# Constrained Poisson Pseudo Maximum Likelihood Estimation of Structural Gravity Models

Michael Pfaffermayr

August 15, 2017

**Abstract**

This paper reconsiders the estimation of structural gravity models, introducing a constrained projection based Poisson pseudo maximum likelihood estimation procedure (constrained PPML) similar to Heyde and Morton (1993) and Falocci, Paniccià and Stanghellini (2009). The constrained PPML approach provides tests and confidence intervals for counterfactual predictions that are unavailable under the commonly used PPML approach with exporter and importer dummies. The paper establishes the asymptotic distribution of the constrained PPML estimator as well as that of the comparative static predictions and the implied percentage changes. Monte Carlo simulations provide encouraging results for medium sized samples. The estimation procedure is applied to a structural gravity model of 59 countries providing standard errors and confidence intervals of the impact of common language, contiguity and country borders on bilateral trade and welfare.

Keywords: Constrained Poisson Pseudo Maximum Likelihood Estimation; International Trade, Gravity Equation, Structural Estimation
JEL: F10, F15, C13, C50

# 1 Introduction

The analysis of the impact of trade costs on bilateral trade flows, like tariffs or trade costs related to geographical distance or country borders, is commonly based on a structural gravity model (see Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Bergstrand, Egger and Larch, 2013 and Allen, Arkolakis and Takahashi, 2014).[1] The structural gravity model is able to establish consistency of the aggregated bilateral trade flows with the observed country specific gross production and expenditure figures and allows to derive theory consistent comparative static estimates of counterfactuals that account for general equilibrium effects.

A standard approach in a cross-section is to estimate the gravity model in levels with fixed exporter and importer effects using Poisson pseudo maximum likelihood (dummy PPML) as suggested by Santos Silva and Tenreyro (2006). Arvis and Shepherd (2013) and Fally (2015) point out that under a Poisson specification with exporter and importer dummies the predicted trade flows add up to production values and expenditures, respectively, if trade flows are fully observed. Fernandez-Val and Weidner (2016) show that dummy PPML is among the few non-linear two-way models whose slope parameters can be estimated without asymptotic bias. However, dummy PPML is of limited use for the prediction of counterfactuals. Since in a structural gravity model exogenous changes in trade barriers affect multilateral resistances, the exporter and importer effects adjust endogenously, and they change with sample size. So the data generating process (DGP) of the structural gravity model forms a triangular array ruling out standard bootstrap approaches of estimating the standard errors of (counterfactual) predictions.

The set-up of Anderson van Wincoop (2003), who use a restricted non-linear least squares estimator in a log specification, is also applicable in a PPML setting. In their approach, and under constrained PPML as well, it is assumed that the system of multilateral resistances holds in expectation at true structural parameters given observed production and expenditures of the countries. This assumption turns out very useful for both the estimation of structural parameters and the prediction of counterfactuals without further restricting comparative static analysis.

Constrained PPML applies the projection based constrained quasi-likelihood estimator proposed by Heyde and Morton (1993) and Falocci, Paniccià and Stan-

[1] Anderson and van Wincoop (2003) envisage an endowment economy with products differentiated by country origin and bilaterally balanced trade flows. Bergstrand, Egger and Larch (2013) assume that trade is balanced multilaterally, rather than bilaterally, and resort to the Krugman model to specify the structural gravity equation. However, they emphasize that other models, e.g., Eaton and Kortum's (2002) Ricardian model, are observationally equivalent and thus covered by their specification as well. Allen, Arkulakis and Takahashi (2014) introduce an axiomatic approach to formulate a general theory consistent structural gravity model that is compatible with many important theoretical approaches.

ghellini (2009). However in contrast to these contributions, the present approach explicitly allows the number of constraints to grow without bound as sample size increases. Su and Judd (2012) demonstrate that such a constrained optimization approach is equivalent to a nested fixed point procedure that first solves the system of multilateral resistances in an inner loop and estimates the structural model parameters in an outer loop, given the multilateral resistance terms (see also Egger and Staub, 2016). Constrained PPML can also be applied in case of missing trade flows if selection is based on observables as in Egger, Larch, Staub and Winkelmann (2011) and comes up with theory consistent estimates and predictions also in this case.

The proposed constrained PPML estimator for cross-section gravity models has three main advantages over dummy PPML. First, it allows to establish confidence intervals for counterfactual predictions that account for changes in multilateral resistances.[2] Second, constrained PPML always comes up with theory consistent predictions even if a number of trade flows are missing at random. Third, constrained PPML performs better than dummy PPML in small and medium sized samples. Below it is shown, and as well demonstrated by Monte Carlo simulations, that under dummy PPML the estimated variances of the structural parameters are affected by the incidental parameter problem and are asymptotically downward biased, while those derived under constrained PPML are not.

This paper derives the asymptotic distribution of the constrained PPML estimator for structural gravity models and, using the delta method, that of comparative static results. Monte Carlo simulation results show that the constrained projection estimator works reasonably well and comes up with correct inference in medium sized cross-sections. To illustrate the usefulness of the constrained PPML estimator, it is applied to assess impact of a common official language, contiguity and country borders on bilateral trade flows and welfare quantitatively.

# 2 The structural cross-section gravity model

Several approaches are available to estimate structural gravity models. Following Santos Silva and Tenreyro (2006), a popular choice is to analyze bilateral trade flows in levels rather than in logs using PPML. Formally, in a cross-section of $C$ countries the DGP of the structural gravity model in levels generates bilateral

[2] Actually, for solving the system of trade resistances one may use the dummy PPML estimates of the structural parameters. Since these estimates are consistent, the proposed confidence intervals can be calculated without implementing the more complicated, iterative estimation procedure of constrained PPML.

trade flows as

$$x_{ij,C} = Y_W t_{ij}^{1-\sigma} \kappa_{i,C} \Pi_{i,C}^{\sigma-1} \theta_{j,C} P_{j,C}^{\sigma-1} \eta_{ij} = Y_W e^{z'_{ij}\alpha + \beta_{i,C}(\alpha) + \gamma_{j,C}(\alpha)} \eta_{ij}. \tag{1}$$

Bilateral trade flows and the implicit solutions of the system of multilateral resistances depend on trade frictions modelled as $t_{ij}^{1-\sigma} = e^{z'_{ij}\alpha}$. $\kappa_{i,C}$ denotes the share of country $i$ in world production $Y_W$, while $\theta_{j,C}$ refers to country $j$'s expenditure as share of $Y_W$. In general, gross production and expenditures of a country may differ from each other ($\kappa_{i,C} \neq \theta_{j,C}$) so that trade may not be balanced at the country level. The disturbances $\eta_{ij}$ are assumed to be independently distributed with $E[\eta_{ij}|z_{ij}] = 1$, but possibly heteroskedastic. Multilateral multilateral resistances enter the model in normalized form as $e^{\beta_{i,C}(\alpha)} = \kappa_{i,C}\Pi_{i,C}(\alpha)^{\sigma-1}$ and $e^{\gamma_{j,C}(\alpha)} = \theta_{j,C}P_{j,C}(\alpha)^{\sigma-1}$, respectively. This notation emphasizes that $\Pi_{i,C}(\alpha)$ and $P_{j,C}(\alpha)$ are not parameters to be estimated, rather they depend on the parameter vector $\alpha$, referring to barriers of bilateral trade, and on the number of countries in the sample.

For estimation, the structural gravity model can be reformulated with additive disturbances:

$$s_{ij,C} = m_{ij}(\vartheta_C) + \varepsilon_{ij}, \quad \varepsilon_{ij} = m_{ij}(\vartheta_C)\,(\eta_{ij} - 1)\,, \tag{2}$$

where $m_{ij}(\vartheta_C) = e^{z'_{ij}\alpha + \beta_{i,C}(\alpha) + \gamma_{j,C}(\alpha)}$ and $\vartheta_C = [\alpha', \beta'_C(\alpha), \gamma'_C(\alpha)]'$ to abbreviate notation. Santos Silva and Windmeijer (1997) show that the multiplicative and additive error Poisson models are observationally equivalent if the explanatory variables are exogenous. Both lead to the same estimators as the estimation procedure is based on a conditional mean assumptions only. Under IV-estimation, this equivalence breaks down, however.

For the estimation of structural gravity models the econometric specification needs to be precise on the stochastic properties of the system of multilateral resistances. Following Anderson and van Wincoop (2003, p. 179, eq. 21), Anderson and Yotov (2010, pp. 2260, eq. 5-6) and Baier and Bergstrand (2009, p. 79, eq. 8) a deterministic approach assumes that the system of multilateral resistances holds in expectation with the countries' gross production and expenditures taken as given and treated as non-stochastic. This assumption is justified if gross production, expenditures and aggregate country specific exports to and imports from the world come from different sources (e.g. UNIDO's Industrial Statistics Database or OECD's STAN database) than bilateral trade flow data. These aggregates are typically collected in country specific censuses that are unrelated to the sources of bilateral trade flow data.[3]

$$\kappa_{i,C} = E\left[\sum_{j=1}^{C} m_{ij}(\vartheta_C) + \varepsilon_{ij}\right] = \sum_{j=1}^{C} m_{ij}(\vartheta_C) \quad (3)$$

$$\theta_{j,C} = E\left[\sum_{i=1}^{C} m_{ij}(\vartheta_C) + \varepsilon_{ij}\right] = \sum_{i=1}^{C} m_{ij}(\vartheta_C) \quad (4)$$

In this setting observed trade flows do not add-up to gross production and expenditures, especially if domestic within country trade is derived from aggregated country specific data (see Appendix H for details).

