserve Bank of St. Louis Working Paper* 1999-010D.

Deardorff, Alan V. (1998). "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" In J.A. Frankel, ed., *The Regionalization of the World Economy*, University of Chicago Press, 7-32.

Eichengreen, Barry and Douglas A. Irwin (1996). "The Role of History in Bilateral Trade Flows." *NBER Working Papers*, 5565, National Bureau of Economic Research, Inc.

Ghosh, S. and S. Yamarik, (2004). "Are Regional Trade Arrangements Trade-Creating? An Application of Extreme Bounds Analysis." *Journal of International Economics*, 63:2, 369-95.

Keller, Wolfgang (1998). Are International R&D Spillovers Trade-Related? Analyzing Spillovers among Randomly Matched Trade Partners." *European Economic Review*, 42:8, 1469-81.

Keller, Wolfgang (2000). "Do Trade Patterns and Technology Flows Affect Productivity Growth?" *World Bank Economic Review*, vol. 14, no. 1, January 2000, 17-47

Kleijnen, Jack (2004). Design and analysis of Monte Carlo Experiments Tilburg University, Center for Economic Research, Discussion Paper: 17

Leamer, Edward and James Levinsohn (1995). "International Trade Theory: The Evidence." In G. Grossman and K. Rogoff, eds., *The Handbook of International Economics: Vol. III*, Elsevier Science B.V., 1339-1394.

Nason, James M. and John H. Rogers (2006). "The present-value model of the current account has been rejected: Round up the usual suspects." *Journal of International Economics*, vol. 68, no. 1, January 2006, 159-187.

Poyhonen, Pentti (1963). "A Tentative Model for the Volume of Trade Between Countries." *Weltwirtschaftliches Archiv* 90(1), 93-99.

Rose, Andrew K. and T.D. Stanley (2005). "A Meta-Analysis of the Effect of Common Currencies on International Trade." *Journal of Economics Surveys*, 19,3, 347-X.

Rose, Andrew K. (2000). "One Money, One Market: The Effect of Common Currencies on Trade." Economic Policy, 15:30, 7-46. Data posted at http://faculty.haas.berkeley.edu/arose/RecRes.htm .

Tinbergen, Jan (1962). *Shaping The World Economy: Suggestions for an International Economic Policy*. New York: 20th Century Fund.

Wall, Howard I. (1999). "Using the Gravity Model to Estimate the Costs of Protection." *Federal Reserve Bank of St. Louis Review* 81:1 (January/February), 33-40.