Bayesian Estimation on Currency Union Effect*

RONALDO CARPIO[1] AND MEIXIN GUO[†2]

[1]School of Finance, University of International Business and Economics, China

[2]ACCEPT, Tsinghua University, China

Accepted in Review of International Economics

Abstract

This paper uses Bayesian method to estimate European (Monetary) Union effect on trade. The high dimensionality of the parameter space when estimating gravity equations with many dummy variables results in standard hypothesis tests with a large Type I (false positive) error. Bayesian methods are able to handle this problem; they also provide a principled method of model selection that can be applied to different specifications of the dummy variables. Bayesian model selection tests prefer our most unrestricted dummy specification, which includes asymmetric bilateral effects, as well as time-varying, country-specific factors. Our estimate shows a zero Euro effect on trade, but a 14.8% increase in imports for a member of the European Union during 1980-2004.

Keywords: Bayesian Method; Currency Union Effect; Gravity Equation; Model Selection

JEL classification: C1, F13, F14, F45

*The authors are greatly indebted to Paul Bergin, Guido Kuersteiner, and Katheryn Russ for their valuable advice and constant encouragements, and also very grateful to Colin Cameron, Òscar Jordà, and Yong Li for valuable suggestions. The authors also appreciate the comments from Brown bag participants at UC Davis, Tsinghua University and Renming University.

†Corresponding author. Email addresses: rncarpio@yahoo.com (R. Carpio), mxguo@sem.tsinghua.edu.cn (M. Guo).

1 Introduction

Since its origin in the 1960s, the gravity equation has been very successful in explaining trade flows between pairs of countries in terms of the countries' output/expenditures, bilateral trade costs and other factors. Anderson and Van Wincoop (2003) refine the theoretical and empirical foundations of the gravity equation, and note the importance of the multilateral resistance terms.[1] These terms capture the general equilibrium effect of bilateral trade costs and countries' business cycles on bilateral trade flows.[2] A simple way to control for these multilateral reistance terms with cross-sectional data is to add importer and exporter fixed effects (i.e. dummy variables), as proposed in Anderson and Van Wincoop (2003) and Feenstra (2004). If we have panel data on country pairs, there are many ways in which these fixed effects may be included: time-varying or time-invariant country dummies, with or without country-pair dummies. With industry-level or firm-level panel data, the possibilities for interactions between time, country, and industry/firm allow even more complicated specifications.

A key issue in estimating longitudinal gravity models is that the estimates of important parameters seem to vary considerably, depending on which specification of dummy variables is used. For example, consider the effect of currency union on trade; Rose and Stanley (2005) use meta-regression analysis to evaluate a list of estimates under two cases: random and fixed effect models. They find a currency union increase two member countries' trade about 47% with a 95% confidence interval ranging from 20% to 80%. To consider another specific example, there is considerable debate on the significance and size of the Eurozone (EZ) effect on trade.[3] Baldwin and Taglioni (2007)

[1]Anderson and Van Wincoop (2003) derive the gravity model in an Armington world, Eaton and Kortum (2002) obtain a similar model with a Ricardian framework. The model can be extended to include heterogeneous firms; Chaney (2008), Helpman et al. (2008) and others derive a more generalized version of the gravity model.

[2]See Head and Mayer (2013) for more discussions on the modular trade impact and general equilibrium trade impact with respect to the multilateral resistance terms.

[3]The trade effects of free trade agreements (FTA) (such as NAFT, GATT, and WTO etc.), have

show that the EZ effect ranges from 0 to about 40% depending on the specification of the dummies.[4] They provide a thorough summary of the literature, but still leave open the question of which fixed effect specification to use. Different choices of fixed effects have distinct theoretical and empirical implications; in this paper, however, we will use model selection methods that rely on the data alone.[5] Another issue that comes up when estimating the gravity model is the problem of high dimensionality of the parameter space. As the number of countries or years in the sample grows, the number of fixed effects increases, resulting in a parameter space that becomes extraordinarily large. For example, later in the paper, we estimate the Eurozone and European Union (EU) effects on a dataset that has 11,500 observations, but our most unrestricted dummy specification has 2469 degrees of freedom - clearly, the number of observations is far too small to invoke the usual asymptotic properties of least squares (LS) estimators. Consequently, the parameter space of the gravity model becomes extraordinarily large.

This study uses panel data for 22 developed countries during the years 1980-2004 to estimate the trade effects of the EZ and EU, and illustrates how the hierarchical Bayesian method can help avoid the problems described above and test different models. We specify ten empirical models implied by different theories, which combine different groups of dummy variables commonly used in the literature.[6] The unrestricted model

---

also been the subject of debate in Rose (2004), Baier and Bergstrand (2007), Subramanian and Wei (2007) and Head and Mayer (2013).

[4]Rose and Van Wincoop (2001) find a 58% increase in trade because of the integration of the EZ using country fixed effects. Micco et al. (2003) obtain an effect of about 5%-20% based on country-pair and year dummies.

[5]This dummy variable selection problem is related to the (subset) variable selection problem in the statistics literature, which either uses various shrinkage methods or imposes some structure on the data to reduce dimensionality. Representative papers are Shao (1997), Meinshausen and Buhlmann (2006),Chen and Chen (2008), Fan and Lv (2008), Zhang and Huang (2008), and Wang (2010), among many others. Data in genomics and finance are typical examples. Researchers effectively identify key parameters among thousands of predictors, and frequently face the case that the number of observations is less than the number of parameters. In the gravity model literature, however, the sample size is larger than the dimension of regressed coefficients. Meanwhile, trade costs and economic mass variables are of key interest, and little attention is paid to the dummy variables.

[6]This paper uses Bayesian method to compare different models and focuses on the model selection. Bayesian model averaging could be another way to deal with the model uncertainty.

(denoted as the model "Baseline"), includes time-varying importer and exporter fixed effects and asymmetric country pair fixed effects. This is the most general specification and nests all other (restricted) models. The simplest model only controls for nation fixed effects(denoted as the model "NA"). We use the Bayesian version of the likelihood ratio test (the BLR test) proposed by Li et al. (2014*a*) and also information criteria to compare the ten models. The Bayesian model selection statistics all prefer the Baseline model, in which countries do not import significantly more from each other although they use a common currency. This model selection result is consistent with the empirical model recommended in Baier and Bergstrand (2007), who use country pair fixed effects and country-and-time effects to control for the endogeneity of FTAs and the multilateral resistance terms.[7] In particular, the result suggests that asymmetric country pair fixed effects could capture the unobserved heterogeneity among country pairs, and then supports the inclusion of (unobserved) heterogeneous preferences across countries into a structural gravity equation, as proposed in Guo (2015).[8]

We use Monte Carlo simulation to show that standard hypothesis tests (the likelihood ratio (LR) test and the Wald test) based on panel LS suffer from a large type I (false positive) error rate. This is due to the high dimensionality of the parameter space, and the hundreds of constraints associated with these hypothesis tests. Simple dimension adjustments to the LR test statistics are insufficient to correct these size issues, even in the case of homoscedastic errors.[9] Simulation results on the Wald test

[7]Baier and Bergstrand (2007) specify different groups of fixed effects in estimating the effects of FTAs on trade and propose a panel approach with country-pair fixed effects to control for the endogeneity of the FTAs. Similarly, Head and Mayer (2013) recommend the country-pair dummies to account for the endogeneity of EZ. Their recommendation, however, is not theory founded and does not distinguish the roles of importer and exporter in trade. The specification of our "Baseline" model is more general. This paper does not propose a solution to the endogeneity problem directly, but does present a method for selecting dummy variables with econometric techniques.

[8]Guo (2015) changes the assumption on the homogeneous preferences in Anderson and Van Wincoop (2003) to heterogeneous preferences across countries, and derives an augmented gravity model with asymmetric pair dummies and time-varying importer and time-varying exporter fixed effects from a multi-country dynamic stochastic general equilibrium model.

[9]With heteroscedastic errors, the type I error of the revised LR test remains large.

also find that coefficients common to restricted and unrestricted models cannot be distinctly estimated. Taking these results together, We conclude that panel LS is not a credible approach for comparing high-dimensional gravity equations.

The Bayesian approach to inference does not rely on large-sample approximations (see Chamberlain and Imbens (2003)), but provides distributions of all estimates. The coefficients on the EZ and EU dummies are assumed to follow a normal distribution with mean 0.2 and variance 0.25, based on the results in Baldwin and Taglioni (2007) and Head and Mayer (2013);[10] other parameters are assumed to have diffuse priors in the absence of prior information. Here, a hierarchical prior specification breaks down the data structure into two levels of sub-models, which naturally captures the original multi-level system.[11] Consequently, the key advantages of hierarchical Bayesian estimation are that it reduces the number of (key) constraints across models, and avoids the small sample size problem from the LS estimations.

Although the emphasis here is on estimating the currency union or EZ effect in gravity models, the Bayesian method is more broadly applicable to other topics using the gravity equation, such as evaluating the effects of different trade policy, such as FTAs. The Bayesian method can also be used in any situation with a high-dimensional parameter space, or with multi-level panel data, or with many fixed effects. For example, evaluating the return to schooling while controlling for different combinations of individual-school-county-state fixed effects is another case with a high dimensional and hierarchical parameter space. This method could also provide an alternative estimation method for confronting possible over-parameterization, as compared to regular panel LS regressions. Though economists have long known about the small sample size

[10] Head and Mayer (2013) review 329 published papers estimating the EU effect, and find that the EU estimate has a mean (and median) of about 0.14-0.23, with a variance of about 0.25. The literature shows that EZ and EU can increase trade by about 20% on average, and the variance is about 20-30%. We do robustness checks with different priors; the conclusion on the model section remains.

[11] See similar arguments in Burda et al. (2008) for a multinomial choice case.

problem, Bayesian methods have not been commonly used in the trade literature and other applied microeconomic studies so far.[12]

The structure of the paper is organized as follows. Section 2 introduces the gravity equation and ten different specifications for the dummy variables. Section 3 provides various results from the panel LS regressions depending on the choice of dummies, and illustrates the magnitudes of the type I error for the LR test and Wald test. Section 4 presents the Bayesian results. The last section concludes.

2 Gravity Equation and Specifications

This section first presents the theoretical gravity model and ten popular empirical specifications on fixed effects in the gravity equation literature. The subsequent subsections provide the results on EZ and EU effects on trade using LS, and test hypotheses on different groups of dummy variables.

2.1 Structural Gravity Model

The structural gravity equation in Anderson and Van Wincoop (2003) augmented with heterogeneous bilateral preferences in consumption baskets for each importer i and exporter k can be shown to yield,

[12]To my knowledge, Ranjan and Tobias (2007) is the only exception using Bayesian method to estimate gravity equations in the trade literature besides Guo (2015). They apply the Bayesian approach to properly handle zero-values in trade between countries and a non-linear relationship between contract enforcements and trade. Instead, Silva and Tenreyro (2006) propose to use the Poisson pseudo-maximum-likelihood (PPML) to control for the heteroskedasticity resulted from the massive zero trade values. Here our data have very few zeros observations in bilateral imports. For applied microeconomic empirical studies, Chamberlain and Imbens (2003) propose a nonparametric Bayesian approach for two cases with high-dimensional parameter spaces: educational choice and quantile regression, due to the parameter uncertainty (the first case) and the violation of the traditional asymptotical distribution assumption (the later one).

$$im_t^{ik} \equiv \frac{IM_t^{ik} * WOUT_t}{EXP_t^i * OUT_t^k} = \alpha^{ik} \left(\frac{P_t^i \Pi_t^k}{\tau_t^{ik}} \right)^{\eta - 1} \tag{1}$$

where the multilateral resistance for exporter/good k is defined as

$$\left(\Pi_t^k\right)^{1-\eta} = \sum_{j=1}^{N} \left(\frac{P_t^j}{\tau_t^{jk}} \right)^{\eta-1} \alpha^{jk} \frac{EXP_t^j}{WOUT_t}, \tag{2}$$

and the aggregate price level (the multilateral resistance for importer i) is defined as

$$P_t^i = \left(\sum_{k=1}^{N} \alpha^{ik} \left[\tau_t^{ik} Q_t^k \right]^{1-\eta} \right)^{\frac{1}{1-\eta}}. \tag{3}$$

where Q_t^k is price level for a good k produced in country k.

The gravity equation above specifies that the bilateral imports IM_t^{ik} (country i imports good k from country k) are positively influenced by world output "$WOUT_t$", importers' expenditure shares of world output "$EXP_t^i/WOUT_t$", exporters' output shares of world output "$OUT_t^k/WOUT_t$", and importer's preference on good k "α^{ik}", but are impeded by trade costs "τ_t^{ik}". To estimate this gravity model, the specifications of dummy variables/fixed effects to control for the two "multilateral (gravitational inconstant) trade resistance terms" (MLR_t^{ik}, a weighted average trade costs), i.e. $MLR_t^{ik} = (P_t^i \Pi_t^k)^{\eta-1}$ is debatable in the literature.

The main difference between our structural gravity model and traditional gravity model is the share α^{ik}[13]. These preferences weights can be explained as the bilateral

[13] In section 2.3, Head and Mayer (2013) derive a series of similar structural gravity equations from various demand-side and/or supply-side theories without the heterogeneous bilateral preferences α^{ik}. In the demand-side theories, Anderson and Yotov (2010) use β_i^k to differentiate the products by place of origin i and sector/class k in the utility function. The parameter β_i^k is regarded as a share or quality parameter, and specific to an importer (and a good) but common to all her trade partners, which corresponds to assume $\alpha^{ik} = \alpha^i$ (k in this paper is defined as an importer). The βs are either canceled out in the structural gravity equation or absorbed in the country fixed effects in the estimation. In the supply-side theories, they also mention βs reflect the absolute advantages in productivity draw distributions across sectors. Here, following Guo (2015), the heterogeneous bilateral preferences α^{ik} are different from theirs. See section 3 of Guo (2015) for details.

preference shock in the gravity equation with heterogenous consumers discussed in Head and Mayer (2013). They also can be interpreted as the bilateral fixed trade costs in the model with heterogenous firms as in Chaney (2008). Indeed, they capture all unobservable heterogeneity in country pairs (such as asymmetric country-pair preferences and trade barriers in the data), regarded as a generalization of "Canadian-U.S. asymmetric border effects" in Bergstrand et al. (2013)[14].