In contrast, Bergstrand, Egger, Larch (2013, p. 113, eq. 10) and Egger and Nigai (2015, p. 88, eq. 2.3 and 3.2) suggest a stochastic specification, establishing the system of multilateral resistances as

$$\kappa_{i,C} = \sum_{j=1}^{C} s_{ij,C} = \sum_{j=1}^{C} m_{ij}(\vartheta_C) + \varepsilon_{ij} \quad (5)$$

$$\theta_{j,C} = \sum_{i=1}^{C} s_{ij,C} = \sum_{i=1}^{C} m_{ij}(\vartheta_C) + \varepsilon_{ij}. \quad (6)$$

At given gross production and expenditure figures this assumption implies restrictions on the distribution of the disturbances, namely $\sum_{i=1}^{C} \varepsilon_{ij} = \sum_{j=1}^{C} \varepsilon_{ij} = 0$. This is the relevant case if domestic trade flows are derived as residuals, $s_{ii,C} = \kappa_{i,C} - \sum_{j=1, j\neq i}^{C} s_{ij,C}$, or if trade flows are calibrated to the system of multilateral resistances ex-ante such as in WIOD. To sum-up, the econometric specification of the structural gravity as laid out above covers both cases when imposing the adding-up constraints (3) and (4), and in the stochastic case additionally the restrictions on the disturbances.

In the absence of any trade barriers ($\alpha = 0$) it holds that $\Pi_{i,C}(0) = 1$ and $P_{j,C}(0) = 1$, while $e^{\beta_{i,C}(\alpha)} = \kappa_{i,C}$ and $e^{\gamma_{j,C}(\alpha)} = \theta_{j,C}$. Since the solution of the system of multilateral resistances is unique up to a constant, without loss of generality $\beta_{C,C}(\alpha)$ is normalized to 0. For estimation, trade flows are further normalized by world expenditures, i.e., $s_{ij,C} = x_{ij,C}/Y_W$ so that $\sum_{i=1}^{C}\sum_{j=1}^{C} s_{ij,C} = 1$ (see Allen, Arkolakis and Takahashi, 2014). The normalization implies that there is no constant in the model. Actually, without further structural assumptions on the DGP, e.g. on endowments in case of an endowment model or on labor markets and tech-

[3] Actually, trade flow data are based on international transactions of goods, while gross production, expenditures and aggregate exports and imports come from censuses that allocate these figures to industries based on the main activities of firms.

nology (Krugman, 1979), $Y_W$ remains unspecified. As the countries' production and expenditures are assumed to be exogenously given, $Y_W$ is given as well and it is assumed to grow at the rate of the number of country pairs, $C^2$. So without loss of generality $Y_W$ may be written as $c_W C^2$ for some constant $c_W$. Note, the data generating process for $s_{ij,C}$ depends on the number of countries and thus forms a triangular array as indicated by the index $C$.

Furthermore, the present approach allows some trade flows to be unobserved due to randomly missing data, while all of them enter the system of multilateral resistances. Specifically, under the assumption of selection on observables (see e.g. Egger, Larch, Staub and Winkelmann, 2011) the structural gravity model with missing trade flows may be seen as the second (outcome) part of a two-part model. Let $V$ denote the corresponding selection matrix, which is derived from the identity matrix by setting the diagonal elements to 1 if a trade flow is observed and 0 otherwise. In matrix form, the model can be then compactly formulated as

$$\begin{aligned} V s_C &= V(m(\vartheta_C) + \varepsilon) \\ D' m(\vartheta_C) - \theta_C &= 0, \end{aligned} \tag{7}$$

where $s_C = (s_{11,C}, ..., s_{CC,C})'$, $\theta_C = (\kappa_{1,C}, ...., \kappa_{C-1,C}, \theta_{1,C}, ..., \theta_{C,C})'$, $m(\vartheta_C) = (m_{11}(\alpha), ..., m_{CC}(\alpha))'$. Exporter and importer dummies are collected in the design matrix $D = [D_x, D_m]$ and $W = [Z, D]$ contains all right hand side variables including exporter and importer dummies.[4]

# 3 Dummy PPML and constrained PPML

In the following $\vartheta_{C,0} = [\alpha_0', \phi_C(\alpha_0)']'$ with $\phi_C(\alpha_0) = [\beta_C'(\alpha_0), \gamma_C'(\alpha_0)]'$ denotes the true parameter vector of the population model with dimension $(K + 2C - 1 \times 1)$. The corresponding constrained estimates are denoted by a hat. Constrained PPML uses the restricted Poisson log-likelihood:

$$\begin{aligned} \ln L^C(\vartheta_C|V) &= \frac{Y_W}{C^2} \sum_{j=1}^{C} \sum_{i=1}^{C} v_{ij} (s_{ij,C} (\ln (m_{ij}(\vartheta_C)) + \ln Y_W) - m_{ij}(\vartheta_C)) \\ &\quad + \lambda' \left(D' m(\vartheta_C) - \theta_C\right), \end{aligned} \tag{8}$$

where $\lambda$ denotes the $(2C - 1 \times 1)$ vector of Lagrange multipliers. Conditional on $V$, the score of the restricted Poisson pseudo-likelihood is given as (ignoring the

[4]Principally, zero trade flows can be accounted for by adjusting the system of trade resistances. For this define the selection matrix $S$, which is derived from the identity matrix by skipping all rows referring to zero trade flows. Then, the system of trade resistances can be written as $D'S'Sm(\vartheta_C) - \theta_C = 0$.

scaling factor $\frac{Y_W}{C^2}$ which does not depend on $\alpha$)

$$\begin{aligned}
\frac{\partial \ln L^C(\vartheta_C|V)}{\partial \alpha} &= Z'V(s_C - m(\vartheta_C)) + Z'M(\vartheta_C)D\lambda && (9)\\
\frac{\partial \ln L^C(\vartheta_C|V)}{\partial \phi_C} &= D'V(s_C - m(\vartheta_C)) + D'M(\vartheta_C)D\lambda && (10)\\
\frac{\partial \ln L^C(\vartheta_C|V)}{\partial \lambda} &= D'm(\vartheta_C) - \theta_C, && (11)
\end{aligned}$$

since

$$\frac{\partial D'm(\vartheta_C) - \theta_C}{\partial \vartheta_C} = W'M(\vartheta_C)D,$$

with $M(\vartheta_C) = diag(m(\vartheta_C))$. For the the unconstrained dummy PPML the score is given as

$$\begin{aligned}
\frac{\partial \ln L^U(\vartheta_C|V)}{\partial \alpha} &= Z'V(s_C - m(\vartheta_C)) && (12)\\
\frac{\partial \ln L^U(\vartheta_C|V)}{\partial \phi_C} &= D'V(s_C - m(\vartheta_C)) && (13)
\end{aligned}$$

with corresponding estimator by $\overline{\vartheta}_C$.

Arvis and Shepherd (2013) and Fally (2015) show that the PPML estimation procedure automatically guarantees that the predicted trade flows add-up to the trading countries' production and expenditures, respectively, if the model includes exporter and importer fixed effects and all $C^2$-country pair observations are used for estimation ($V = I_{C_2}$, in (13)). Under missing trade flows this property of PPML is lost, however.

Constrained PPML implies that predicted bilateral trade flows always add up to exporter production value and importer expenditures even if some trade flows are missing, since the constraint (11) is imposed. In contrast, from (13) it follows that under the dummy PPML $D'Vm(\overline{\vartheta}_C) = \theta_C + (D'Vm(\vartheta_{C,0}) - \theta_C) + D'\varepsilon$. Thereby, the assumed true DGP, $s_C = m(\vartheta_{C,0}) + \varepsilon$ and $D'm(\vartheta_{C,0}) = \theta_C$, has been inserted. Therefore, it holds that $D'Vm(\overline{\vartheta}_C) \neq \theta_C$ even at $D'\varepsilon = 0$, if some non-zero trade flows remain unobserved.

If data are fully observable and generated such that $D's_C = \theta_C$, e.g., in GTAP or WIOD, (10) implies that $\lambda = 0$ and the score reduces to

$$\frac{\partial \ln L^C(\alpha, \phi_C(\alpha))}{\partial \alpha} = Z'(s_C - m(\alpha, \phi_C(\alpha)), \tag{14}$$

where $\phi_C(\alpha)$ solves the system of multilateral resistances $D'm(\alpha, \phi_C(\alpha)) - \theta_C = 0$.

In this scenario, at true parameters the score of constrained PPML is given as $W'\varepsilon = (Z, D)'\varepsilon$ and the restriction $D'\varepsilon = 0$. Hence, the variance-covariance matrix of the score is singular and disturbances will not be independent by construction. The derivation of asymptotic distribution of $\widehat{\alpha}$ thus has to take account of this restriction and the implied pattern of dependence of the disturbances.

Following Falocci, Paniccià and Stanghellini (2009)[5] and Heyde and Morton (1993), constrained PPML estimation can be designed as an iterative projection based estimation procedure that solves the constrained ML-maximization problem defined in (9)-(11).

**Proposition 1** *(Constrained PPML): Assume iteration $r$ yields $\widehat{\vartheta}_{C,r}$ and define*

$$\begin{aligned} \widehat{F}_r &= D'M(\widehat{\vartheta}_{C,r})W \\ \widehat{G}_r &= W'VM(\widehat{\vartheta}_{C,r})W, \end{aligned}$$

*where $\widehat{G}_r$ is non-singular. Iteration $r+1$ obtains*

$$\begin{aligned} \widehat{\vartheta}_{C,r+1} &= \widehat{\vartheta}_{C,r} + \left(\widehat{G}_r^{-1} - \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) \\ &+ \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\theta_C - D'm(\widehat{\vartheta}_{C,r})\right). \end{aligned}$$

**Proof.** See the Appendix C. ■

**Remark 1** *Upon convergence, at $\widehat{\vartheta}_{C,r+1} = \widehat{\vartheta}_{C,r}$, the constrained PPML estimator guarantees the adding-up constraint to hold even in case of missing trade flows, since in this case $0 = \widehat{F}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) = \theta_C - D'm(\widehat{\vartheta}_{C,r})$. However, it does not force the estimated score to zero. Rather, upon convergence the estimator leads to*

$$\widehat{G}_r^{-1}\left(I_{K+2C-1} - \widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)\left(W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right)\right) = 0.$$

*where the involved protection matrix is of rank $K$.*

This estimator is easy to implement and may use the parameter estimates of the dummy PPML estimator as starting values. Su and Judd (2012) demonstrate that this constrained optimization approach yields the same estimates as the nested fixed-point procedure, which solves the system of multilateral resistances in an inner loop and estimates the structural model parameters at given multilateral resistance parameters in the outer loop.

[5]These authors develop a constrained iterative maximum likelihood to estimate a log-linear gravity model of tourism flows under accounting constraints.

Based on the restricted maximum likelihood (8) it is possible to derive the limit distribution of the constrained PPML estimate $\widehat{\alpha}$ that fully respects the general equilibrium restrictions and accounts for the functional dependence of the multilateral resistance terms on the slope parameters $\alpha$. The asymptotic variance-covariance of the constrained PPML estimator differs from that of the dummy PPML estimator under the assumption that the system of multilateral resistances holds in expectation. Specifically, Heyde and Morton (1993) demonstrate that in a setting without incidental parameters the asymptotic distribution of the projection based estimator is singular normal (see their Corollary, p. 758).

The present approach is more general, since it establishes the limit distribution of $\widehat{\alpha}$ by projecting out the importer and exporter effects similar to Fernandez-Val and Weidner (2016) and allowing the number of constraints, $2C-1$, to grow without limit. The set of assumptions to establish consistency of $\widehat{\alpha}$ and its limit distribution is stated in detail in Appendix A. Here, it suffices to mention that the assumptions maintain that $m_{ij}(\alpha)$ and $\theta_{i,C}$ converge to zero at rates $C^2$ and $C$, respectively, as the trade flows are normalized by world production so that $\sum_{j=1}^{C}\sum_{i=1}^{C} s_{ij,C} = 1$. This is guaranteed by assuming that expected trade flows are bounded from above and below, i.e., $c_a/C^2 < m_{ij}(\alpha, \phi_C(\alpha)) < (1-c_a)/C^2$ for a constant $c_a < 0.5$, an assumptions similar to Condition S made by Berry, Linton and Pakes (2007) in the analysis of (mixed) logit estimators for market share models. $\varepsilon_{ij}$ is assumed to be independently distributed with variance $\sigma_{ij}^2/C^4$.[6] Naturally, the derivation of the asymptotic distribution of the constrained PPML estimator has to account for that normalization. Lastly, one has to assume that under the normalization $\beta_{C,C} = 0$ the solution of the system of multilateral resistances is unique (which is demonstrated Appendix B), the absence of multicollinearity in the explanatory variables and the existence of several limiting matrices.

**Proposition 2** *Under the set of assumptions specified in Appendix A it holds that*

*(i) Both dummy and constrained PPML estimates of* $\alpha_0$ *are consistent, i.e.* $\overline{\alpha} \xrightarrow{P} \alpha_0$ *and* $\widehat{\alpha} \xrightarrow{P} \alpha_0$.

*(ii)* $C(\hat{\alpha} - \alpha_0) \xrightarrow{d} N\left(0, B_0^{-1} A_0 \Omega_\varepsilon A_0' B_0^{-1}\right)$,

*where* $\Omega_\varepsilon = diag(\sigma_{ij}^2)$, $A_0 \Omega_\varepsilon A_0' = p\lim_{C\to\infty} \frac{1}{C^2} A(\alpha^*)\varepsilon\varepsilon' A(\alpha^*)'$, $B_0 = p\lim_{C\to\infty} B(\alpha^*)$

[6]This assumption is for convenience. Actually, it is sufficient to assume that $E[\varepsilon_{ij}^2] = O(C^{-4})$.

*with $\alpha^*$ lying in between $\widehat{\alpha}$ and $\alpha_0$, element by element, and*

$$
\begin{aligned}
M(\alpha) &= diag(m_{ij}(\alpha, \phi_C(\alpha))) \\
G(\alpha) &= W'VM(\alpha)W \\
F(\alpha) &= D'M(\alpha)W \\
A(\alpha) &= C^2 \left[I_K, 0_{K\times 2C-1}\right] \left[I - F(\alpha)' \left(F(\alpha)G(\alpha)^{-1}F(\alpha)'\right)^{-1} F(\alpha)G(\alpha)^{-1}\right] W'V \\
B(\alpha) &= Z'V\left[M(\alpha) - M(\alpha)D(D'M(\alpha)D)^{-1}D'M(\alpha)\right]Z.
\end{aligned}
$$

*(iii) For estimation one uses $\widehat{B} = B(\widehat{\alpha}) \xrightarrow{p} B_0$ and $\frac{1}{C^2}A(\widehat{\alpha})diag(\widehat{\varepsilon}\widehat{\varepsilon}')A(\widehat{\alpha})' \xrightarrow{p} A_0\Omega_\varepsilon A_0'$.*

**Proof.** See the Appendices D and E. ■

The covariance matrix of $\hat{\alpha}$ is easy to calculate, once the results of the iterative estimation procedure as outlined in Proposition 1 are available. Upon convergence constrained PPML delivers the estimates of $M(\widehat{\alpha})$, $F(\widehat{\alpha})$ and $G(\widehat{\alpha})$. Plugging in the estimated residuals $\widehat{\varepsilon}$ one can use $\widehat{Var}\,(\hat{\alpha}) = \frac{C^2-1}{C^2}B(\widehat{\alpha})\left(\frac{1}{C^4}A(\widehat{\alpha})\ diag(\widehat{\varepsilon}\widehat{\varepsilon}')A(\widehat{\alpha})'\right)B(\widehat{\alpha})$ for inference in finite samples. Clearly, plugging in the consistent dummy PPML-estimators $\overline{\alpha}$ does the job as well.

**Remark 2** *Both the dummy and the constrained PPML estimates of $\alpha_0$ are consistent and asymptotically unbiased (see also the no bias result of Fernandez-Val and Weidner, 2016, for dummy Poisson ML-estimators).*[7] *However, the limit distribution of the dummy PPML estimator $\overline{\alpha}$ is different, if some trade flows are missing and the DGP actually obeys the restrictions imposed by the system of multilateral resistances. The reason is that the dummy PPML estimator neglects the functional dependence of the multilateral resistance terms on structural parameters and that the predicted trade flows do not add-up to production and expenditures in case of missing trade flows. In Appendix G it is shown that the limit distribution of the dummy PPML estimator is based on*

$$
B(\alpha) = Z'V\left[\overline{M}(\alpha) - \overline{M}(\alpha)D(D'V\overline{M}(\alpha)D)^{-1}D'V\overline{M}(\alpha))\right]VZ, \tag{15}
$$

*with $\overline{\phi}_C(\alpha)$ solving $D'V(s_C - m(\alpha, \overline{\phi}_C(\alpha))) = 0$ and $\overline{M}(\alpha) = M(\alpha, \overline{\phi}_C(\alpha))$. The limiting variance of the score of dummy PPML $A_0\Omega_\varepsilon A_0'$ makes use of*

$$
A(\alpha) = C^2 Z'\left[I_{C^2} - \overline{M}(\alpha)VD(D'V\overline{M}(\alpha)D)^{-1}D'\right]V. \tag{16}
$$

[7] Proofs of the following remarks are given in Appendix G.