Because of potential endogenous economic mass variables and the non-stationarity issue, we use the import ratio im_t^{ik} instead the import level as the dependent variable. Since memberships in the European Union (EU) entailed economic reforms that could be expected to raise bilateral trade themselves, Baldwin and Taglioni (2007) evaluate the effect of European Monetary Union (Eurozone) separately from EU policies. Following Baldwin and Taglioni (2007), we include variables on EZ and EU in the gravity model as well as other standard set of controls representing trade costs and fixed effects. Our most general model, the "Baseline", is shown below after we take logs on equation (1),

$$w_t^{ik} = cons + \beta_{EZ} EZ + \beta_{EU} EU + \delta^{ik} + \theta_t^i + \phi_t^k + \sum_{j=1}^{J} \gamma_{jt} g_j^{ik} + \varepsilon_t^{ik}. \tag{4}$$

The dependent variable, bilateral import ratio, is defined as $w_t^{ik} = log(im_t^{ik})$,[15] and the asymmetric country pair dummies δ^{ik} are used to capture the preferences α^{ik}. Both time-varying importer and exporter dummies are used to control for the MLR_t^{ik}, i.e. $log(MLR_t^{ik}) = \theta_t^i + \phi_t^k$.

The form of bilateral trade costs is assumed as follows,

[14]Bergstrand et al. (2013) provide strong evidence for the "asymmetric border effects", which means the direct trade-impeding effect of the Canadian-U.S. border is much larger for Canadian imports from the U.S. relative to the U.S. imports from Canada.

[15]See appendix A for another two versions of estimation equations commonly used in the literature. This paper focuses on the results for bilateral import ratios to avoid a possible endogeneity problem on the economic mass variables.

$$\left(\tau_t^{ik}\right)^{1-\eta} \equiv \prod_{j=1}^{J} \left(G_{tj}^{ik}\right)^{\gamma_{jt}}. \tag{5}$$

Taking logs of both sides of the equation obtains the log of trade costs, $\sum_{j=1}^{J} \gamma_{jt} g_{tj}^{ik}$, where variables $g_{tj}^{ik} (\equiv log\left(G_{tj}^{ik}\right))$, include logs of bilateral distances (log(dist.)) and four dummies for border (contig.), common official language (comlang.), Eurozone membership (EZ), and European Union membership (EU). The variable "EZ" is equal to 1 if both countries use Euro in trade; and the variable "EU" is equal to 1 if the two countries belong to the European Union.

2.2 Constraints and Implications

Table 1 lists another nine variations of equation (4) based on different assumptions on MLR_t^{ik} in the literature, which are all nested in the "Baseline" model.[16] These restricted models arise from combinations of three key restrictions with respect to the dummy variables, which have different theoretical implications on the estimated Euro effect and European Union effect on trade.

The first restriction is on country pair dummies, excluding these asymmetric pair dummies ($\delta^{ik} = 0$) or imposing symmetric restrictions on pair dummies ($\delta^{ik} = \delta^{ki}$) in the regression. The asymmetric county pair dummies are constant over time, and variables on EU and EZ are time-varying since countries become members in different years. Including these pair dummies help absorb the effects of all constant bilateral relationships between members on trade flows so that variables on EU and EZ can precisely capture how bilateral trade of a specific country-pair changes over time due to

[16] The trade cost variables are redundant because of the multi-collearity with the asymmetric country pair dummies; time-varying importer and exporter dummies also drop one for each country due to the same reason. In this paper with 22-country and 25-year panel data, the Baseline model drops total 72 dummies, including 3 trade costs variables, 25 time-varying importer dummies for USA (θ_t^{USA}), 1 time-varying exporter dummy in 2004 for USA, and another 43 asymmetric country pair dummies.

the adoption of a customs union and common currency. The estimated Euro effect is a within effect and symmetric/asymmetric if we use symmetric/asymmetric country-pair fixed effects. Without the pair dummies, the estimation may suffer from the omitted variable problem or endogeniety issue (Baier and Bergstrand (2007)) since these pair dummies capture all unobserved bilateral trade relations and correlated with EZ and EU variables. An exception is that these unobserved bilateral relations are homogeneous across country-pairs. Anderson and Van Wincoop (2003) and the followers assume homogeneous preferences and unobserved trade costs in their model and do not control for the asymmetric pair dummies.

The second one focuses on the multilateral resistance terms. Due to the elimination of currency differences among countries, a substantial diversion of trade away from the rest of the world would occur theoretically, and then the multilateral resistance/remoteness terms change over time accordingly. Therefore, the time-varying not constant multilateral resistance terms should be included in the estimations based on Anderson and Van Wincoop (2003); the estimated Euro effect is more close to a between effect (a cross-sectional comparison). Frankel (2010), however, argues that such large trade diversion from currency unions is not robust in the data and then the multilateral resistance terms and price index are relative constant over years due to constant trade costs and expenditure shares in Equations (2) and (3). We can impose and test the constraint on the multilateral resistance terms by data, which takes the forms $log(MLR_t^{ik}) = \theta^i + \phi^k$ or $log(MLR_t^{ik}) = \mu_t + \theta^i + \phi^k$.[17] We also can assume that the multilateral resistance is country-pair specific but constant, the constraint will have $log(MLR_t^{ik}) = \mu_t + \widetilde{\zeta^{ik}}$ so that estimated country pair dummies ζ^{ik} is the sum of two parts, i.e. $\zeta^{ik} = \delta^{ik} + \widetilde{\zeta^{ik}}$.

The third constraint ignores the different roles for importers and exporters, and

[17] Note that country fixed effects, such as θ^i and ϕ^k, are collinear with the country-pair dummies.

contains nation dummies only, such as $log(MLR_t^{ik}) = \theta^i + \theta^k$ or $log(MLR_t^{ik}) = \theta_t^i + \theta_t^k$ ($\theta^k = \phi^k$ or $\theta_t^k = \phi_t^k$). This constraint assumes that trade costs are symmetric for importers and exporters and that countries all have a balanced trade or very small shares of trade balance over GDP every period. The symmetric trade costs assumption is inconsistent to Waugh (2010), who finds that poor countries face higher export costs than rich counties and this asymmetric trade costs are quantitatively important to explain the large international income differences. The small trade balance assumption implies that international borrowing and lending markets are not important for countries to smooth consumptions.

3 Eurozone Effect or European Union Effect?

3.1 Standard Panel Regressions on EZ and EU Effects

We collect the annual data for 22 OECD countries during 1980-2004 following Baldwin and Taglioni (2007)(appendix A) to estimate the two effects on trade. Countries participated the EU and EZ in different years. There are 14 countries in EU by year 1995 and 8 non-EU countries. Among the EU group, 4 countries, Denmark, Greece, Sweden, and United Kingdom did not use euro by year 1999. The EU and EZ effects on trade can be distinguished by the variations of memberships across countries and years. All models are given in table 1 and the Baseline model includes time-varying importer and exporter fixed effects, and asymmetric country-pair dummies.[18].

Table 2 shows the results from the panel LS estimations. Using euros in trade increases the bilateral import ratio 18% on average, and the effect varies from -0.4% (insignificant) to 51.1% (significant).[19] Compared to the EZ effect, the EU effect on

[18]Note that Baldwin and Taglioni (2007) use the import level as the dependent variable, and their results are replicated and available upon request.

[19]The number 51.1% is equal to $e^{(0.413)} - 1$. Other percentages below are also calculated similarly.

import is more stable and varies from 16.3% to 25.5%. In the Baseline model, the EZ policy does not affect trade significantly while the EU membership has a significant and positive effect on imports, 25.5% more.[20] So do models, columns of "YIMYEX-Pair" and "YNAPair", with symmetric pair dummies. Estimations with time-varying country fixed effects only (columns of "YIMYEX" and "YNA"), however, show that the significant effect of the EZ on increasing import ratios is 51.1%, compared to non-EZ members. Estimations with time-invariant country fixed effects only (columns of "NA", "NAYear", "IMEX" and "IMEXYear") show that both the EZ and EU have large effects in promoting imports, 34-35% and 16-18% respectively. To sum up, these results illustrate that the EZ and EU effects vary significantly with the choice of dummy variables though the standard errors of the coefficients on the EZ and EU variables are similar across different specifications.[21]

From the above discussions, we conclude: 1) the choice of time-varying or constant country fixed effects does matter significantly for the estimations of the EZ and EU effects on import ratios; 2) models with either symmetric or asymmetric country pair dummies reach similar results; 3) similar estimates are produced whether the model isolates the role of importing/exporting country or not.

3.2 Hypothesis Tests and Large Type I Error

Based on the estimates in Table 2, we use four hypothesis tests to distinguish the ten models, and their statistics are given in table 3. Results from the classical LR test (the dimension adjusted LR test suggested by Italianer (1985)) are listed in the LR1 (LR2) column.[22] The LR test and F test assume i.i.d. error terms. The "Wald_NW" column

[20]Results with import levels yield qualitatively similar conclusions with those with import ratios.

[21]The results from MLE shown in the appendix remain similar to the LS estimations.

[22]Italianer (1985) finds that the LR test statistic is chi-squared distributed with the correction factor m/N, where N is the number of observations and m is equal to $(N - r - 0.5 * d_n)$ with the number of restrictions r and the dimension d_n of the restricted model (the null hypothesis). See appendix B.1

takes heteroskedasticity into account, and provides Wald statistics using Newey-West standard errors with 2 lags, which are robust to heteroskedasticity and autocorrelation (HAC) (Newey and West (1987, 1994)).[23] All four tests reject the null hypothesis in the first nine combinations of the null and alternative models. That is, the "Baseline" model cannot be rejected. In other words, based on the LS results, this data sample supports the model with asymmetric country-pair and time-varying importer/exporter fixed effects, where the EZ has no effect on the import ratios but the EU increases the import ratio by 25.5%. This conclusion, however, is not credible since these tests suffer from large size distortions (the type I error, the rate of rejecting the null when the null is true). The following discussions provides the limitations of the LS estimations and the size distortions of these tests.

The preceding regressions include a large number of dummy variables, especially for the Baseline model, and the hypothesis tests have hundreds of constraints (the number of different parameters in the null and alternative models, the degree of freedom (DF) for the test, the column of "DF" in table 3). Though the hypothesis tests all support the Baseline model, statistical inference may suffer from a small sample size problem due to the unusually high dimensionality of the parameter space and the extremely large group of constraints associated with the hypothesis tests. Evans and Savin (1982) and Italianer (1985) show that the finite sample distribution of the statistic is biased towards the conventional large sample asymptotic chi-square distribution. Then the statistics (or the critical values for the 5% significant level) need to be adjusted.

Monte Carlo simulations with i.i.d errors show a large size distortion for the classical LR test (the LR1 column with spherical errors) in Table 4,[24] more than 50% on average. Using the same consistent but biased estimated variance as in LR1 column, the Wald

and table 4 for calculations on LR2.

[23]The conclusion remains up to 5-lags.

[24]See appendix B.1 for simulation details for columns of LR1, LR2, Wald1, Wald2 and F. Considering the HAC, the size distortions are even larger.

test (the Wald1 column) has an even larger size distortion, 77% on average. In contrast, the F test and the second Wald test (the Wald2 column) with dimension adjusted consistent and unbiased estimated variance have normal sizes. For example, in the first combination (1) (Comb.(1): "NA" vs "Baseline"), the sizes for the LR1, Wald1 and Wald2 test are 61.5%, 93.7% and 6.7% respectively. In other words, we reject the true null model with time-invariant nation dummies slightly more often (1.7% more than 5%) using the Wald2 test, but too often for the LR1 and Wald1 tests. Supplementary rejection rates of the null model is provided in figure 1, which presents the power curves of these tests.[25] The dimension adjusted LR kernel density (LR2) based on Italianer (1985) performs better than the LR1 test but its type I error remains high (more than 10%) for the Comb.(7) and (8), which test the significance of country-pair dummies. These large size occurs because the distributions of the tests with small finite samples differ from the asymptotic ones.[26]

3.3 How Different Are These Ten Specifications?

Do different models estimate consistent coefficients though only one model has the true data generating process? We do MC simulations to estimate the coefficients common to the null and alternative models given the true coefficients are from the null model. Table 5 presents the type I error of the Wald test[27] on the estimated common coefficients from the null (restricted) and the alternative (unrestricted) models against the true coefficients with homoscedastic error terms. The null hypothesis in Table 5 is

[25]See appendix B.1 for the details. For the Comb. (1) ("NA" vs "Baseline"), the null hypotheses $H0$ have 1469 constraints, including $\delta^{ik} = 0$ and $\Delta = \phi_t^k - \theta^k = \theta_t^k - \theta^k = 0$. We assume that δ and Δ follow normal distributions with a zero mean and a common variance σ_Δ^2 and obtain the power curves by increasing the σ_Δ^2 from zero to 0.05.

[26]See appendix B.1 for more discussions.

[27]From here on, we use the Wald2 statistics as the Wald test with a small and normal size. Table 5 provides the results with homoscedasticity. In the appendix, tables B.2-B.5 also provide the results with misspecified heteroscedasticity.

$H0 : \hat{\mathbf{B}}_m = \mathbf{B}_0$ for $m \in \{a, n\}$, where the true coefficients $\mathbf{B}_0$ are initially estimated from the null model and included in both the null and alternative models. The Wald statistic is calculated by the following formula,

$$Wald = (\hat{\mathbf{B}}_m - \mathbf{B}_0) * \left[\widehat{var(\mathbf{B}_m)}\right]^{-1} * (\hat{\mathbf{B}}_m - \mathbf{B}_0). \tag{6}$$

where $\widehat{var(\mathbf{B}_m)} = \hat{\sigma}_m^2 * (X_m' X_m)^{-1}$, $\hat{\sigma}_m^2 = RSS_m/(N - K_m)$, and $m \in \{a, n\}$. Then the rejection rate (size) of the test is the frequency of rejection over 1000 simulations. The rejection rates for both the null and alternative models with homoscedasticity are all close to 5% in the large column "All coefficients in the Null Model". In particular, the second large column focuses on the EZ and EU effects, which gives a similar conclusion on the sizes, around 5%. That is, both the null and alternative models can estimate consistent Euro and EU effects though only the null model is true.