*In contrast, under constrained PPML $\phi_C(\alpha)$ always solves $D'(s_C - m(\alpha, \phi_C(\alpha))) = 0$ and it follows (simplifying the expressions in Proposition 2) that*

$$
\begin{aligned}
B(\alpha) &= Z'V\left[M(\alpha) - M(\alpha)D(D'M(\alpha)D)^{-1}D'M(\alpha)\right]Z. && (17)\\
A(\alpha) &= C^2 Z'\left[I_{C^2} - M(\alpha)D\left(F(\alpha)G(\alpha)^{-1}F(\alpha)'\right)^{-1}F(\alpha)G(\alpha)^{-1}W'\right]V && (18)
\end{aligned}
$$

*This result illustrates that the difference in the limit distribution of $\widehat{\alpha}$ and $\overline{\alpha}$ arises from the usage of different projection matrices. While the dummy PPML estimator applies $I_{C^2} - \overline{M}(\alpha)VD(D'V\overline{M}(\alpha)D)^{-1}D'$ to project out the exporter and importer dummies, the constrained PPML estimator applies a different projection that does not involve the selection matrix $V$ and that forces the estimated parameters of the exporter and importer dummies to obey the restrictions of the system of multilateral resistances without error.*

*Constrained and dummy PPML lead to the same limit distribution of the estimated structural parameters under fully observed trade flows. In fact, in this case $\phi_C(\overline{\alpha}) = \phi_C(\widehat{\alpha})$ solving $D'(s_C - m(\widehat{\alpha}, \phi_C(\widehat{\alpha}))) = 0$ (see Appendix G for details on the derivation) and in both cases*

$$B(\alpha) = Z'\left[M(\alpha) - M(\alpha)D(D'M(\alpha)D)^{-1}D'M(\alpha)\right]Z, \tag{19}$$

*and*

$$A(\alpha) = C^2 Z'\left[I_{C^2} - M(\alpha)D(D'M(\alpha)D)^{-1}D'\right]. \tag{20}$$

*This finding mirrors the well known result of Arvis and Shepherd (2013) and Fally (2015).*

**Remark 3** *An important difference between the constrained PPML and the dummy PPML arises in the estimation of $Var(\hat{\alpha})$ and $Var(\overline{\alpha})$. Since dummy PPML requires the estimation of $2C - 1$ parameters of the importer and exporter dummies, $\widehat{Var(\overline{\alpha})}$ is asymptotically downward biased of order $C^{-1}$. Constrained PPML does not exhibit this incidental parameter problem as the exporter and importer dummies are treated as functions of $\alpha$. (See Chesher and Jewitt (1987), Cribari–Neto, Ferrari and Cordeiro (2000), and Imbens and Kolesar (2016)). Appendix G provides a proof for the case of fully observed trade flows). Monte Carlo evidence shown below confirms this finding.*

**Remark 4** *If trade flows are fully observed and data are generated such that $D's_C = \theta_C$, as in GTAP or WIOD, there is the additional implicit restriction $D'\varepsilon = 0$. Moreover, it immediately follows that the score at true parameters reduces to $C^2 Z'\varepsilon$. By construction the disturbances will not be independent in this case. More importantly, the variance-covariance matrix of their scores is singular* *and the limit distribution of $\bar{\alpha}$ and $\widehat{\alpha}$, derived above is inappropriate. To account for this singularity a simple specification would be to assume that $C^2\varepsilon_{ij}$ is independently distributed as $(0, \sigma_{ij}^2)$ for $i \neq j$ and $i \neq C$, while $\varepsilon_{ii} = -\sum_{j=1,j\neq i}^{C} \varepsilon_{ij}$ for $i < C$ and $\varepsilon_{Cj} = -\sum_{j=1,j\neq C}^{C} \varepsilon_{Cj}$, $j = 1, ..., C$. In this case, data can be partitioned so that $\varepsilon = (\varepsilon_R', \varepsilon_U')'$, where $\varepsilon_R$ includes $\varepsilon_{ii}$, $i = 1, .., C-1$ and $\varepsilon_{Cj}$, $j = 1, .., C$, while $\varepsilon_U$ comprises the remaining unconstrained disturbances. The matrices $D$ and $Z$ are partitioned in the same way. Thereby, $D_R$ is an invertible $2C-1 \times 2C-1$ matrix, while $D_U$ has dimension $2C-1 \times (C-1)^2$. Then one can decompose the restrictions accordingly as*

$$\begin{aligned} D'\varepsilon &= D_R'\varepsilon_R + D_U'\varepsilon_U = 0 \\ \varepsilon_R &= -D_R'^{-1} D_U' \varepsilon_U, \end{aligned} \tag{21}$$

*and apply Proposition 2 with $A(\alpha) = C^2 \left[ Z_U' - Z_R D_R'^{-1} D_U' \right]$ and $\Omega_\varepsilon = E[\varepsilon_U \varepsilon_U']$.*

**Remark 5** *Instead of assuming heteroskedastic disturbances, one may use the clustering approach of Cameron, Gelbach and Miller (2011) to account for dependence of the disturbances within exporters and importers, e.g., induced by unobserved random exporter and importer effects (see also Egger and Tarlea, 2015, for an application in a panel framework to gravity models). In this case, one defines selector matrices that take the value of 1 if any two observations belong to the same cluster of exporting or importing countries, respectively. Using $D_x$ to select the exporter specific and $D_m$ the importer specific cluster and denoting the Hadarmard element-wise product by $\circ$, one obtains under this more general assumption on the disturbances*

$$A_0 \Omega_\varepsilon A_0 = p \lim_{C\to\infty} \tfrac{1}{C^3} A(\alpha^*) \left( \varepsilon\varepsilon' \circ (D_x D_x' + D_m D_m' - I_{C^2}) \right) A(\alpha^*). \tag{22}$$

*The variance-covariance matrix of the estimated parameters can again be estimated consistently, by plugging in the estimated residuals of constrained PPML for disturbances $\varepsilon$. However, the rate of convergence is lower and one needs to normalize $\hat{\alpha} - \alpha_0$ by $C^{\frac{1}{2}}$ rather than by $C$ (see Cameron, Gelbach and Miller, 2011, p. 247-248).*

# 4 Counterfactual predictions

The comparative static analysis is based on predicted trade flows obtained under counterfactual changes of the explanatory variables. This section concentrates on conditional general equilibrium effects treating production and expenditure shares as fixed (see Larch and Yotov, 2016).[8] The derivation of the asymptotic distribution of the estimated counterfactual trade flows uses selection matrices that pick out a finite number of countries by skipping the corresponding rows of the non-selected ones from $I_{C^2}$ or generate means for a finite set of trade flows of groups of countries.

**Proposition 3** *(Counterfactual prediction) The set of assumptions for this proposition is given in Appendix A. Let* $V_\alpha = B_0^{-1} A_0 \Omega_\varepsilon A_0' B_0^{-1}$.

*(i) Define the normalized* $(s \times C^2)$ *selection matrix* $S$, $s < K$, *so that* $SM(\alpha_0, Z)$ *possesses typical non-zero elements* $C^2 m_{ij}(\alpha_{0,} z_{ij})$ *and let*

$$\begin{aligned}
\Gamma_0^c &= \lim_{C\to\infty} SM(\alpha_0, Z^c)[I_{C^2} - D\left(D'M(\alpha_0, Z^c)D\right)^{-1} D'M(\alpha_0, Z^c)]Z^c \\
\Gamma_0 &= \lim_{C\to\infty} SM(\alpha_0, Z)[I_{C^2} - D\left(D'M(\alpha_0, Z)D\right)^{-1} D'M(\alpha_0, Z)]Z.
\end{aligned}$$

*Then, it follows that*

$$\begin{aligned}
&CS(m(\widehat{\alpha}, Z^c) - m(\alpha_0, Z^c)) \xrightarrow{d} N(0, \Gamma_0^c V_\alpha \Gamma_0^{c\prime}) \\
&CS(m(\widehat{\alpha}, Z) - m(\alpha_0, Z)) \xrightarrow{d} N(0, \Gamma_0 V_\alpha \Gamma_0') \\
&CS\Delta m(\widehat{\alpha}, Z^c) \xrightarrow{d} N(0, (\Gamma^c - \Gamma)\, V_\alpha\, (\Gamma_0^c - \Gamma_0)')
\end{aligned}$$

*and* $\widehat{\Gamma}^c - \Gamma_0^c = o_p(1)$ *and* $\widehat{\Gamma} - \Gamma_0 = o_p(1)$.

*(ii) Define the* $(r \times C^2)$ *selection matrix* $R$, $r \leq K$, *so that* $RM(\alpha_0, Z)^{-1}$ *has non-zero elements* $m_{ij}(\alpha_{0,} z_{ij})^{-1}$ *and let*

$$\begin{aligned}
\Upsilon_0^c &= \lim_{C\to\infty} RM(\alpha_0, Z)^{-1} M(\alpha_0, Z^c)[I_{C^2} - D\left(D'M(\alpha_0, Z^c)D\right)^{-1} D'M(\alpha_0, Z^c)]Z^c \\
\Upsilon_0 &= \lim_{C\to\infty} R[I_{C^2} - D\left(D'M(\alpha_0, Z)D\right)^{-1} D'M(\alpha_0, Z)]Z.
\end{aligned}$$

*It follows that*

$$CRM(\widehat{\alpha}, Z)^{-1} m(\widehat{\alpha}, Z^c) \xrightarrow{d} N(0, (\Upsilon_0^c - \Upsilon_0)\, V_\alpha\, (\Upsilon_0^c - \Upsilon_0)')$$

*and* $\left(\widehat{\Upsilon}^c - \Upsilon_0^c\right) = o_p(1)$ *and* $\left(\widehat{\Upsilon} - \Upsilon_0\right) = o_p(1)$.

**Proof.** See Appendix F. ■

[8] Counterfactual predictions of full (endowment) general equilibrium effects with endogenously adjusting production and expenditures can be derived in a similar way.

**Remark 6** *It has to be emphasized that this procedure only needs consistent estimates of the structural parameters* $\alpha$*, that may also come from dummy PPML. However,* $\widehat{V}_\alpha$ *has to be calculated according to Proposition 2 to avoid asymptotic bias of the standard errors of the counterfactual predictions.*

# 5 Monte Carlo simulations

The Monte Carlo simulation experiments assesse the quality of the asymptotic results of Propositions 2 and 3 as an approximation in medium sized cross-sections. The simulations are based on a simplified structural gravity model that is specified as

$$
\begin{aligned}
s_{ij,C} &= e^{-0.4z_{ij,1}-0.75z_{ij,2}+\beta_{i,C}+\gamma_{j,C}}\eta_{ij} \\
\kappa_{i,C} &= \sum_{j=1}^{C} e^{-0.4z_{ij,1}-0.75z_{ij,2}+\beta_{i,C}+\gamma_{j,C}} \\
\theta_{j,C} &= \sum_{i=1}^{C} e^{-0.4z_{ij,1}-0.75z_{ij,2}+\beta_{i,C}+\gamma_{j,C}},
\end{aligned}
$$

where $C^2\eta_{ij}$ is iid $N(1,\sigma^2)$ and $\eta$ enters in multiplicative form. [9] The explanatory variables are taken from CEPII's database. $z_{ij,1}$ is a dummy for a common official language. The second explanatory variable, $z_{ij,2}$, refers to log weighted distance. $z_{ij,1}$ is zero for intra-country trade flows ($i = j$) and both explanatory variables remain fixed in repeated samples. Production and expenditure shares come from GTAP.[10] Data are sorted by country size so that the sample always includes the $C$ largest ones. The multilateral resistance terms, $\beta_{i,C}$ and $\gamma_{j,C}$, are derived as solutions to the system of multilateral resistances at true parameter values $\alpha_0 = (-0.4, -0.75)$. Thus the system of multilateral resistances holds in expectation, but not for the generated trade flows which are subject to measurement error. This design guarantees that the $E[s_{ij,C}]$ is uniformly bounded. The normality assumption on $\varepsilon_{ij}$ violates the uniform boundedness assumption and may produce negative realizations of $s_{ij,C}$. In all the Monte Carlo runs this did not occur, however.

The Monte Carlo experiments consider $C \in \{40, 60\}$ and set $\sigma = \frac{20}{C^2}$ to account for the assumption that $\sigma^2$ decreases with sample size (Assumption Part I.3). Un-

[9] Assuming non-normal disturbances, where $C^2\eta_{ij} = \sigma\left(1 + \frac{\zeta_{ij}-10}{20}\right)$, $\zeta_{ij}$ being iid $\chi^2(10)$, yields similar results, which are available upon request.

[10] For simplicity GDP shares are used for the simulations. For the Monte Carlo experiments this choice is irrelevant and does distort the results.

der these assumptions the standard errors of the estimated parameters are similar to those found in the literature. Furthermore, in a set of experiments 50 percent of the trade flows remain unobserved, while all observations on the explanatory variables and the production and expenditure shares are available. The third set of experiments considers the case, where only domestic trade flows are missing.

Table 1: Monte Carlo simulation results I: Simulated standard errors, estimated standard errors and 95% coverage rates of structural parameters under constrained and dummy PPML

| Countries | Missings | Sim. std. | | Est. std. | | 95% Coverage rate | |
|---|---|---|---|---|---|---|---|
| | | $\alpha_1$ | $\alpha_2$ | $\alpha_1$ | $\alpha_2$ | $\alpha_1$ | $\alpha_2$ |
| Constrained PPML | | | | | | | |
| 40 | 0 | 0.0117 | 0.0030 | 0.0110 | 0.0028 | 0.938 | 0.939 |
| 40 | 50% | 0.0147 | 0.0043 | 0.0136 | 0.0040 | 0.924 | 0.919 |
| 40 | domestic | 0.0089 | 0.0019 | 0.0087 | 0.0018 | 0.939 | 0.944 |
| | | | | | | | |
| 60 | 0 | 0.0048 | 0.0013 | 0.0046 | 0.0012 | 0.944 | 0.940 |
| 60 | 50% | 0.0078 | 0.0020 | 0.0070 | 0.0018 | 0.924 | 0.929 |
| 60 | domestic | 0.0037 | 0.0008 | 0.0036 | 0.0008 | 0.941 | 0.945 |
| | | | | | | | |
| Dummy PPML | | | | | | | |
| 40 | 0 | 0.0117 | 0.0030 | 0.0072 | 0.0014 | 0.778 | 0.632 |
| 40 | 50% | 0.0137 | 0.0039 | 0.0088 | 0.0017 | 0.788 | 0.619 |
| 40 | domestic | 0.0112 | 0.0035 | 0.0087 | 0.0023 | 0.872 | 0.801 |
| | | | | | | | |
| 60 | 0 | 0.0048 | 0.0013 | 0.0030 | 0.0006 | 0.780 | 0.646 |
| 60 | 50% | 0.0069 | 0.0018 | 0.0035 | 0.0008 | 0.670 | 0.578 |
| 60 | domestic | 0.0046 | 0.0014 | 0.0035 | 0.0009 | 0.865 | 0.796 |

*Notes*: 5000 Monte Carlo runs. Coverage rate refers to a nominal 95 percent confidence interval using the normal distribution.

Both the dummy and the constrained PPML estimators of the structural parameters are virtually unbiased in all considered experiments with negligible deviations from their true values and the results are thus not reported. Table 1 exhibits the simulated standard errors of the parameter estimates of $\alpha_1$ and $\alpha_2$ and their estimated counterparts for the constrained and the dummy PPML. Thereby, the simulated standard errors are calculated as the standard deviation of the corresponding point estimates in 5000 Monte Carlo runs. The last two columns of Table