In addition, Table 6 provides the rejection rates (powers) of the Wald test on the β_{EZ} and β_{EU} in the null model when the true data generating process deviates from the null model.[28] For example, in the Comb. (1), we impose the restrictions $\delta^{ik} = 0$, $\Delta = \phi_t^k - \theta^k = \theta_t^k - \theta^k = 0$ on the model "Baseline" to get the model "NA". We assume that δ and Δ follow normal distributions with a zero mean and a common variance σ_Δ^2. When the variance σ_Δ^2 increases to 0.05, we should reject the null model more often. Surprisingly, the rejection rate of the coefficients on the EZ and EU effects estimated from the null "NA" remains low, 4.7% though the the model "NA" is wrong. In other words, the low dimensional models (time-invariant fixed effects) are sufficient to provide consistent estimators for EZ and EU effects if the variances of the time-varying importer and exporter fixed effects are small. In other cases except the Comb. (7), the rejection rates all remain low, around 5%, though the null model is wrong. These results show

[28]See the details in appendix B.1.

that the Wald test has a low power on the type II error.

4 Bayesian Method

Previous results from the hypothesis tests support the Baseline model with asymmetric country pair and time-varying importer and exporter fixed effects for this specific data sample. These tests rely on the asymptotical probability theory. Due to limited sample size, the large dimension of the parameter space can influence the asymptotical distributions of the estimates and leads to a large type I error. Bayesian framework allows to cope with hierarchical/multilevel models without encountering the problems of "overfitting"(Gelman et al. (2004)). The Bayesian method can estimate the distributions of all parameters, including the distributions of coefficients on EU and EZ and the distributions of the variance across years. The latter provides direct visual evidence on the volatility of these time-varying dummies. The hierarchical structure matches the multi-leveled data and has the advantage of reducing the dimension of the key parameter space and number of constraints to test different models, which avoids the large dimensionality problem in LS and MLE.[29] Section 4.1 provides the hierarchical Bayesian model and its estimation methodology.[30] The next subsection shows the Bayesian results on the model selection.

[29]Theoretically, feasible general least squares (FGLS) on these models can be used to obtain the likelihood and then to do hypothesis tests. But the unknown variance matrix (13255x13255) makes FGLS impractical unless we assume an analytical formula for the inverse of the variance matrix. Hausman and Kuersteiner (2008) find that FGLS is biased when error terms with homoscedasticity across groups but unconstrained covariance matrix within a group.

[30]This paper tries to do model selection and comparison by Bayesian method instead of Bayesian model averaging.

4.1 Estimation Models and Priors

The general two-level hierarchical linear Bayesian model with normal prior distributions for regression coefficients and Gamma/Wishart prior distributions for variance coefficients is given as follows (Gelman et al. (2004), Gelman (2006), and Koop et al. (2007)),

$$\mathbf{Y} = \mathbf{X} * \mathbf{B}_1 + \varepsilon \tag{7}$$

$$\begin{aligned}
\mathbf{Y} = \mathbf{X}\mathbf{B}_1 + \varepsilon, \quad E(\mathbf{Y}) \quad &= \overline{\mathbf{Y}} \\
\text{where } \overline{\mathbf{Y}}| \quad \mathbf{B}_1, \Sigma_1 \quad &\sim N(\mathbf{X} * \mathbf{B}_1, \Sigma_1) \\
\mathbf{B}_1| \quad \mathbf{B}_2, \Sigma_2 \quad &\sim N(W * \mathbf{B}_2, \Sigma_2) \\
\mathbf{B}_2| \quad \mathbf{B}_3, \Sigma_3 \quad &\sim N(Z * \mathbf{B}_3, \Sigma_3) \\
\Sigma_{1,2,3}^{-1}| \quad \mathbf{B}_{1,2,3} \quad &\sim Gamma/Wishart,
\end{aligned}$$

where the $\mathbf{B}_2$ is the key/hyper parameter vector. This Normal-Wishart prior distribution gives analytical posterior (conditional) distributions for all parameters and has an advantage for estimation. In this paper, the hierarchical structure of the Baseline model can be specified as follows,

$$\begin{aligned}
w_t^{ik} \quad &= \quad cons + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + \theta_t^i + \phi_t^k + \varepsilon_t^{ik} \qquad (8) \\
w_t^{ik}|\Theta \quad &\sim \quad N\left(\overline{w_t^{ik}}, \left(\sigma^{ik}\right)^2\right) \\
\text{where } \overline{w_t^{ik}} \quad &= \quad cons + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + \theta_t^i + \phi_t^k \\
\frac{1}{\left(\sigma^{ik}\right)^2} \quad &\sim \quad G(a_1, a_2),
\end{aligned}$$

where the error term has a simple heteroskedastic form without serial correlations. The choice on this type of error term mainly relies on the fact that heterogeneity of cross-sectional country pair significantly affects the estimation of the EZ and EU effects based on the previous section 3.[31]

The specific values and prior distributions for the parameters are listed below,[32]

$cons \sim N(0, 1000)$, $\beta_{EZ} \sim N(0.2, 0.25)$, and $\beta_{EU} \sim N(0.2, 0.25)$;

$\delta^{ik} \sim N\left(p, \sigma_p^2\right)$ with mean $p \sim N(0, 1000)$ and variance $\frac{1}{\sigma_p^2} \sim G(a_1, a_2)$;

$\theta_t^i \sim N\left(\theta^i, \sigma_{\theta t}^2\right)$ with mean $\theta^i \sim N(0, 1000)$ and variance $\frac{1}{\sigma_{\theta t}^2} \sim G(a_1, a_2)$;

$\phi_t^k \sim N\left(\phi^k, \sigma_{\phi t}^2\right)$ with mean $\phi^k \sim N(0, 1000)$ and variance $\frac{1}{\sigma_{\phi t}^2} \sim G(a_1, a_2)$;

where the shape $a_1 = 0.001$ and the rate $a_2 = 1000$. β_{EZ} and β_{EU} use the informative priors based on Baldwin and Taglioni (2007) and Head and Mayer (2013) while priors of other parameters are assumed to be noninformative.[33] The time-invariant country pair fixed effects δ^{ik} are assumed to have a common mean p and variance σ_p^2 to capture the common features of these 22 OECD countries. The time-varying importer and exporter fixed effects, the multilateral resistance terms, θ_t^i and ϕ_t^i are expected to have heterogeneous means θ^i and ϕ^i across countries and variances $\sigma_{\theta t}^2$ and $\sigma_{\phi t}^2$ across years. They are used to control for the business cycle properties and shocks for each country in each year. Volatile and/or large posterior distributions of $\sigma_{\theta t}^2$ and $\sigma_{\phi t}^2$ across years, as well as volatile posterior distributions of θ^i and ϕ^i, provide direct evidence to support the models with time-varying country fixed effects. Depending on the specifications of

[31] With a heteroskedastic and serially correlated variance and covariance structure, the time-varying component of the data would be absorbed in the error ε_t^{ik} instead of the country fixed effects θ_t^i and ϕ_t^k; consequently, the model with time-invariant fixed effect is more easily preferred.

[32] This hierarchical linear model is a special case of mixed linear models in Cameron and Trivedi (2005) with randomly varying intercepts (p774). Here the priors for the parameters $\mathbf{B}_1$ (and $\mathbf{B}_2$) are assumed to be independent; their posteriors are correlated.

[33] We also try other priors on β_{EZ} and β_{EU}. First, we use the diffuse priors on the two parameters. With diffuse priors, Bayesian results show that the Baseline model is still preferred and that the magnitude of the EZ effect is similar. We also try priors of the hyperparameters that are similar to Ranjan and Tobias (2007), who uses the hierarchical Bayesian method to estimate a non-parametric threshold Tobit gravity model because of too many zeros in the trade data. The conclusions on model selection and the EZ effect remain; the import data in this paper has no zeros.

these dummies, models and priors vary. For example, the model "IMEX" with time-invariant fixed effects imposes the constraints: $\theta_t^i = \theta^i$, $\phi_t^i = \phi^i$, $\delta^{ik} = 0$, $\sigma_{\theta t}^2 = 0$ and $\sigma_{\phi t}^2 = 0$ on the Baseline model, assumes $\theta^i \sim N(0, 1000)$ and $\phi^i \sim N(0, 1000)$, and includes the diffuse priors of the coefficients γ_j on the three trade costs variables.

The estimation algorithm is the Markov Chain Monte Carlo Simulation via Gibbs Sampler.[34] Two convergence tests are used here to determine the burning and draw times: 1) Gelman-Rubin statistic (BGR) with $|R - 1| < 0.05$ for single parameter and multiple parameters (Gelman (2006));[35] 2) Geweke chi-squared test (Geweke (1992)).[36] Most of the parameters in the small models like "NA", "NAYear", "IMEX", "IMEXYear" and "PairYear" converge after 5000 burn-in times based on the two tests. The large models, however, have more than one thousand parameters and converge slowly. After 100,000 burn-in times, parameters in all models have converged based on BGR statistics for 50000 draws (thin 10) from either single chain or multiple chains. The second level (key) parameters (β_2) in large models converged based on Geweke Chi-squared statistics.

4.2 Bayesian Results

Table 7 shows the estimated coefficients and standard deviations on variables EZ and EU for the ten models using the Bayesian method. The average EZ effect in increasing import ratios is 15%. The EZ effects are significantly positive in seven models: "NA", "NAYear", "IMEX", "IMEXYear", "YIMYEX" and "PairYear"; their magnitudes are a little smaller than those using the LS in Table 2. In another three models: "Baseline", "YIMYEXPair", and "YNAPair", using euros has no significant effect on trade. In

[34] See appendix B.5 for the five estimation steps.
[35] Matlab code reference: the GNU General Public License.
[36] Matlab code reference: James P. LeSage, Dept of Economics Texas State University-San Marcos, jlesage@spatial-econometrics.com.

contrast, an EU membership is consistently estimated to have a significantly positive effect on import ratio, 17% more on average. The variation of EZ and EU effects due to the different choices on dummy variables remain as the LS and MLE estimations.

Since Bayes factor/odds ratio suffers from Jeffreys' concern and Jeffreys'-Lindleys' paradox,[37] we use the Bayesian approach for the LR test (BLR test) proposed by Li et al. (2014a) to compare models. When the likelihood function is available in closed-form and equations are estimated by MCMC, this BLR test statistic is defined by the difference of the posterior means of the log-likelihood values for the null and alternative models,

$$T = \overline{\widehat{LL\left(y \mid \Theta_a\right)}} - \overline{\widehat{LL\left(y \mid \Theta_n\right)}} \tag{9}$$

where the $LL\left(y \mid \Theta_m\right)$ $(m \in \{u, r\})$ is the log-likelihood value (the unrestricted or restricted model respectively). The symbol "-" refers to the mean value. This BLR statistic asymptotically follows $\chi 2(p) - p$, where p is the number of constraints.

We also use three information criteria: Akaike Information Criterion(AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC) to select models. The three criteria are developed based on the posterior log-likelihood with a penalty on the number of dimensions. DIC is initially developed by Spiegelhalter et al. (2002); and Celeux et al. (n.d.) provides eight different versions of DICs for latent variable models. The key interest of the ten competing models is how to specify the (latent) dummy variables; so we treat all dummy variables in the model as parameters. Then

[37] See Gelman et al. (2004) (page 185-186) and Li et al. (2014a) for extensive discussions.

we use the conditional DIC_7 in Celeux et al. (n.d.), which is calculated as below,[38]

$$\begin{aligned}
DIC_7 &= \overline{\widehat{D\left(y,\Theta\right)}} + p_D \\
\text{where } p_D &= \overline{\widehat{D\left(y,\Theta\right)}} - D\left(y,\overline{\widehat{\Theta}}\right), \\
\overline{\widehat{D\left(y,\Theta\right)}} &= -2 * \overline{\widehat{LL\left(y \mid \Theta\right)}} \text{ and } D\left(y,\overline{\widehat{\Theta}}\right) = -2 * LL\left(y \mid \overline{\widehat{\Theta}}\right).
\end{aligned}$$

The $D\left(y,\Theta\right)$ is the Bayesian deviance, a goodness of fit (a measure of surprise or uncertainty). The posterior mean deviance $\overline{\widehat{D\left(y,\Theta\right)}}$ is equal to -2 times the mean of posterior log-likelihood $\overline{\widehat{LL\left(y \mid \Theta\right)}}$, and the deviance $D\left(y,\overline{\widehat{\Theta}}\right)$ uses the log-likelihood calculated by the mean of posterior parameters $\overline{\widehat{\Theta}}$. The p_D, represents the effective dimension proposed by Spiegelhalter et al. (2002), a penalty with a larger number of parameters. The smaller number of DIC implies a better fit of the model. Another two commonly used criteria are AIC and BIC (Congdon (2005) and Iliopoulos et al. (2007)) shown as below,

$$\begin{aligned}
AIC &= D\left(y,\overline{\widehat{\Theta}}\right) + 2 * d \\
BIC &= D\left(y,\overline{\widehat{\Theta}}\right) + d * log\left(N\right)
\end{aligned}$$

where d is the (effective) number of estimated parameters and N is the number of observations. Smaller values in AIC and BIC indicate better fit of the model.