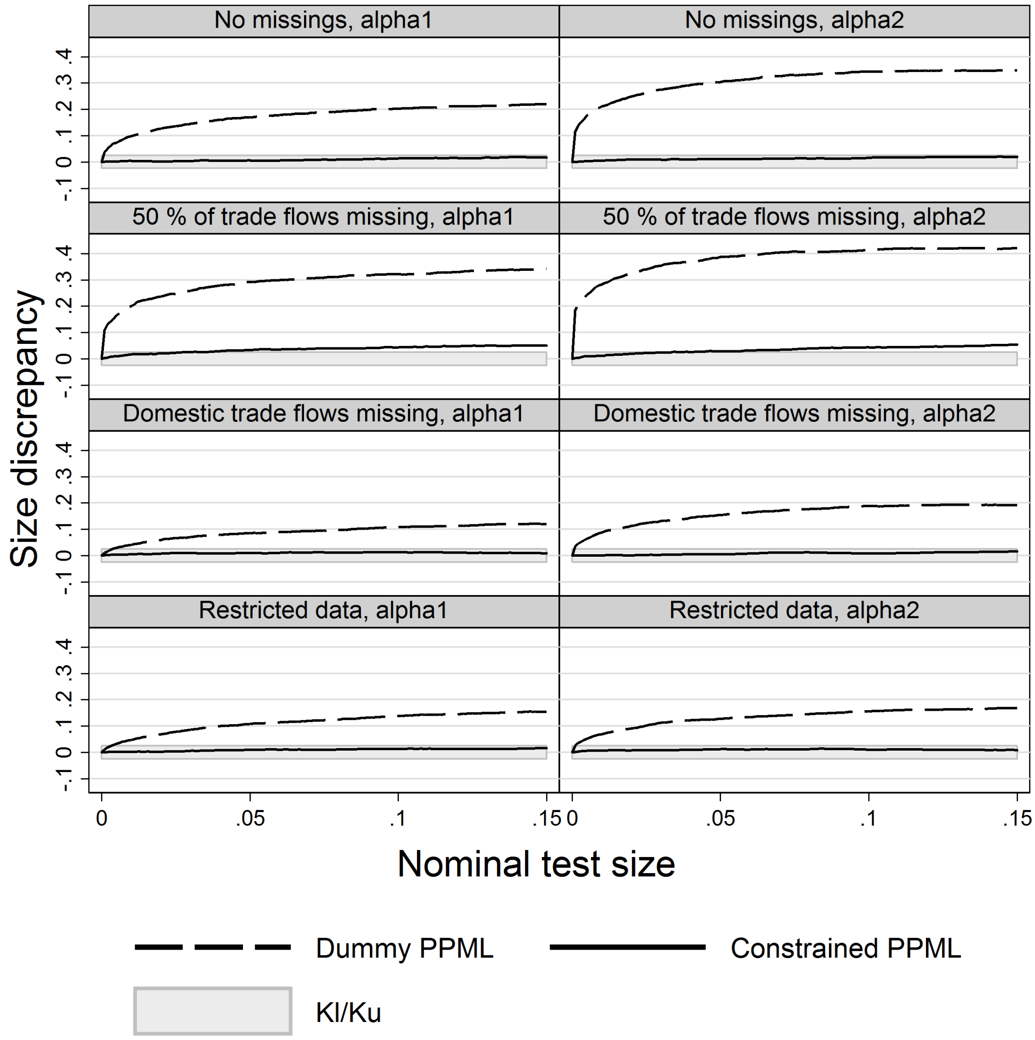
1 display the coverage rates of 95%-confidence intervals to check the validity of the asymptotic distribution of $\widehat{\alpha}$ as derived in Proposition 2 in finite samples. For constrained PPML the estimated standard errors are quite close to their simulated counterparts in almost all experiments and in most of cases the simulated coverage rates come close to their nominal values. Considering the uncertainty induced by the Monte Carlo simulation, a 99% confidence interval of the simulated coverage rates is $[0.942, 0.958]$. The simulated coverage rates are below the lower bound of this interval by a small margin, especially at $C = 40$ and in case of 50% missing values.

In contrast to constrained PPML but in line with the findings in Remark 3, the standard errors of the structural parameters estimated by dummy PPML are severely downward biased. The bias calculations given in Appendix G show that under homoskedastic disturbances $\eta$, as a assumed in the Monte Carlo design, the proportionate bias is driven by the leverage only. Assuming that the true parameters are known, the lower and the upper bound of the proportionate bias are calculated as $[-6.13, 0.00]$ percentage points for the standard error of $\widehat{\alpha}_1$ estimated by constrained PPML. Under dummy PPML this interval widens to $[-74.30, -0.77]$ percentage points. The Monte Carlo simulations confirm this finding. The corresponding simulated coverage ratios reported in Table 1 are substantially and significantly below their the nominal value of 95%. For example, for $C = 40$ and fully observed trade flows these coverage rates amount to 0.778 and 0.632 for $\alpha_1$ and $\alpha_2$, respectively. With the data at hand the leverage and, therefore, the lower bound of the bias of dummy PPML increases only marginally at $C = 60$.

The size discrepancy plot in Figure 1 for $C = 60$ is based on Davidson and MacKinnon (1998) and displays the difference between the actual and nominal size of a t-test for $H_0 : \alpha_1 = -0.4$ and $H_0 : \alpha_2 = -0.75$ over a range of significance levels. Formally, the size discrepancy curve is derived from the empirical cumulative distribution function of the p-values $p_r$ defined as $F(q) = \frac{1}{R}\sum_{r=1}^{R} I(p_r \leq q)$, where $R$ is the number of Monte Carlo replications. Figure 1 exhibits the plots of $F(q) - q$ against $q$ under the assumption that $H_0$ actually holds. In addition, the Kolmogorov and Smirnov test shows whether $F(q) - q$ differs significantly from 0 (Davidson and MacKinnon 1998, p. 11) and the size of the t-tests is distorted (see the Kl/Ku-band in Figure 1). The results indicate that in case of fully observed trade flows or only domestic trade flows missing the t-tests based on the constrained PPML are correctly sized and located within the 1% Kolmogorov and Smirnov band. If 50% of trade flows are missing the t-tests are marginally oversized at nominal test size above 0.075% indicating that the sample should include 40 countries or more to obtain properly sized t-tests.

In contrast, the t-tests based on the dummy PPML are considerable oversized in all experiments. As shown above the reason is that the estimated variances are

Figure 1: Size discrepancy plot,
dummy PPML vs. constrained PPML




Notes: C=60, sigma/C^2=0.022

prone to asymptotic bias induced by the incidental parameter problem. For example, for $C = 60$ the actual size of a t-test for $\alpha_1$ based on the dummy PPML amounts to 0.22 under fully observed trade flows and to 0.34 in case of 50 percent missing trade flows, if a nominal significance of 0.05 assumed. The size distortion of t-tests based on the dummy PPML is more pronounced if 50% of trade flows are missing, but turn out smaller if only domestic trade flows are missing.

Table 2: Monte Carlo simulation results II: 95% coverage rates counterfactual changes under constrained PPML

| Countries | Missings | Country pairs | | | |
|---|---|---|---|---|---|
| | | 1,1 | 1,2 | 2,1 | 2,2 |
| Absolute Change | | | | | |
| 40 | 0 | 0.940 | 0.939 | 0.937 | 0.926 |
| 40 | 50% | 0.923 | 0.925 | 0.924 | 0.910 |
| 40 | domestic | 0.945 | 0.940 | 0.941 | 0.938 |
| | | | | | |
| 60 | 0 | 0.931 | 0.937 | 0.944 | 0.932 |
| 60 | 50% | 0.907 | 0.911 | 0.919 | 0.906 |
| 60 | domestic | 0.927 | 0.930 | 0.936 | 0.924 |
| | | | | | |
| Relative Change | | | | | |
| 40 | 0 | 0.950 | 0.937 | 0.948 | 0.951 |
| 40 | 50% | 0.939 | 0.922 | 0.936 | 0.939 |
| 40 | domestic | 0.952 | 0.938 | 0.948 | 0.953 |
| | | | | | |
| 60 | 0 | 0.950 | 0.954 | 0.951 | 0.957 |
| 60 | 50% | 0.933 | 0.937 | 0.933 | 0.941 |
| 60 | domestic | 0.943 | 0.951 | 0.948 | 0.949 |

*Notes*: 5000 Monte Carlo runs. The coverage rate refers to a nominal 95 percent confidence interval using the normal distribution.

Table 2 displays the results of the estimated impact when the dummy for common official language is counterfactually set to 0 and constrained PPML is used.[11] The reported figures in the Table 2 refer to the change in trade flows within and between the smallest two countries.[12] Across the board, the estimated

[11] Detailed simulation results for dummy PPML are available upon request.

[12] These country pairs exhibit different languages. So the common language dummy has been set to 1 in the data base to allow for both direct and indirect effects of the counterfactual change.

standard errors are well in line with the simulated ones. For absolute changes the simulated coverage rates are slightly below their nominal values, while for relative changes and $C = 60$ the simulated coverage rates are located always in the 99% confidence interval under fully observed trade flows or in case of missing domestic trade flows. However, the deviations of the coverage rates from their nominal values are somewhat larger with 50% missing values.

Table 3: Monte Carlo simulation results III: 95% coverage rates of counterfactual changes with restricted observed trade flows

| Countries | Country pairs | | | |
|---|---|---|---|---|
| | 1,1 | 1,2 | 2,1 | 2,2 |
| Absolute Change | | | | |
| 40 | 0.954 | 0.945 | 0.941 | 0.936 |
| 60 | 0.922 | 0.940 | 0.941 | 0.919 |
| | | | | |
| Relative Change | | | | |
| 40 | 0.950 | 0.946 | 0.947 | 0.952 |
| 60 | 0.943 | 0.942 | 0.943 | 0.943 |

*Notes*: 5000 Monte Carlo runs. The coverage rate refers to a nominal 95 percent confidence interval using the normal distribution.

Table 3 reports simulated coverage ratios of the 95%-confidence intervals for the case when the observed trade flows add-up to given exporters' production and importers' expenditures and the variance-covariance matrix of the disturbances is singular. Trade flows and the corresponding disturbances are generated according to the specification in (21). To avoid negative trade flows, especially domestic ones in small countries, disturbances are specified as distributed independently truncated normal so that the bounded support assumption of the disturbances is fulfilled. This reduces the variance of disturbances and the standard errors of estimated parameters. In order to obtain realistic t-values the log distance has been scaled by a factor 100. Results show that coverage rates are slightly below their nominal values, especially at $C = 60$, in case of absolute changes. For relative changes the simulated coverage rates always lie in the 99%-confidence interval $[0.942, 0.958]$. The last row of graphs in Figure 1 indicates that t-test based on constrained PPML are correctly sized in this setting, while dummy PPML leads to oversized tests.

Overall, the simulation results show that the limit distribution of the constrained PPML comes up with approximately correct standard errors and coverage rates of the 95% confidence intervals as well as correctly sized tests if the system of multilateral resistances holds in expectation. Also the estimated standard errors as derived by the delta method in Proposition 3 allow proper inference and provide approximately correct confidence intervals for comparative static experiments in reasonably large cross-sections.

# 6 The impact of common language, contiguity and country borders on bilateral trade flows

As an illustration of the constrained PPML estimation procedure this section considers the impact of common language, contiguity and national borders on international trade and welfare following the approach of Costinot and Rodríguez-Clare (2014). The sample comprises 59 countries and refers to the year 2006. Data on goods trade and compatible figures on total manufacturing production, aggregate exports and imports are taken from OECD's Stan database. The latter data are augmented by information from CEPII's Trade, Production and Bilateral Protection (TradeProd) database as well as aggregated Comtrade data to impute a few missing data points on total manufacturing production, aggregate exports or imports.[13] Then all production and expenditure data for the 59 countries are available. Details on the calculation of domestic trade as well as on the corrections for trade with the rest of the world and trade imbalances are given in Appendix H.

For the analysis of the impact of common language and contiguity the missing trade flows as well as domestic trade are treated as true missings, but are implicitly predicted by constrained PPML. The estimation of border effects includes data on domestic trade flows that have been calculated as residuals from aggregate production exports and imports setting 48 missing bilateral trade flows to zero so that the data fulfill the above mentioned aggregation restrictions of the disturbances.[14] In both cases trade flows, production and expenditures are normalized by world output and sum to 1.

Lastly, the measures of population weighted distance and the dummies for contiguity, official common language, colony and common colonizer are from Mayer and Zignago (2011). All explanatory variables are defined as trade barriers so that all their parameters exhibit negative signs. For example, the variable border takes the value 1 if exporter and importer countries are different, while it is 0 for within country flows.

[13] Details on the interpolation procedure are available upon request.

[14] Setting these values to missings rather than to 0 yields nearly identical estimation results.

Table 4: Parameter estimates, dummy and constrained PPML

| | Dummy PPML | | Constrained PPML | |
|---|---|---|---|---|
| | $\alpha$ | t-value | $\alpha$ | t-value |
| (A) Without domestic trade flows | | | | |
| Country border | - | - | - | - |
| No common official language | $-0.25$ | $-3.34^{***}$ | $-0.40$ | $-4.99^{***}$ |
| No contiguity | $-0.27$ | $-4.09^{***}$ | $-0.33$ | $-3.59^{***}$ |
| No colony | 0.14 | 1.34 | 0.05 | 0.50 |
| No common Colonizer | $-0.37$ | $-1.37$ | $-1.30$ | $-12.10^{***}$ |
| Log distance | $-0.89$ | $-27.64^{***}$ | $-0.91$ | $-22.21^{***}$ |
| Pseudo-$R^2$ | 0.968 | | 0.997 | |
| (B) With domestic trade flows | | | | |
| Country border | $-1.42$ | $-4.97^{***}$ | $-1.42$ | $-4.60^{***}$ |
| No common official language | $-0.33$ | $-3.46^{***}$ | $-0.33$ | $-2.58^{***}$ |
| No contiguity | $-0.43$ | $-4.28^{***}$ | $-0.43$ | $-2.76^{***}$ |
| No colony | 0.11 | 1.11 | 0.11 | 0.58 |
| No common Colonizer | $-0.01$ | $-0.02$ | $-0.01$ | $-0.03$ |
| Log distance | $-0.91$ | $-20.25^{***}$ | $-0.91$ | $-17.77^{***}$ |
| Pseudo-$R^2$ | 0.997 | | 0.997 | |

*Notes*: The estimates in Panel A are based on 3374 observations and those in Panel B on 3481. Pseudo-$R^2$ is defined as the correlation of observed an predicted values. *** significant at 1 %.

The parameter estimates reported in panel (A) of Table 4 exclude domestic trade flows and 49 missing trade flows. The parameter estimates differ substantially between dummy and constrained PPML. With exception of the dummy for colony, which is insignificant, all parameter estimates turn out higher in absolute value under constrained PPML. Moreover, dummy PPML does a bad job in predicting out of sample, since it substantially underestimates domestic trade of the large countries. Aggregating the normalized trade flows predicted by dummy PPML yields a total of 0.56, casting some doubt on the consistency of dummy PPML estimates when domestic trade flows are missing. In contrast, per construction under constrained PPML the predictions sum up to 1 coming up with more realistic predictions of domestic trade flows.

Panel B of Table 4 refers to the case of fully observed trade flows. In this setting dummy PPML and constrained PPML yield identical point estimates, while their estimated standard errors differ. In line with the theoretical findings and those of the Monte Carlo simulations the standard errors tend to be lower under dummy PPML leading to higher t-values. Since, with exception of the dummies for colony and common colonizer all parameter are estimated very precisely, the conclusions from the t-tests do not change, however. Under fully observed trade flows, it is possible to estimate border effects which turn out to be substantial and similar in size to those found in the literature. Overall, the results support the findings of Yotov (2012), who argues that the impact of international economic integration should be measured against that of internal markets by including domestic trade flows.

The results for the counterfactual changes of the common language, contiguity and common borders dummies are reported in Table 5 and refer to a conditional general equilibrium setting (see Larch and Yotov, 2016). This exercise allows for third country effects via changes in multilateral resistance terms, while output and expenditures remain unchanged. To provide a summary of the estimated effects, the sample is split in large and small countries (below and above the median of the value of gross production) and average percentage changes of trade flows within and between these groups are reported. The calculation of the welfare effects in Table 6 follows Arkolakis, Costinot and Rodríguez-Clare (2012) and is based on $(\widehat{s}^{C}_{ii,C}/\widehat{s}_{ii,C})^{(1/1-\sigma)}$, where the elasticity of substitution $\sigma$ is assumed to be 6.982, the preferred estimate in Bergstrand, Egger and Larch (2013, Table 1). The corresponding standard errors are derived using the delta method.

Overall, the estimated direct effects considerably overestimate the impact of the reduction in trade barriers. In the counterfactual scenario with all trade barriers removed that arise from a lack of a common language we see a substantial increase in trade flows within the group of large countries (19.84%). However, despite the precise estimation with a t-value of 4.01, the 95%−confidence interval turns out quite large amounting to $[10.14, 29.53]$ percentage points. The increase is smaller in case of trade flows that involve small countries. However, with a reduction of 22.9%, $[-37.77, -13.02]$, the impact on domestic trade is more pronounced in the group of small countries, which translates into larger welfare gains of small coun-

Table 5: Comparative statics

| | Direct change in % | | | | Total change in % | | | |
|---|---|---|---|---|---|---|---|---|
| | Impact | t-value | [95% Conf. | Interval] | Impact | t-value | [95% Conf. | interval] |
| (A) Common language | | | | | | | | |
| Dom-small | 0.00 | - | - | - | −22.90 | −4.55*** | −32.77 | −13.02 |
| Dom-large | 0.00 | - | - | - | −8.72 | −2.94*** | −14.53 | −2.90 |
| Small-small | 31.32 | 6.14*** | 21.32 | 41.32 | 2.60 | 3.48*** | 1.14 | 4.07 |
| Large-large | 30.08 | 6.14*** | 20.48 | 39.69 | 19.84 | 4.01*** | 10.14 | 29.53 |
| Small-large | 31.01 | 6.14*** | 21.11 | 40.91 | 14.01 | 4.71*** | 8.18 | 19.83 |
| Large-small | 30.65 | 6.14*** | 20.87 | 40.44 | 11.19 | 5.54*** | 7.23 | 15.14 |
| (B) Contiguity | | | | | | | | |
| Dom-small | 0.00 | - | - | - | −19.63 | −2.61*** | −34.38 | −4.89 |
| Dom-large | 0.00 | - | - | - | −6.28 | −1.60 | −13.99 | 1.44 |
| Small-small | 26.72 | 3.62*** | 12.24 | 41.20 | −2.82 | −3.80*** | −4.28 | −1.37 |
| Large-large | 25.97 | 3.63*** | 11.94 | 40.00 | 14.13 | 2.12** | 1.09 | 27.17 |
| Small-large | 27.01 | 3.65*** | 12.50 | 41.51 | 11.97 | 3.15*** | 4.53 | 19.42 |
| Large-small | 27.08 | 3.67*** | 12.63 | 41.53 | 10.22 | 3.52*** | 4.53 | 15.92 |
| (C) Border | | | | | | | | |
| Dom-small | 0.00 | - | - | - | −69.82 | −8.79*** | −85.39 | −54.26 |
| Dom-large | 0.00 | - | - | - | −36.84 | −3.74*** | −56.16 | −17.52 |
| Small-small | 75.74 | 10.15*** | 61.11 | 90.37 | 3.84 | 0.70 | −6.92 | 14.60 |
| Large-large | 75.74 | 10.15*** | 61.11 | 90.37 | 91.77 | 7.17*** | 66.69 | 116.86 |
| Small-large | 75.74 | 10.15*** | 61.11 | 90.37 | 44.45 | 7.58*** | 32.95 | 55.95 |
| Large-small | 75.74 | 10.15*** | 61.11 | 90.37 | 38.73 | 10.76*** | 31.68 | 45.79 |

*Notes*: ** significant at 5 %, *** significant at 1 %.

tries amounting to 4.44%, [2.21, 6.68]. In comparison the group of large countries would achieve a welfare improvement of 1.54%, [0.46, 2.82] on average (see Table 6, Panel A). Reducing the trade barriers to those between neighboring countries lead to a somewhat smaller direct impact (see Tables 5 and 6, Panel B). The main difference is now that trade flows among the group of small countries decrease by 2.82%, [1.37, 4.28] on average. Again the average welfare effects exhibited in Table 6 are substantial for the group of small countries amounting to 3.72%, [0.54, 6.90], while they are insignificant for the group of large countries.