Table 7 presents the Bayesian statistics of the ten models. The BLR column shows the statistics of the BLR test, and the alternative model is the Baseline model. These test statistics are all larger than the critical values associated with dimension at 1% significant level, and reject the null model. Hence, the "Baseline" model with asymmetric

[38]Though Celeux et al. (n.d.) recommend DIC_3 and DIC_4, Li et al. (2014*b*) find that only three versions of DICs: DIC_1, DIC_2 and DIC_7, are coherent. Although Li et al. (2014*b*) also find DIC_7 has a few theoretical problems, it is computationally convenient to use DIC_7, as explained in Spiegelhalter et al. (2002).

country-pair fixed effects as well as time-variant import and exporter fixed effects, is supported by the BLR test. Additionally, all three criteria give the lowest values for the Baseline model. These Bayesian statistics all favor the Baseline model to other nine restricted models. In the Baseline model, the posterior mean of the EZ effect on trade is negative but insignificant, whereas the posterior mean of the EU effect is positive and significant. In other words, two countries using the Euro does not increase imports with each other; but membership in the European Union does help increase imports by 14.8% within the union.[39]

We expect a positively significant effect of EU on trade since the EU is a unique economics and political union. The initial organization, European Economic Community, was created in 1958 among six initiated countries: Belgium, Germany, France, Italy, Luxembourg and the Netherlands. Then, 22 other members joined and a huge single market has been created and continue to grow. The estimated positive EU effect is 14.8%, close to the mean of EU estimates (17.35%) in Head and Mayer (2013).

The insignificant EZ effect may come from three reasons. Since EU membership has enhanced trade a great deal together among the EU countries, the EZ effect on promoting more trade could be limited. Second, most factors that form the EU also plausibly cause the formation of European Monetary Union, such as countries' GDPs, historical relationships, relative factor endowments and geographic distances.[40] We observe a positive correlation between the EU and EZ variables. The average correlation over the whole sample is 0.384, but varies considerably across country-pairs (from zero to 0.688). We also find that variable EU varies much more than variable EZ across time and country pairs. The overall standard deviation of EU dummy (0.449) is twice of that of EZ dummy (0.225).[41] Thus, Due to the positive correlation of the two policies

[39]The posterior means and variances of other hyper parameters are shown in the Table B.6.

[40]See Baier and Bergstrand (2004) and Baier and Bergstrand (2007) for more discussions on the economic determinants of FTAs.

[41]The between standard deviation of the EU dummy (0.385) is about 4 times of that of the EZ

and the smaller variation of the EZ, the effect of EZ could be partly captured by the EU effect. Furthermore, agents in our 22 OECD countries can access international financial markets freely. We expect very low transaction costs of converting currency across these countries so that using a common currency "Euro" reduces very small barriers technically and has no significant effect on trade flows. Baldwin and Taglioni (2007) also find negative but insignificant effect of Eurozone on trade once they control for the country pair and nation-year fixed effects.

The supportive evidence for the Baseline model can also be found in Figure 2, which depicts fairly volatile posterior 95% credible intervals of the variances of time-varying fixed effects. The posterior means of the two variances $\sigma^2_{\theta t}$ and $\sigma^2_{\phi t}$ show a U-shaped trend; the yearly world shocks decrease from year 1980, reach the trough in the mid 1990s, and increase from the late 1990s.

The Baseline model is derived by augmenting the theory of Anderson and Van Wincoop (2003) with heterogeneous shares on products in the consumption bundle. The importer-year and exporter-year fixed effects capture the "multilateral resistance terms", the essential requirement of the theoretical gravity equation in Anderson and Van Wincoop (2003). The country pair fixed effects further help econometrically resolve the endogeneity of the EU and EZ policies addressed in Baier and Bergstrand (2007). Therefore, the model selection result identifies the necessity of both groups of fixed effects in the gravity equation although most empirical studies in gravity equation commonly considers either of them only.

dummy (0.097). The within standard deviations of the two dummies are close, 0.232 on variable EU and 0.203 on variable EZ.

5 Conclusion

This paper studies how different specifications of the bilateral trade relations and the multilateral resistance terms in the gravity equation influence the estimated effect of trade policies such as EZ and EU. We have considered ten commonly used gravity models, and have shown that the choice of dummy variables affects the magnitude of the estimated currency union effect on bilateral imports. Three groups of dummies, which are included in the gravity equations to control for individual country and country-pair fixed effects, are compared: asymmetric vs. symmetric country-pair dummies, time-varying vs. time-invariant country dummies, and separate importer/exporter vs. nation dummies. Depending on the choice of dummies, EZ and EU effects on trade during the period 1980-2004 vary greatly using the LS and MLE methods, from -0.4% to 51%. Based on the LS results, the conventional Wald test, LR test and F test are used to assess the necessity of the different dummies, but these tests have large size distortions. Two factors contribute to this large size distortion. First, the high dimensionality of the parameter space leads to biased asymptotical chi-square distributions for the LR test and Wald tests. Second, the large number (hundreds) of constraints associated with the hypothesis tests drive the size distortion's sensitivity to the dimension adjustment method on the test statistics.

Panel LS estimation results and Monte Carlo simulations on size distortions more or less show that symmetric and asymmetric pair fixed effects obtain similar EZ and EU effects, and that separating the role of importer and exporter in the estimations also does not significantly change the coefficients compared to the model with symmetric nation dummies. The choice of time-varying vs. time-invariant country dummies, however, affects the estimations considerably. In another words, the choice of time-varying vs. constant country dummies is the least clear-cut.

Then we provide alternative econometric method: Bayesian method to re-estimate all ten models and do model selection using Bayesian statistics. The Bayesian method provides the distributions of all parameters and allows investigation of the variances of those time-varying country fixed effects. The Bayesian likelihood ratio test and three information criteria clearly show that the unrestricted Baseline model with asymmetric pair fixed effects, time-varying importer and exporter fixed effects, is favored among all ten models. The EZ effect on trade disappears but the EU effect on Trade remains significant and positive, 14.8%. The volatile posterior distributions of the variance parameters in the Baseline model provide evidence to support the time-varying importer and exporter effects. This model selection result identifies the necessity of the country pair effects and importer-time and exporter-time fixed effects in the gravity equation.

References

Anderson, J. E. and Yotov, Y. V. (2010), 'The changing incidence of geography', *The American economic review* **100**(5), 2157–2186.

Anderson, J. and Van Wincoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American Economic Review* **93**(1), 170–192.

Aviat, A. and Coeurdacier, N. (2007), 'The geography of trade in goods and asset holdings', *Journal of International Economics* **71**(1), 22–51.

Baier, S. L. and Bergstrand, J. H. (2004), 'Economic determinants of free trade agreements', *Journal of International Economics* **64**(1), 29 – 63.
URL: *http://www.sciencedirect.com/science/article/pii/S0022199603000795*

Baier, S. L. and Bergstrand, J. H. (2007), 'Do free trade agreements actually increase members' international trade?', *Journal of International Economics* **71**(1), 72 – 95.

Baldwin, R. and Taglioni, D. (2007), 'Trade effects of the euro: a comparison of estimators', *Journal of Economic Integration* **22**(3), 780–818.

Bergstrand, J. H., Egger, P. and Larch, M. (2013), 'Gravity redux: Estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs', *Journal of International Economics* **89**(1), 110 – 121.
URL: *http://www.sciencedirect.com/science/article/pii/S0022199612000955*

Burda, M., Harding, M. and Hausman, J. (2008), 'A Bayesian mixed logit-probit model for multinomial choice', *Journal of Econometrics* **147**(2), 232–246.

Cameron, C. A. and Trivedi, P. (2005), 'Microeconometrics: Methods and applications'.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (n.d.), 'Deviance information criteria for missing data models', *Bayesian Analysis* (4), 651–673.

Chamberlain, G. and Imbens, G. (2003), 'Nonparametric applications of Bayesian inference', *Journal of Business and Economic Statistics* **21**(1), 12–18.

Chaney, T. (2008), 'Distorted gravity: The intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–21.

Chen, J. and Chen, Z. (2008), 'Extended Bayesian information criteria for model selection with large model spaces', *Biometrika* **95**(3), 759.

Congdon, P. (2005), *Bayesian Models for Categorical Data*, Chichester, UK John Wiley and Sons Ltd.

Eaton, J. and Kortum, S. (2002), 'Technology, geography, and trade', *Econometrica* **70**(5).

Evans, G. and Savin, N. (1982), 'Conflict among the criteria revisited: The w, lr and lm tests', *Econometrica* **50**(3), 737–48.

Fan, J. and Lv, J. (2008), 'Sure independence screening for ultrahigh dimensional feature space', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 849–911.

Feenstra, R. (2004), *Advanced International Trade: Theory and Evidence*.

Frankel, J. (2010), The estimated trade effects of the euro: Why are they below those from historical monetary unions among smaller countries?, *in* 'Europe and the Euro', University of Chicago Press, pp. 169–212.

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models', *Bayesian Analysis* **1**(3), 515–533.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004), *Bayesian Data Analysis*, New York, NY: Chapman and Hall.

Geweke, J. (1992), 'Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments', *Bayesian statistics* **4**(2), 169–193.

Guo, M. (2015), 'Financial frictions, trade costs and international consumption risk sharing', *Tsinghua University. Manuscript* .

Hausman, J. and Kuersteiner, G. (2008), 'Difference in difference meets generalized least squares: Higher order properties of hypotheses tests', *Journal of Econometrics* **144**(2), 371–391.

Head, K. and Mayer, T. (2013), 'Gravity Equations: Workhorse,Toolkit, and Cookbook'.

Helpman, E., Melitz, M. and Rubinstein, Y. (2008), 'Estimating trade flows: Trading partners and trading volumes', *The Quarterly Journal of Economics* **123**(2), 441–487.

Iliopoulos, G., Kateri, M. and Ntzoufras, I. (2007), 'Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table', *Computational Statistics and Data Analysis* **51**(9), 4643–4655.

Italianer, A. (1985), 'A small-sample correction for the likelihood ratio test', *Economics Letters* **19**(4), 315–317.

Koop, G., Poirier, D. and Tobias, J. (2007), *Bayesian econometric methods*, Cambridge Univ Pr.

Li, Y., Zeng, T. and Yu, J. (2014*a*), 'A new approach to Bayesian hypothesis testing', *Journal of Econometrics* **178, Part 3**(0), 602 – 612.

Li, Y., Zeng, T. and Yu, J. (2014*b*), 'Robust Deviance Information Criterion for Latent Variable Models'.

Meinshausen, N. and Buhlmann, P. (2006), 'High-dimensional graphs and variable selection with the lasso', *The Annals of Statistics* **34**(3), 1436–1462.

Micco, A., Stein, E., Ordoñez, G., Midelfart, K. and Viaene, J. (2003), 'The currency union effect on trade: early evidence from EMU', *Economic Policy* **18**(37), 317–356.

Newey, W. and West, K. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–08.

Newey, W. and West, K. (1994), 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies* **61**(4), 631–53.

Ranjan, P. and Tobias, L. (2007), 'Bayesian inference for the gravity model', *Journal of Applied Econometrics* **22**(4), 817–838.

Rose, A. (2004), 'Do we really know that the wto increases trade?', *American Economic Review* **94**(1), 98–114.

Rose, A. and Stanley, T. (2005), 'A meta-analysis of the effect of common currencies on international trade', *Meta-regression analysis: issues of publication bias in economics* p. 53.

Rose, A. and Van Wincoop, E. (2001), 'National money as a barrier to international trade: The real case for currency union', *American Economic Review* **91**(2), 386–390.

Shao, J. (1997), 'An asymptotic theory for linear model selection', *Statistica Sinica* **7**, 221–242.

Silva, J. M. C. S. and Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and Statistics* **88**(4), 641–658.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society. Series B, Statistical Methodology* pp. 583–639.

Subramanian, A. and Wei, S. (2007), 'The wto promotes trade, strongly but unevenly', *Journal of International Economics* **72**(1), 151–175.

Wang, H. (2010), 'Factor Profiling for Ultra High Dimensional Variable Selection', *Peking University, Working Paper* .

Waugh, M. E. (2010), 'International trade and income differences', *American Economic Review* **100**(5), 2093–2124.

Zhang, C. and Huang, J. (2008), 'The sparsity and bias of the Lasso selection in high-dimensional linear regression', *Annals of Statistics* **36**(4), 1567–1594.

Table 1: List of equations: bilateral import ratios

Models	Equations
Baseline	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + \theta_t^i + \phi_t^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
NA	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \theta^i + \theta^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
NAYear	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \mu_t + \theta^i + \theta^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
YNA	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \theta_t^i + \theta_t^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
YNAPair	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \zeta^{ik} + \theta_t^i + \theta_t^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
IMEX	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \theta^i + \phi^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
IMEXYear	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \mu_t + \theta^i + \phi^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
YIMYEX	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \theta_t^i + \phi_t^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
YIMYEXPair	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \zeta^{ik} + \theta_t^i + \phi_t^k + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$
PairYear	$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \zeta^{ik} + \mu_t + \sum_{j=1}^{3}\gamma_j g_j^{ik} + \varepsilon_t^{ik}$

Table 2: Eurozone effect and European Union effect on the log bilateral import ratio by LS: 1980-2004

Var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dummies	Baseline	NA	NAYear	YNA	YNAPair	IMEX	IMEXYear	YIMYEX	YIMYEXPair	PairYear
Time-varying	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Imp. & Exp.	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No
Nation	No	Yes	Yes	Yes	Yes	No	No	No	No	No
Year	No	No	No	Yes	No	No	Yes	No	No	Yes
Pair	Asym.	No	No	No	Sym.	No	No	No	sym.	Sym.
EZ	-0.004	0.290***	0.302***	0.413***	-0.004	0.290***	0.302***	0.413***	-0.004	0.139***
	0.055	0.052	0.057	0.110	0.053	0.052	0.058	0.112	0.055	0.038
EU	0.227***	0.164**	0.151**	0.171	0.227***	0.164**	0.151**	0.171	0.227***	0.163***
	0.056	0.068	0.074	0.109	0.054	0.068	0.074	0.112	0.055	0.038
log(dist)		-0.894***	-0.895***	-0.890***		-0.894***	-0.895***	-0.890***		
		0.064	0.064	0.067		0.064	0.064	0.069		
contig.		0.212*	0.212*	0.212*		0.212*	0.212*	0.212		
		0.125	0.125	0.127		0.125	0.125	0.130		
comlang.		0.423***	0.421***	0.424***		0.423***	0.421***	0.424***		
		0.105	0.105	0.107		0.105	0.105	0.109		
locked_EX										1.293***
										0.054
locked_IM										1.121***
										0.054
Observations	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550
A.R2	0.954	0.755	0.758	0.759	0.858	0.793	0.797	0.798	0.904	0.855

*** significant at 1%; ** significant at 5 %; * significant at 10%. The standard errors reported below the coefficient estimates for all models are clustered on time-invariant country pairs. "LL" is the log-likelihood based on i.i.d. errors. "EZ" variable is the dummy for Eurozone effect, taking one when both countries are in the Eurozone and zero when at least one country is not in the Eurozone. "EU" variable is the dummy for European Union, taking one if country pairs are members of European Union (tracing back to European Coal and Steel Community and Treaty of Rome).

Table 3: Hypotheses testing: 1980-2004

Comb.	H0	H1	Wald_NW	LR1	LR2	CV(Chi2)	F-Test	CV(F)	DF
(1)	NA	Baseline	89171.9	21003.3	18307.4	1559.3	35.3	1.1	1469
(2)	NAYear	Baseline	85753.8	20809.9	18160.4	1534.5	35.2	1.1	1445
(3)	YNA	Baseline	72244	20282.7	18142.9	985.4	51.2	1.1	941
(4)	YNAPair	Baseline	20894.8	13898	12555.7	799.2	31.9	1.1	735
(5)	IMEX	Baseline	69295.7	19023.1	16598.7	1537.6	29.1	1.1	1448
(6)	IMEXYear	Baseline	65866.8	18793.1	16417.5	1512.9	28.9	1.1	1424
(7)	YIMYEX	Baseline	48478.1	17646	16185.5	464.6	87.2	1.1	416
(8)	YIMYEXPair	Baseline	7589.4	8837.9	8185.2	244.8	55.0	1.2	210
(9)	PairYear	Baseline	27390.9	14737.9	12993.6	1321.0	21	1.1	1238
(10)	IMEX	YIMYEX	1716.4	1377.1	1251.2	1107.8	1.3	1.1	1032
(11)	IMEXyear	YIMYEX	1104.9	1147.1	1043.4	1083.0	1.1	1.1	1008
(12)	YNA	YIMYEX	1412.1	2636.6	2453.4	579.4	5.1	1.1	525
(13)	YNAPair	YIMYEXPair	2914.1	5060.1	4663.4	579.4	10.7	1.1	525

The LR statistics (LR1 and LR2), log of likelihood ratio for the null and the alternative models are asymptotically distributed as chi squared with the degrees of freedom (DF) as given assuming i.i.d. in the error terms. LR1 is traditionally calculated, but the LR2 uses the method in Italianer (1985) to adjust the dimensions (see appendix B.1). Assuming HAC in the error term, the chi square distributed Wald test statistics are shown in the column "Wald_NW", using the Newey-West kernel for panel data with 2 lags. The "DF" column is the difference of the dimensions in two models (the number of constraints imposed on the alternative models). The "CV(Chi2)" and "CV(F)" columns show the critical value at the 5% significance level for Chi-square distribution and F distribution respectively. All null hypotheses are rejected at the 5% significance level.

Table 4: Actual size of the tests with homoscedasticity

Comb.	H0	H1	DF	LR1	LR2	Wald1	Wald2	F
(1)	NA	Baseline	1469	61.5	0	97.3	6.7	5.7
(2)	NAYear	Baseline	1445	62.7	0	97.5	7.1	5.6
(3)	YNA	Baseline	941	66	1.9	90.1	5.4	4.3
(4)	YNAPair	Baseline	735	63.3	5.2	83.9	5.5	4.7
(5)	IMEX	Baseline	1448	62.1	0	97.6	6.9	5.5
(6)	IMEXYear	Baseline	1424	62.7	0	97.5	6.9	5.3
(7)	YIMYEX	Baseline	416	53.4	10.7	65	4.7	4.5
(8)	YIMYEXPair	Baseline	210	37.1	12.5	41.5	4.4	4.3
(9)	PairYear	Baseline	1238	65.7	0.2	95.5	6.2	5.1
(10)	IMEX	YIMYEX	1032	32.9	0.4	71.8	5.7	4.8
(11)	IMEXyear	YIMYEX	1008	33.5	0.6	71.2	6	5.3
(12)	YNA	YIMYEX	525	34	5.3	50.7	5.7	4.9
(13)	YNAPair	YIMYEXPair	525	47.6	7.3	62.5	5.7	4.9

See appendix B.1 for simulations of Wald1, Wald2, F, LR1 and LR2 tests with the homoscedastic errors.

Table 5: Size of the Wald test, H0: $\hat{\mathbf{B}}_m = \mathbf{B}_0$, where $m \in \{n, a\}$

Comb.	Null (n)	Alternative (a)	All Coefficients in Null			β_{EU} and β_{EZ} only	
			DF	n	a	n	a
(1)	NA	Baseline	27	4.5	4	4.1	4.7
(2)	NAYear	Baseline	51	4.8	4.3	4.2	4.7
(3)	YNA	Baseline	555	4.7	6	5.3	4.7
(4)	YNAPair	Baseline	761	6	5.9	4.7	4.7
(5)	IMEX	Baseline	48	4.5	5.3	4.1	4.7
(6)	IMEXYear	Baseline	72	4.7	4.1	4.2	4.7
(7)	YIMYEX	Baseline	1080	5.7	6.2	5.3	4.7
(8)	YIMYEXPair	Baseline	1286	7.2	7.3	4.7	4.7
(9)	PairYear	Baseline	258	5	4.2	4.5	4.7
(10)	IMEX	YIMYEX	48	4.5	4.9	4.1	5.3
(11)	IMEXyear	YIMYEX	72	4.7	5	4.2	5.3
(12)	YNA	YIMYEX	555	4.7	6.2	5.3	5.3
(13)	YNAPair	YIMYEXPair	761	6	6.3	4.7	4.7

The simulations assume the homoscedastic errors with the Wald2 test. The subscripts n and a represents the null and alternative models respectively. See appendix B.1 and appendix B.4 for more details.

Table 6: Power of the Wald test on β_{EU} and β_{EZ} only estimated from the null model: H0: $\hat{\mathbf{B}}_n = \mathbf{B}_0$

Comb.	Null (n)	Alternative (a)	Variance of restricted coefficients, σ^2_Δ					
			0	0.01	0.02	0.03	0.04	0.05
(1)	NA	Baseline	4.1	4.3	4.4	4.4	4.5	4.7
(2)	NAYear	Baseline	4.2	4.1	4.2	4.5	4.7	4.6
(3)	YNA	Baseline	5.3	5.2	5.2	5.6	6.1	6.1
(4)	YNAPair	Baseline	4.7	4.7	4.7	4.7	4.7	4.7
(5)	IMEX	Baseline	4.1	4.2	4.3	4.5	4.3	5.1
(6)	IMEXYear	Baseline	4.2	4.5	4.9	4.5	4.8	4.9
(7)	YIMYEX	Baseline	5.3	6.1	6.5	8	9.6	11.3
(8)	YIMYEXPair	Baseline	4.7	4.7	4.7	4.7	4.7	4.7
(9)	PairYear	Baseline	4.5	4.2	4.5	4.4	4.4	4.4
(10)	IMEX	YIMYEX	4.1	4.2	4	4.1	4.5	5
(11)	IMEXyear	YIMYEX	4.2	4.2	4.5	4.5	4.5	4.5
(12)	YNA	YIMYEX	5.3	5.3	5.3	5.3	5.3	5.3
(13)	YNAPair	YIMYEXPair	4.7	4.7	4.7	4.7	4.7	4.7

The simulations assume homoscedastic errors. The subscripts n and a represent the null and alternative models, respectively. In theory, the larger σ^2_Δ is (a higher deviation from the null model), the more rejections on the null model. But this prediction does not hold in the simulations because the power of the test is too low. See appendix B.1 and appendix B.4 for more details.

Table 7: Model selection by BLR test and three information criteria and estimates on EZ and EU effects

Models	DF	$\overline{\widehat{LL}}$	$\widehat{LL}$	BLR	DIC	AIC	BIC	$\overline{\widehat{r2}}$	$\widetilde{\widehat{\beta}}_{EZ}$	std(β_{EZ})	$\widetilde{\widehat{\beta}}_{EU}$	std(β_{EU})
Baseline	2053	4233	5193	—	-6548	-6280	8819	0.919	-0.009	0.014	0.138**	0.012
NA	489	-6904	-6656	11137	14302	14291	17887	0.731	0.248**	0.016	0.145**	0.013
NAYear	513	-6517	-6259	10750	13548	13544	17317	0.733	0.185**	0.017	0.141**	0.013
YNA	1064	-5663	-5300	9896	12050	12728	20554	0.733	0.158**	0.040	0.131**	0.020
YNAPair	1272	-5633	-5143	9866	12243	12831	22186	0.820	0.010	0.039	0.211**	0.033
IMEX	510	-5922	-5665	10155	12358	12350	16101	0.765	0.221**	0.015	0.193**	0.013
IMEXYear	534	-5438	-5169	9671	11414	11406	15333	0.764	0.137**	0.014	0.203**	0.013
YIMYEX	1635	-3435	-2733	7668	8274	8736	20761	0.776	0.114**	0.019	0.180**	0.015
YIMYEXPair	1843	1572	2383	2661	-1523	-1079	12475	0.861	0.005	0.015	0.141**	0.014
PairYear	722	-2845	-2562	7078	6255	6567	11877	0.791	0.092**	0.012	0.198**	0.012

This table presents results on model selection using the Bayesian method, as well as the posterior means of EZ and EU effects on trade and their standard deviation (SE). The alternative model in the BLR test is the Baseline model. Burn-in time: 100,000; thin: 10; simulation: 50,000. ** significant at 1%.

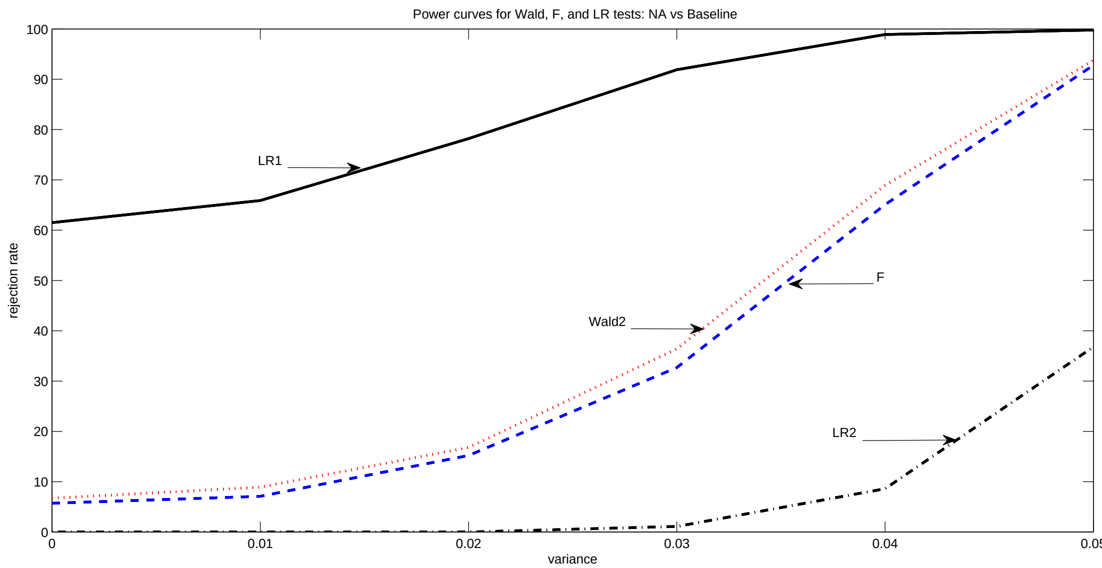

Figure 1: Power curves of three hypothesis tests: NA vs Baseline

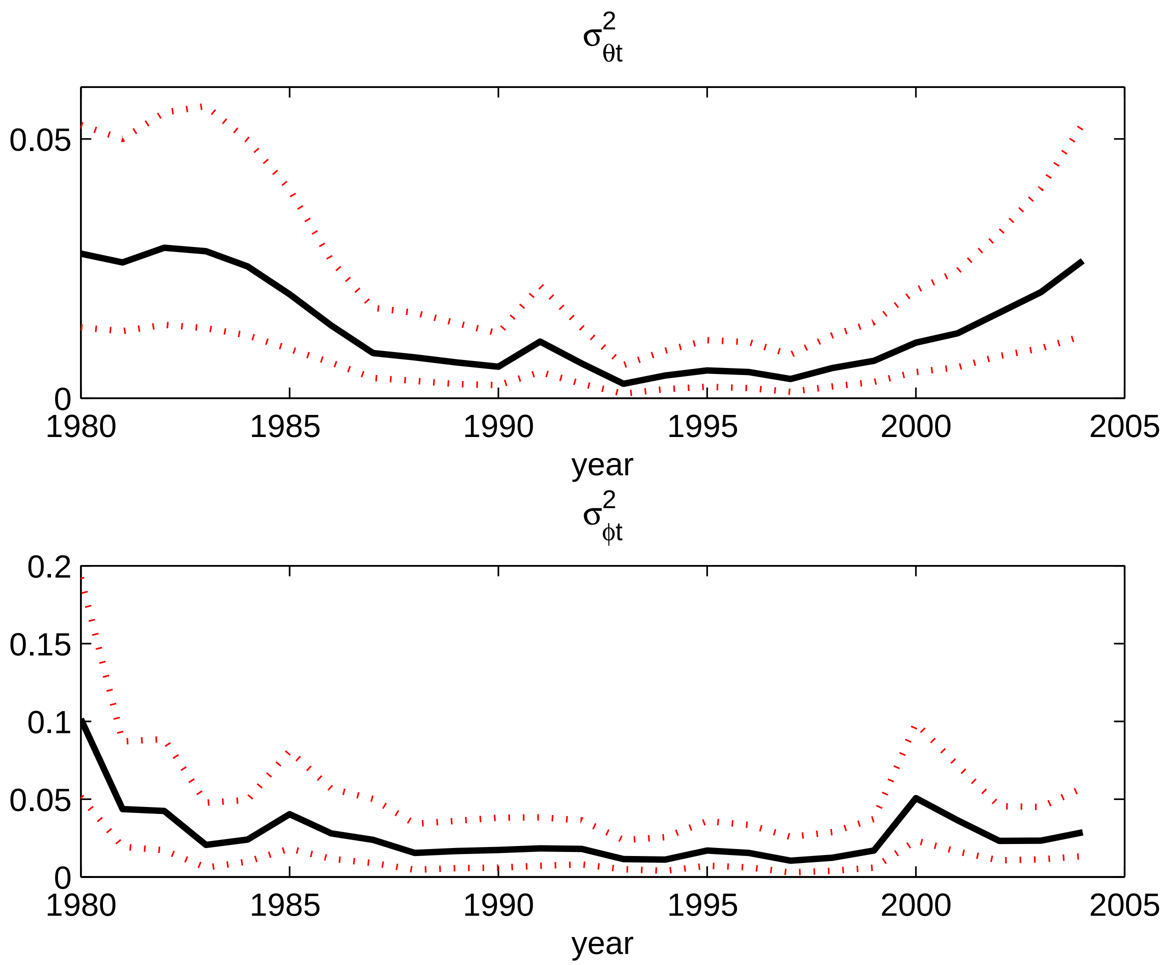

Figure 2: Posterior medians and the 95% credible intervals for $\sigma^2_{\theta t}$ and $\sigma^2_{\phi t}$ in the Baseline model

Appendices

A Gravity Models and Data

Two main choices for dependent variable have been considered by researchers in estimating the gravity equation: trade levels and ratios. The level dependent variable can be the log of bilateral (unidirectional) imports/exports, or the average/sum of imports and exports between countries, whereas the latter one suffers from the silver medal error proposed by Baldwin and Taglioni (2007). These trade flow data can be measured by current dollar or deflated by price index (US CPI). However, estimations with deflated trade values suffer from the bronze medal error shown in Baldwin and Taglioni (2007). The model with log of bilateral import levels is shown as below,

$$lim_t^{ik} = cons + lyy + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + \theta_t^i + \phi_t^k + \sum_{j=1}^{J} \gamma_{jt} g_{jt}^{ik} + \varepsilon_t^{ik}. \tag{10}$$

The dependent variable, lim_t^{ik}, is the log of bilateral import levels, $log(IM_t^{ik})$, which is determined by heterogeneous preferences ($\delta^{ik} \equiv log(\alpha^{ik})$) , the product of importers' expenditures and exporters' outputs ($lyy \equiv log(EXP_t^i * OUT_t^k)$), trade costs ($g_{jt}^{ik} \equiv log(\tau_t^{ik})$) and fixed effects. The trade costs g_{jt}^{ik} include log of distance, dummies for border, common language, land-lock, Eurozone and European Union. With symmetric conditions, $\tau_t^{ik} = \tau_t^{ki}$ and $\alpha^i(k) = \alpha^k(i)$, trade balance for each country is zero and total output is equal to total expenditure, $EXP_t^i = OUT_t^i$, so that $lyy = log(OUT_t^i * OUT_t^k)$ if replacing expenditure EXP_t^i using output OUT_t^i.

This type of estimation has a potential endogeneity problem because the economic mass data "lyy" are included in the explanatory variables. Therefore, researchers use the second choice: the log of the bilateral import ratio— imports divided by the product of the importer's expenditure and exporter's output as in Anderson and Van Wincoop (2003) with cross-section data. This estimation restricts the unit effect of economic mass variables on bilateral trade. The model using bilateral import ratios in Anderson and Van Wincoop (2003) and Aviat and Coeurdacier (2007) is given below,

$$limr_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + \theta_t^i + \phi_t^k + \sum_{j=1}^{J} \gamma_{jt} g_{jt}^{ik} + \varepsilon_t^{ik}, \tag{11}$$

where the dependent variable is defined as

$$limr_t^{ik} = log(\frac{IM_t^{ik}}{OUT_t^i * OUT_t^k}) \text{ or } log(\frac{IM_t^{ik}}{EXP_t^i * OUT_t^k}).$$

The version of bilateral import ratios may be non-stationary with long panel data.

Hence, as in section 3 of Guo (2015), this paper uses the import ratios in equation (4) of section 2.1

$$limr_t^{ik} = log(\frac{IM_t^{ik} * WOUT_t}{OUT_t^i * OUT_t^k}).$$

The data contain 22 OECD countries. There are fourteen countries in EU by 1995—AUT, BEL-LUX, DEU, DNK, ESP, FIN, FRA, GBR, GRC, IRL, ITA, NLD, PRT, and SWE, among which four countries did not join in the EZ in 2000—DNK, GBR, GRC, and SWE. Another eight countries—AUS, CAN, CHE, JPN, USA, ISL, NOR, and NZL— do not belong to EU. The data source is listed as below,

1) Current dollar value of bilateral import/export data: IMF DOTS.

2) Current dollar value of GDP: WDI and IMF DOTS (robustness check).

3) Current dollar value of private consumption expenditure: World Bank's World Development Indicators (WDI).

4) Bilateral trade costs variables: distance, dummies for border connection, landlock, and common language are taken from CEPII. Geodesic (great circle) distances are measured as kilometers between capital cities.[42]

5) EU and EZ dummies: constructed by author following the dates of countries' participation in the European Union and Eurozone respectively.

Compared with the above LS results, Table B.1 provides results estimated by MLE, assuming i.i.d. normally distributed country pair random effects. The MLE results provide a robustness check on the variation of EZ and EU effects due to different choices of fixed effects. On average, the MLE results show the EZ effect on increasing the import ratio is 7.9%, compared with 25.4% using the LS in Table 2. The magnitude of the EZ effect depends on the choice of time-varying or time-invariant country fixed effects. Estimations with time-varying fixed effects in columns of "Baseline", "YIMYEXPair", "YIMYEX", "YNAPair", and "YNA", do not provide evidence to support a significant EZ effect, but show a significant effect for EU membership (15% more). The model with year and pair dummies ("PairYear") used by Micco et al. (2003) show a 12% increase in import ratio due to the currency union. In contrast, models with time-invariant country fixed effects in columns of "NA", "NAYear", "IMEX" and "IMEXYear", shows that both EZ and EU variables significantly affect imports, around 21% and 25 % more respectively. If we use the simple difference-in-difference method (DID) and take year 1999 as the breaking point, estimations with time-invariant country fixed effects show that using euros can increase 24% import ratios. After controlling for the time-varying country fixed effect, however, the EZ effect from DID drops to 9%.

[42]http://www.cepii.franglaisgraph /bdd/distances.htm

B Monte Carlo Simulations

B.1 Type I Error under Homoscedastic Errors

All models with different groups of dummy variables are nested in the baseline model. We use the baseline model (the alternative) and the model "IMEX" (the null) as an example to illustrate the Monte Carlo simulations for size distortion presented in table 4

$$w_t^{ik} = cons + \beta_{EZ}EZ + \beta_{EU}EU + \delta^{ik} + (\widetilde{\theta_t^i} + \widetilde{\theta^i}) + (\widetilde{\phi_t^k} + \widetilde{\phi^k}) + \sum_{j=1}^{3}\gamma_{jt}g_j^{ik} + \varepsilon_t^{ik},$$

where $\theta_t^i = \widetilde{\theta_t^i} + \widetilde{\theta^i}$ and $\phi_t^k = \widetilde{\phi_t^k} + \widetilde{\phi^k}$. In order to obtain the model "IMEX", we need to impose the following 1448 restrictions on the baseline model: $\delta^{ik} = 0$, $\widetilde{\theta_t^i} = 0$, and $\widetilde{\phi_t^k} = 0$. All simulations are performed 1000 times.

1) Obtain the coefficients B_0 ($\widetilde{\theta^i}$, $\widetilde{\phi^k}$, β_{EZ}, β_{EU}, γ_{jt} and the constant intercept) and variance σ_0^2 ($=var(\epsilon_0)$) based on the model $y = XB_0 + \epsilon_0$. We estimate the coefficients from the model "IMEX" using the real data shown in appendix A. The dependent variable is import ratio $log\left(\frac{(1+IM_t^{ik})*WOUT_t}{EXP_t^i*OUT_t^k}\right)$. The coefficients on trade costs γ_{jt} are listed in table 2. The variance is the mean of the squared residual.

2) Simulate the dependent variables $\hat{y}$ for 1000 times given B_0, σ_0^2, and the covariates X from the model "IMEX". The random sample comes from the random draws of the error term.

3) Fit the simulated $\hat{y}$ using both the null and alternative models ($y = XB_m + \epsilon_j$ and $m \in \{n, a\}$, the subscript"n" and "a" is represented the the null and alternative models respectively.) and obtain the estimated coefficients ($\hat{B_n}$ and $\hat{B_a}$) and variance ($\hat{\sigma_n^2}$ and $\hat{\sigma_a^2}$) for 1000 times assuming i.i.d.

4) Calculate the statistics for the LR test and rejection rate (size). We use the formula

$$LR1 = N * \left[log(\hat{\sigma_n^2}) - log(\hat{\sigma_a^2})\right]$$

to calculate the statistic for the LR test,"LR1", where $\hat{\sigma_j^2} = RSS_j/N$ and RSS_j is the residual sum of squares of model m. Following Italianer (1985), the "LR2" adjusts the dimensions of the models (footnote 22); that is

$$LR2 = LR1 * (N - r - K_n/2) = LR1 * (11550 - 1448 - 0.5 * 48)/11550.$$

The statistic is chi-squared distributed with 1448 degrees of freedom and the critical value at 5% significant level is 1537.639. The size is the percentage of rejecting the null model "IMEX" with 1000 simulations when the null model "IMEX" is true.

5) The power curve. The difference between model "IMEX" and the baseline model includes the $\sigma_{\widetilde{\theta}}$ and $\sigma_{\widetilde{\phi}}$, and σ_{δ}. For example, $\widetilde{\theta_t^i} = \theta_t^i - \widetilde{\theta^i}$ with zero mean and $var\left(\widetilde{\theta_t^i}\right) = \sigma^2_{\widetilde{\theta}}$. The standard deviation $\sigma_{\widetilde{\theta}}$ is equal to zero in model "IMEX"; so do $\sigma_{\widetilde{\phi}}$ and σ_{δ}. By increasing the $\sigma^2_{\widetilde{\theta}}$, $\sigma^2_{\widetilde{\phi}}$, and σ^2_{δ} by the same scale, i.e. 0.01, the rejection rate of the null model "IMEX" goes up. The power curve plots the rejection rate along with the increasing variance.

6) Calculate the statistic for the F test. The F test can be used to compare models with homoscedastic error terms. In table 4 the F test statistic for null and alternative models is calculated as

$$F = \frac{(RSS_n - RSS_a)/(K_a - K_n)}{RSS_a/(N - K_a)},$$

where RSS is the residual sum of squares and K is the number of estimated coefficients. The F statistic has the degrees of freedom ($K_a - K_n = 1496 - 48 = 1448$ and $N - K_a = 11550 - 1496 = 10054$) and the critical value for the significant level 5% is 1.067.

7) Calculate the statistic for the $Wald$ test. The constraint matrix R_n can be constructed using the conditions $\delta^{ik} = 0$, $\widetilde{\theta_t^i} = 0$, and $\widetilde{\phi_t^k} = 0$. The Wald statistic is calculated as follows

$$Wald = (R_n \hat{B}_a)' * [R_n var(\hat{B}_a) R_n']^{-1} * (R_n \hat{B}_a),$$

where $var(\hat{B}_a) = \hat{\sigma_a^2} * (X_a' * X_a)^{-1}$. We use the consistent and biased estimate $\hat{\sigma_a^2} = RSS_a/N$ to calculate the Wald1 statistics, and use consistent and unbiased estimate $\hat{\sigma_a^2} = RSS_a/(N - K_a)$ to calculate (dimension adjusted) Wald2 statistics. The statistic is chi-squared distributed with degrees of freedom 1448 and the critical value at 5% significant level is 1537.639. The size is the percentage to reject the null model "IMEX" with 1000 simulations when the null model "IMEX" is true.

Two figures provide direct evidence for the biased distributions of the hypothesis tests due to the high dimensionality. We take the Comb. (1) as an example to illustrate the mechanism. Figure B.1 provides the chi-square densities with DF 1469 for three cases (appendix B.1): 1) the solid red line is the ideal theoretical kernel density, drawing 1000 observations from the Chi-square distribution directly; 2) the dashed black line plots the empirical LR kernel density (LR1) based on the null model "NA" and the baseline model using 1000 simulations ; 3) the blue dash-dot line plots dimension adjusted LR kernel density (LR2) based on Italianer (1985). The vertical red line is equal to 1559.3, the theoretical critical value (CV) at the 5% significant level. Similarly, figure B.2 portrays the densities for the Wald tests.

With DF 1469, the empirical Chi-squared distribution for the LR and Wald tests (the dash black lines) in Figure B.1 and Figure B.2 are biased compared to the ideal theoretical distribution (the red solid line). These biased asymptotical chi-square distributions occur because of the high dimensions in these models and a large number of constraints associated with the hypothesis tests. Using the conventional CV at the 5% significant level (1559.3), the empirical LR1 and Wald1 tests both have a large size distortion. In Figure B.1, the dash black LR1 line has a 62.1 % rejection rate on the null model "NA". A small adjustment proposed by Italianer (1985) (the weight is equal to $0.872 = \frac{11550-1469-0.5*27}{11550}$) shown in the dash-dot blue LR2 line reduces this large size to zero. A value, such as 1560, is changed into 1360.32 with the adjustment (the weight is equal to 0.872), which is no longer significant compared with the CV 1559.3. This simple adjustment does not work well for the LR test. Similarly in Figure B.2 for the Wald tests, the size is 97.6% for the Wald1 test, and decreases to 6.9% for the Wald2 test after the dimension adjustment (the weight is equal to $0.873 = \frac{11550-1469}{11550}$).

B.2 Errors with Heteroskedasticity and Autocorrelations

In table 3, we calculate the Wald statistics using Newey-West standard errors with 2 lags, robust to the heteroskedasticity and autocorrelation (HAC). The Monte Carlo simulations assume HAC error terms for a specific importer i and exporter k pair (462 pairs) and specify three parametric forms for the HAC. The conclusions on three hypothesis tests and size distortions (with misspecification or not) are robust to the choices on HAC.

The first HAC, "HAC1", in tables B.2-B.5 takes the form as below,

$$\epsilon_t^{ik} = g^{ik} + \nu_t^{ik} \qquad \nu_t^{ik} = b_\nu \nu_{t-1}^{ik} + \mu_t^{ik} \qquad var(g^{ik}) = \sigma_{g^{ik}}^2 \qquad var(\mu_t^{ik}) = \sigma_\mu^2$$

There is no contemporaneous correlation across country pairs. This parametric assumption considers the role of fixed effect in the variance covariance matrix Ξ $(= var(\epsilon))$. The matrix Ξ $(= var(\epsilon))$ is a block diagonal matrix with Ω^{ik} (462 pairs) for one specific importer i and exporter k pair and Ω^{ik} has the following form,

$$\Omega^{ik} = \sigma_{g^{ik}}^2 \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & 1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} + \frac{\sigma_\mu^2}{1-b_\nu^2} \begin{bmatrix} 1 & b_\nu & \cdots & b_\nu^{T-1} \\ b_\nu & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & b_\nu \\ b_\nu^{T-1} & \cdots & b_\nu & 1 \end{bmatrix}.$$

The second HAC, "HAC2", does not take the fixed affect into account and takes the form,

$$\epsilon_t^{ik} = b\epsilon_{t-1}^{ik} + \upsilon_t^{ik}, \text{ and } var(\upsilon_t^{ik}) = \sigma_{\upsilon^{ik}}^2$$

The variance Ξ is a block diagonal matrix with Ω^{ik}, where

$$\Omega^{ik} = \frac{\sigma^2_{\upsilon^{ik}_t}}{1-b^2}\begin{bmatrix} 1 & b & \cdots & b^{T-1} \\ b & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & b \\ b^{T-1} & \cdots & b & 1 \end{bmatrix}.$$

Considering both cases in HAC1 and HAC2 leads to the third HAC form, "HAC3", which takes the form,

$$\epsilon^{ik}_t = g^{ik} + \nu^{ik}_t \qquad \nu^{ik}_t = b_\nu \nu^{ik}_{t-1} + \mu^{ik}_t \qquad var(g^{ik}) = \sigma^2_{g^{ik}} \qquad var(\mu^{ik}_t) = \sigma^2_{\mu^{ik}}$$

The variance Ξ is a block diagonal matrix with Ω^{ik}, where

$$\Omega^{ik} = \sigma^2_{g^{ik}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & 1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} + \frac{\sigma^2_{\mu^{ik}}}{1-b^2_\nu}\begin{bmatrix} 1 & b_\nu & \cdots & b^{T-1}_\nu \\ b_\nu & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & b_\nu \\ b^{T-1}_\nu & \cdots & b_\nu & 1 \end{bmatrix}.$$

With HAC2, the size distortions are larger than those with HAC1 and conclusions remain. With HAC3, the size distortions are close to either HAC1 or HAC2 depending on the null hypothesis models. Simulations with only heteroskedastic errors without serial correlation gives similar results too. The results with HAC1 only are reported in the paper to save space.

The last HAC, "HAC4", in tables B.2-B.5 is White-type heteroscedastic. The variance Ξ_0 is diagonal matrix with $(\sigma^{ik})^2$ $(=\varepsilon^{ik}_t)$ for a specific importer-exporter group. Zeros are for all non-diagonal elements. This is assumed in the Bayesian model specification.

B.3 Monte Carlo Simulations for the Misspecified Case

Since the data show HAC, we consider four HAC forms in the paper (appendix B.2). The first three HACs obtain qualitatively similar results and we mainly focus on "HAC1". The HAC1 takes the form as below,

$$\epsilon^{ik}_t = g^{ik} + \nu^{ik}_t \qquad \nu^{ik}_t = b_\nu \nu^{ik}_{t-1} + \mu^{ik}_t \qquad var(g^{ik}) = \sigma^2_{g^{ik}} \qquad var(\mu^{ik}_t) = \sigma^2_{\mu}$$

There is no contemporaneous correlation across county pairs.[43] This parametric assumption considers the heterogeneous fixed effect in the variance covariance matrix Ξ $(= var(\epsilon))$, which is a block diagonal matrix with Ω^{ik} (462 pairs) for one specific country pair (importer i and exporter k) and Ω^{ik} has the following form,

[43]Models with contemporaneous correlation across county pairs can be estimated by spacial regression.

$$\Omega^{ik} = \sigma^2_{g^{ik}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & 1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} + \frac{\sigma^2_\mu}{1-b^2_\nu} \begin{bmatrix} 1 & b_\nu & \cdots & b_\nu^{T-1} \\ b_\nu & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & b_\nu \\ b_\nu^{T-1} & \cdots & b_\nu & 1 \end{bmatrix}.$$

The HAC4 is White-type heteroscedastic. The variance Ξ_0 is diagonal matrix with $(\sigma^{ik})^2$ $(=\varepsilon_t^{ik})$ for a specific importer-exporter group. Zeros are for all non-diagonal elements. This is assumed in the Bayesian model specification.

In simulations, we consider the case of misspecifications on the error structure, which the true errors have HAC but the estimations do not control for HAC (assuming spherical errors), noted as HAC(M). In table B.2, compared with the case of "HOMO", the case of misspecification HAC1(M) has overwhelmingly higher rejection rates for both the null and alternative models; most of the values are 100%. Because of the misspecification, both the null and alternative models cannot estimate consistent coefficients and are rejected easily by the Wald test. Surprisingly, some of the rejection rates are very small, and several are even less than 5% for the case of misspecification HAC4(M). Particularly, for the last five combinations the rates in rejecting the true null model are higher than those in rejecting the alternative model though on average the former one is smaller than the later one. In sum, the country-pair specific variance structure remarkably influences the estimations of the coefficients. Without controlling for the true HAC, both the null and alternative models cannot obtain consistent coefficients except few specifications in HAC4(M).

We continue to use the combination of the null model "IMEX" and the alternative baseline model as an example to illustrate the Monte Carlo simulations on the misspecified case (HAC1(M)) in tables B.2-B.5. The misspecification refers (no controlling for HAC) to the fact that the simulated data have HAC in the error term, but the regressions ignore the HAC and assume homoscedastic error terms to estimate the variance-covariance matrix of the coefficients.

1) Obtain the coefficients B_0 and variance Ξ_0 $(= var(\epsilon_0))$ based on $(y = XB_0 + \epsilon_0)$. The (estimated) variance covariance matrix Ω^{ik} has the form either HAC1 or HAC2 or HAC3 in appendix B.2.

2) Simulate the dependent variables $\hat{y}$ for 1000 times given B_0, Ξ_0, and covariates in the null model "IMEX". The random sample comes from the random draws of the error term.

3) Fit the simulated data into models, same as in the appendix B.1 assuming homoscedasticity.

4) Calculate the statistics for three tests, including "LR1" and "LR2" for the LR test, "F" for the F test and "Wald1" and "Wald2" for the Wald test. Then obtain the rejection rates (size) for each test, which follows the appendix B.1 assuming homoscedasticity.

B.4 Monte Carlo Simulations for the Wald Test on B_0

Tables 5, B.4, B.3 and B.5 show the Wald hypothesis tests (Wald2) on the estimated coefficients from both the null (n) and alternative (a) with respect to the artificial B_0. The "HAC1(M)" refers to the misspecification case discussed in appendix B.3 without controlling for the HAC1. Particularly, tables B.3 and B.5 provide details for EZ and EU effects, a subset of the B_0. We continue using the same example to illustrate the simulation.

1) Obtain the coefficients B_0 and variance, either homoscedasticity σ_0^2 or heteroskedasticity Ξ_0 based on ($y = XB_0 + \epsilon_0$) as in append B.1 and B.3.

2) Simulate the dependent variables $\hat{y}$ for 1000 times given B_0,σ_0^2 or Ξ_0, and covariates in the null model (n). The random sample comes from the random draws of the error term.

3) Fit the simulated data into models, same as in the appendix B.1 if with homoscedasticity. With HAC, we transform the $\hat{y}$ by multiply the cholesky decomposition of the variance matrix Ξ_0, which has no misspecification. The case with HAC and misspecification is the fact that the simulated data have HAC in the error term, but the regressions assume homoscedastic error terms.

4) Calculate the statistic for the Wald test in table 5 for both null and alternative models. The null hypothesis in the Wald test is $H0 : \hat{B}_m = B_0$ for $m \in \{a, n\}$, and Wald statistic (Wald2) is calculated by following formula

$$Wald = (\hat{B}_m - B_0) * \left[\widehat{var(\hat{B}_m)}\right]^{-1} * (\hat{B}_m - B_0),$$

where $\widehat{var(\hat{B}_m)} = \hat{\sigma_m^2} * (X_m' X_m)^{-1}$ and $\hat{\sigma_m^2} = RSS_m/(N - K_m)$. Then obtain the rejection rates (size) for the test under different assumptions of the error terms.

5) Obtain the rejection rates (size) in table B.3 for both null and alternative models based on the choice of the subset of the coefficients B_0.

6) Calculate the Wald statistics in table B.4. The null hypothesis is $H0 : \bar{\hat{B}}_m = B_0$ for $m \in \{a, c\}$, the statistic is calculated as

$$Wald = (\bar{\hat{B}}_m - B_0) * \left[\widehat{var(\bar{\hat{B}}_m)}\right]^{-1} * (\bar{\hat{B}}_m - B_0),$$

where the variance covariance matrix is

$$\widehat{var(\bar{\hat{B}}_m)} = \bar{\hat{\sigma_m^2}} * (X_m' X_m)^{-1}/1000.$$

with mean of the estimated variance $\bar{\hat{\sigma}}_m^2 = \overline{RSS}_m/(N - K_m)$ and mean of the sum of squared residual $\overline{RSS}_m$. The Wald statistic follows chi-squared distribution given degrees of freedom $K = 48$ and the critical value for the significant level 5% are 65.17.

B.5 Steps of MCMC via Gibbs

Markov Chain Monte Carlo Simulation via Gibbs Sampler has the following five steps:

Step 1. Give initial values for the variances, $\Sigma_{1,2}$ ($(\sigma^{ik})^2$, σ_p^2, $\sigma_{\theta t}^2$, $\sigma_{\phi t}^2$), and the second level parameters $\mathbf{B}_2$ (p, θ^i, ϕ^k);

Step 2. (update the first level parameters) Draw values from the posterior distributions for the first level parameters $\mathbf{B}_1$ ($cons, \beta_{EZ}, \beta_{EU}$, θ_t^i, ϕ_t^k, and δ^{ik}), given $\Sigma_{1,2}$ and $\mathbf{B}_2$;

Step 3. (update all the variance) Draw values from the posterior distributions for the variances $\Sigma_{1,2}$ ($(\sigma^{ik})^2$, σ_p^2, $\sigma_{\theta t}^2$, and $\sigma_{\phi t}^2$), given $\mathbf{B}_1$ in Step 2 and $\mathbf{B}_2$ in Step 1.

Step 4. (update the second level parameters) Draw values from the posterior distributions for the parameters $\mathbf{B}_2$, given $\mathbf{B}_1$ in Step 2 and $\Sigma_{1,2}$ in Step 3.

Step 5. Repeat from the second step until the chain converges.

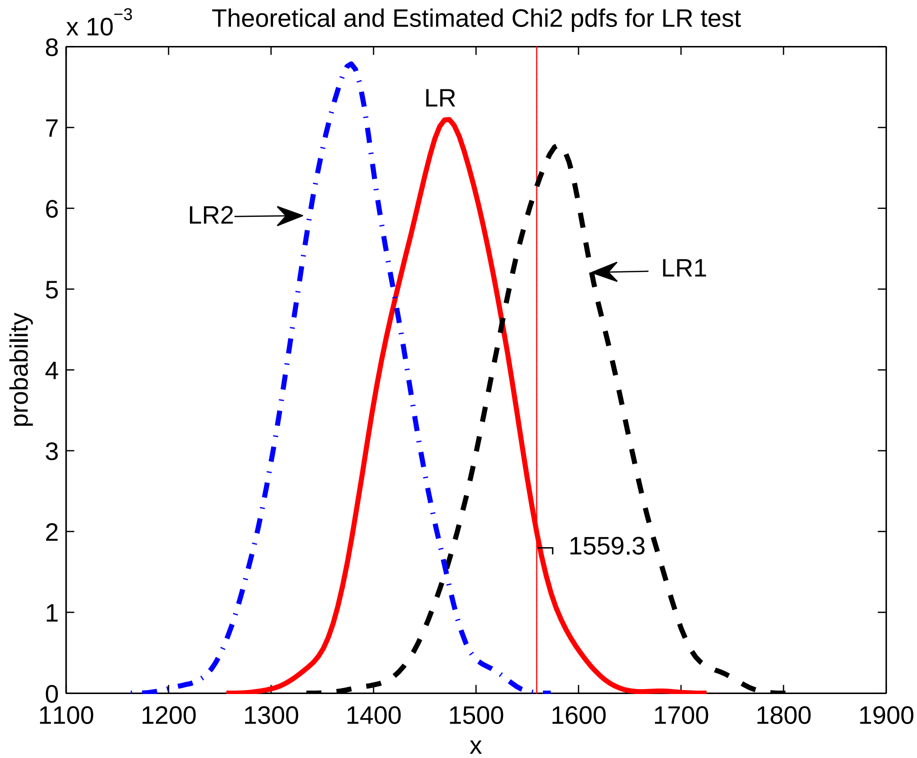

Figure B.1: Theoretical and empirical Chi-squared distributions with DF 1469

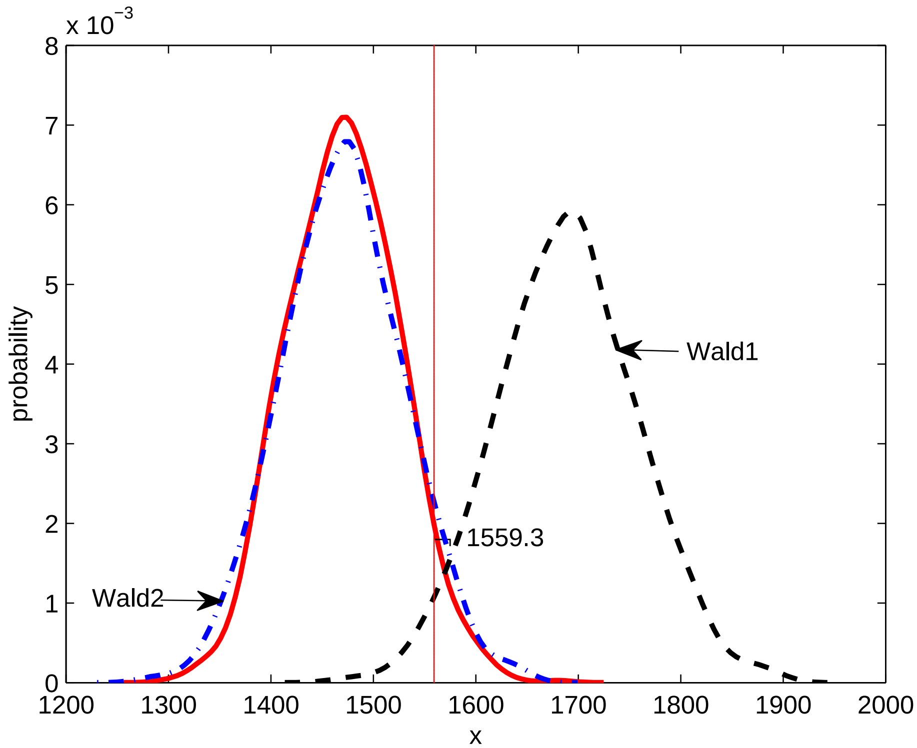

Figure B.2: Kernel densities for Wald1 and Wald2: NA vs Baseline

Table B.1: Eurozone effect and European Union effect on the log of bilateral import ratio by MLE: 1980-2004

Var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dummies	Baseline	NA	NAYear	YNA	YNAPair	IMEX	IMEXYear	YIMYEX	YIMYEXPair	PairYear
Time-varying	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Imp. & Exp.	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No
Nation	No	Yes	Yes	Yes	Yes	No	No	No	No	No
Year	No	No	No	Yes	No	No	Yes	No	No	Yes
Pair	Asym.	No	No	No	Sym.	No	No	No	sym.	Sym.
EZ	-0.004	0.158***	0.141***	0.001	-0.004	0.158***	0.142***	0.002	-0.004	0.139***
	0.023	0.016	0.017	0.025	0.025	0.016	0.017	0.023	0.023	0.017
EU	0.227***	0.191***	0.163***	0.227***	0.227***	0.191***	0.163***	0.226***	0.227***	0.163***
	0.018	0.014	0.014	0.020	0.020	0.014	0.014	0.018	0.018	0.015
log(dist)		-0.892***	-0.897***	-0.890***		-0.892***	-0.897***	-0.890***		
		0.052	0.052	0.053		0.047	0.047	0.047		
contig.		0.211*	0.212*	0.210*		0.211*	0.212*	0.210*		
		0.122	0.122	0.123		0.109	0.109	0.109		
comlang.		0.426***	0.422***	0.430***		0.426***	0.422***	0.430***		
		0.104	0.104	0.104		0.092	0.092	0.092		
locked_EX										-1.236***
										0.372
locked_IM										-1.409***
										0.372
Obs	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550	11,550
R2-p	0.968	0.115	0.188	0.409	0.445	0.126	0.199	0.629	0.682	0.224
sigma_u	0.002	0.542	0.543	0.545	0.367	0.481	0.481	0.484	0.268	0.364
sigma_e	0.255	0.328	0.317	0.286	0.286	0.328	0.317	0.260	0.260	0.317
rho	0.008	0.733	0.746	0.784	0.621	0.683	0.697	0.776	0.515	0.569

See table note in Table 2. MLE assumes the random effects, i.e. $\epsilon_t^{ik} = u^{ik} + e_t^{ik}$, $var(u^{ik}) = \sigma_u^2$ and $var(e_t^{ik}) = \sigma_e^2$.

Table B.2: Size of the Wald test (Wald2), H0: $\hat{\mathbf{B}}_m = \mathbf{B}_0$, where $m \in \{n, a\}$

Comb.	Null (n)	Alternative (a)	DF	n			a		
				HOMO	HAC1(M)	HAC4(M)	HOMO	HAC1(M)	HAC4(M)
(1)	NA	Baseline	27	4.5	100	5.1	4	100	21.4
(2)	NAYear	Baseline	51	4.8	100	3.7	4.3	100	4.4
(3)	YNA	Baseline	555	4.7	48.7	8	6	100	11.3
(4)	YNAPair	Baseline	761	6	100	14.9	5.9	100	17.3
(5)	IMEX	Baseline	48	4.5	100	6.9	5.3	100	29.8
(6)	IMEXYear	Baseline	72	4.7	100	7.3	4.1	100	13.2
(7)	YIMYEX	Baseline	1080	5.7	45.3	10.4	6.2	100	22.1
(8)	YIMYEXPair	Baseline	1286	7.2	100	15.7	7.3	100	24.1
(9)	PairYear	Baseline	258	5	100	14.7	4.2	100	7.1
(10)	IMEX	YIMYEX	48	4.5	100	6.9	4.9	89.4	7.7
(11)	IMEXyear	YIMYEX	72	4.7	100	7.3	5	77.9	2.7
(12)	YNA	YIMYEX	555	4.7	48.7	8	6.2	44.7	6.2
(13)	YNAPair	YIMYEXPair	761	6	100	14.9	6.3	100	9.1

Appendix B.4 provides the calculations on the size distortions. The subscripts n and a represents the null and alternative models respectively. See appendix B.1 for the Monte Carlo simulations for the homoscedasticity case (the "HOMO" columns); see appendix B.2 and appendix B.3 for "HAC1(M)" and "HAC4(M)" columns.

Table B.3: Actual size of the Wald test (Wald2), H0: $\beta_{EZ}^{\bar{\hat{m}}} = \beta_{EZ}^{0}$ and $\beta_{EU}^{\bar{\hat{m}}} = \beta_{EU}^{0}$, where $m \in \{n, a\}$

Comb.	Null (n)	Alternative (a)	n			a		
			HOMO	HAC1(M)	HAC4(M)	HOMO	HAC1(M)	HAC4(M)
(1)	NA	Baseline	4.1	58.2	1.9	4.7	48.6	2.6
(2)	NAYear	Baseline	4.2	60	2.5	4.7	49.3	2.7
(3)	YNA	Baseline	5.3	74.2	4.9	4.7	48.6	2.7
(4)	YNAPair	Baseline	4.7	13.5	3.2	4.7	46.9	3.2
(5)	IMEX	Baseline	4.1	58.8	1.8	4.7	48	3.3
(6)	IMEXYear	Baseline	4.2	62.6	2.7	4.7	48.9	3.1
(7)	YIMYEX	Baseline	5.3	73.2	6.1	4.7	47.4	3
(8)	YIMYEXPair	Baseline	4.7	19.2	3.3	4.7	44.8	3.3
(9)	PairYear	Baseline	4.5	19.2	1.5	4.7	48.2	3.3
(10)	IMEX	YIMYEX	4.1	58.8	1.8	5.3	73.6	5.9
(11)	IMEXyear	YIMYEX	4.2	62.6	2.7	5.3	74.1	5.7
(12)	YNA	YIMYEX	5.3	74.2	4.9	5.3	74.1	4.8
(13)	YNAPair	YIMYEXPair	4.7	13.5	3.2	4.7	14.5	3.2

This table focuses on the EZ and EU coefficients particularly. See table notes in table B.2.

Table B.4: Wald statistics (Wald2), H0: $\bar{\hat{\mathbf{B}}}_m = \mathbf{B}_0$, where $m \in \{n, a\}$

Comb.	Null (n)	Alternative (a)	DF	CV(Chi2)	n			a		
					HOMO	HAC1(M)	HAC4(M)	HOMO	HAC1(M)	HAC4(M)
(1)	NA	Baseline	27	40.1	25.5	280.8	18.1	17.9	2356.4	14.4
(2)	NAYear	Baseline	51	68.7	55.9	290.7	51.6	35.7	2671.8	25.1
(3)	YNA	Baseline	555	610.9	507.6	377.7	513.3	518.1	5349.2	503.3
(4)	YNAPair	Baseline	761	826.3	706.4	7412.1	691.7	744.5	22973.4	730.4
(5)	IMEX	Baseline	48	65.2	43.1	618.6	41.7	54.1	2626.1	43.7
(6)	IMEXYear	Baseline	72	92.8	73.6	637.2	74.2	74.5	2682.9	58.3
(7)	YIMYEX	Baseline	1080	1157.6	1016.4	800.9	1021.3	1017	9141.7	1011.4
(8)	YIMYEXPair	Baseline	1286	1370.5	1215.2	7054.1	1182.2	1244.1	18610.6	1239
(9)	PairYear	Baseline	258	296.5	256.8	7093.3	221.4	268.3	14530.8	270.8
(10)	IMEX	YIMYEX	48	65.2	43.1	618.6	41.7	51.7	38.1	75.3
(11)	IMEXyear	YIMYEX	72	92.8	73.6	637.2	74.2	69.9	45.1	89.4
(12)	YNA	YIMYEX	555	610.9	507.6	377.7	513.3	516.1	465.4	499
(13)	YNAPair	YIMYEXPair	761	826.3	706.4	7412.1	691.7	714.9	7631.6	675.8

This table focuses on the mean of the estimated coefficients. See table notes in table B.2.

Table B.5: Wald statistics (Wald2), H0: $\beta^{\bar{\hat{m}}}_{EZ} = \beta^{0}_{EZ}$ and $\beta^{\bar{\hat{m}}}_{EU} = \beta^{0}_{EU}$, where $m \in \{n, a\}$

Comb.	Null (n)	Alternative (a)	CV(chi2)	n			a		
				HOMO	HAC1(M)	HAC4(M)	HOMO	HAC1(M)	HAC4(M)
(1)	NA	Baseline	6.0	5.7	4.0	2.2	1.1	7.4	0.2
(2)	NAYear	Baseline	6.0	5.0	6.4	1.1	1.1	8.1	0.2
(3)	YNA	Baseline	6.0	7.8	13.9	3.1	1.1	7.1	0.2
(4)	YNAPair	Baseline	6.0	1.1	3.0	1.0	1.1	8.3	1.0
(5)	IMEX	Baseline	6.0	5.7	2.2	4.0	1.1	6.9	0.2
(6)	IMEXYear	Baseline	6.0	5.0	4.0	2.5	1.1	7.3	0.1
(7)	YIMYEX	Baseline	6.0	7.8	11.6	4.8	1.1	8.2	0.1
(8)	YIMYEXPair	Baseline	6.0	1.1	2.8	1.1	1.1	6.1	1.1
(9)	PairYear	Baseline	6.0	0.4	1.0	1.1	1.1	9.1	0.7
(10)	IMEX	YIMYEX	6.0	5.7	2.2	4.0	7.8	4.8	4.7
(11)	IMEXyear	YIMYEX	6.0	5.0	4.0	2.5	7.8	5.9	4.7
(12)	YNA	YIMYEX	6.0	7.8	13.9	3.1	7.8	13.9	3.1
(13)	YNAPair	YIMYEXPair	6.0	1.1	3.0	1.0	1.1	3.1	1.0

This table focuses on the mean of the estimated EZ and EU coefficients, β_{EZ} and β_{EU}, particularly. See table notes in table B.2.

Table B.6: Posterior means and standard deviations of $\mathbf{B_2}$

Var.		mean	std	Var.		mean	std
p		1.669	0.098	σ_p^2		2.292	0.175
θ_i	AUS	1.110	0.052	ϕ_k	AUS	0.400	0.069
	AUT	-1.172	0.074		AUT	-0.123	0.071
	BEL	0.968	0.055		BEL	0.324	0.049
	CAN	0.679	0.060		CAN	0.190	0.049
	CHE	-0.560	0.046		CHE	0.640	0.051
	DEU	-0.049	0.043		DEU	0.309	0.048
	DNK	-0.151	0.091		DNK	0.277	0.056
	ESP	-0.092	0.061		ESP	-0.877	0.067
	FIN	0.375	0.146		FIN	0.840	0.064
	FRA	0.040	0.062		FRA	-0.262	0.055
	GBR	0.908	0.049		GBR	0.848	0.056
	GRC	-0.770	0.141		GRC	-0.926	0.092
	IRL	-0.279	0.143		IRL	1.383	0.070
	ISL	2.080	0.105		ISL	-1.348	0.152
	ITA	0.374	0.037		ITA	0.103	0.053
	JPN	1.710	0.042		JPN	0.682	0.062
	NLD	0.688	0.065		NLD	0.301	0.061
	NOR	0.073	0.079		NOR	-0.341	0.094
	NZL	0.332	0.428		NZL	0.299	0.414
	PRT	-0.071	0.120		PRT	-0.543	0.067
	SWE	-0.099	0.065		SWE	1.051	0.055
	USA	omitted	omitted		USA	0.218	0.047