Table 6: Estimated welfare effects of counterfactual changes in common language, contiguity and country border

| | Change in % | t-value | [95% Conf. | interval] |
|---|---|---|---|---|
| (A) Common language | | | | |
| Small | 4.44 | 3.89*** | 2.21 | 6.68 |
| Large | 1.54 | 2.79*** | 0.46 | 2.62 |
| (B) Contiguity | | | | |
| Small | 3.72 | 2.29*** | 0.54 | 6.90 |
| Large | 1.09 | 1.54 | −0.30 | 2.48 |
| (C) Border | | | | |
| Small | 22.18 | 4.13*** | 11.64 | 32.71 |
| Large | 7.99 | 2.83*** | 2.46 | 13.51 |

*Notes*: *** significant at 1%.

Lastly, Panel C of Table 5 reports the impact of eliminating country borders, while preserving the effects of geography and the other trade barriers. This exercise complements the standard counterfactual experiments of reverting to autarky reported, e.g., in Costinot and Rodríguez-Clare (2014). Abandoning counterfactually country borders yields a direct trade increase of 75.74%, [61.11, 90.37]. Accounting for changes in multilateral resistances reveals considerable large effects on trade flows among the group of large countries amounting to 91.77%, [66.69, 116.86]) on average. For trade flows from small to large and large to small countries the corresponding estimates are 44.45%, [32.95, 55.95] and 38.73%, [31.68, 45.79], respectively. Again the increase is smallest in case of trade flows among the group of small countries which turns out insignificant. As shown in Table 6, removing borders would lead to pronounced welfare effects for the small countries of

$22.18\%, [11.64, 32.71]$, while large countries would experience a much smaller increase of $7.99\%, [2.46, 13.51]$.

# 7 Conclusions

PPML is now widely used for estimation of structural gravity models that impose trade balance restrictions, whose solutions define the exporter- and importer-specific multilateral resistance terms. Introducing exporter and importer dummies in cross-sectional models proved especially useful for estimation to account for these multilateral resistances.

However, for calculating standard errors and confidence intervals of predicted counterfactuals the dummy PPML approach is of limited use. Multilateral resistance terms are not parameters to be estimated, rather they are solutions to the system of multilateral multilateral resistances that depend on the structural parameters of the gravity model. For comparative static analysis exactly this assumption is commonly maintained, namely that the expected values of trade flows adhere to the restrictions imposed by the system of multilateral resistances. This assumption turns out very useful for both estimation of structural parameters and the prediction of counterfactuals. Based on this assumption, constrained PPML leads to a projection based estimator as proposed by Heyde and Morton (1993) in a setting without dummies.

The present contribution establishes the asymptotic distribution of the constrained PPML estimator and, using the delta method, derives asymptotic results for comparative static predictions and the implied percentage changes. Hence, it is possible to test hypotheses on counterfactual changes and to provide confidence intervals, both based on a limit distribution that is unaffected by incidental parameters. Monte Carlo simulations provide encouraging results for medium sized cross-sections.

Lastly, the usefulness of the proposed estimation procedure is illustrated by estimating a structural gravity model for 59 countries, analyzing the impact of common language, contiguity and country borders on bilateral trade and welfare. Results show that the impact is largest for trade flows involving a large trading partner, while small countries gain most in terms of welfare from eliminating these trade barriers. The quantitative estimates of these effects turn out significant in almost all cases, while 95% confidence intervals of the counterfactual predictions are relatively large, despite the precise estimation of the involved parameters.

# Appendix

## A The set of assumptions

Since $\beta_{i,C}(\alpha)$ and $\gamma_{j,C}(\alpha)$ change with sample size, the DGP of the structural gravity model forms a triangular array with index $C$.

**Part I: Consistency**

1) Normalization: $\beta_{iC} = 0$.

2) The parameter space of $\alpha$, $\Theta \subset \mathbb{R}^K$, is compact. $\alpha_0$ is an interior point of $\Theta$.

3) Disturbances: $C^2\varepsilon_{ij}$, $ij = 1, ..., C$ is independently distributed as $(0, \sigma_{ij}^2)$ with $\sigma_{ij}^2 < \overline{\sigma} < \infty$ and bounded support so that $m_{ij}(\alpha_0, \phi_C(\alpha_0)) + \varepsilon_{ij} > 0$. One may also write, $\varepsilon_{ij} = m_{ij}(\alpha_0, \phi_C(\alpha_0))(\eta_{ij} - 1)$ with $\eta_{ij}$ independently distributed as $(1, \sigma_{\eta,ij}^2)$ and $\underline{\sigma}_\eta^2 \leq \sigma_{\eta,ij}^2 \leq \overline{\sigma}_\eta^2$.

4) (i) $c_a/C^2 < m_{ij}(\alpha, \phi_C(\alpha)) < (1 - c_a)/C^2$ for some positive constant $c_a < 0.5$.

   (ii) $Y_W = c_W C^2$ for some positive constant $c_W$.

   (iii) The system of multilateral resistances holds under the true model: $r_C(\alpha_0, \phi_C(\alpha_0)) := D'm(\alpha_0, \phi_C(\alpha_0)) - \theta_C = 0$, where $\theta_C$ is given, non-stochastic and of order $O\left(C^{-1}\right)$.

5) Missings: $\frac{K+2C-1}{C^2} < \frac{\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}}{C^2} \leq 1$.

6) $Z$ possesses full column rank $K$, its elements are uniformly bounded by some constant $c_z$, i.e., $|z_{ij,k}| \leq c_z$ and all elements of $Z$ vary at the bilateral level.

**Part II: Limit distribution of $\widehat{\alpha}$**

1. Let $q_{ij,C}(\alpha) = Y_W\left(m_{ij}(\alpha_0, \phi_C(\alpha_0)) + \varepsilon_{ij}\right) \ln\left(m_{ij}(\alpha, \phi_C(\alpha))Y_W\right) - m_{ij}(\alpha, \phi_C(\alpha))$. $q_{ij,C}(\alpha)$ is twice continuously differentiable at every interior point $\alpha \in \Theta$ for each $\varepsilon_{ij}$ and $z_{ij}$.

2. Let $M(\alpha) = diag(m_{ij}(\alpha, \phi_C(\alpha)) :$

   (i) $G(\alpha) = W'VM(\alpha)VW$ possesses uniformly bounded elements and is invertible for $\alpha \in \Theta'$, where $\Theta'$ is a closed ball around $\alpha_0$ in the interior of $\Theta$. $\|G(\alpha)\| \leq (K + 2C - 1)c_{g,1} < \infty$ and $\|G(\alpha)^{-1}\| \leq (K + 2C - 1)c_{g,2} < \infty$.

(ii) The elements of $\left|D\frac{\partial\phi(\alpha)}{\partial\alpha'}\right| = \left|D\left(D'M(\alpha)D\right)^{-1}D'M(\alpha)Z\right|$ are uniformly bounded by some constant $c_\phi$.

3. Assumption on moments $s_\alpha(\alpha)$ (Billingsley, 1995, Theorem 27.3). Let

$$\begin{aligned} s_\alpha(\alpha)_{(K\times 1)} &= C^2\sum_{i=1}^{C}\sum_{j=1}^{C}\left[I_K, 0_{K\times 2C-1}\right]G^{1/2}(\alpha)Q_{G^{-1/2}F(\alpha)'}G^{-1/2}(\alpha) \\ &\quad * w_{ij(K+2C-1\times 1)}v_{ij}\varepsilon_{ij} = A(\alpha)\varepsilon. \end{aligned}$$

$E\left[\|s_\alpha(\alpha)\|^{2+\delta}|Z,V\right] = o(1)$.

4. Let $Q_{M^{1/2}D}(\alpha) = I_{C^2} - M(\alpha,Z)^{\frac{1}{2}}D\left(D'M(\alpha,Z)D\right)^{-1}D'M(\alpha,Z)^{\frac{1}{2}}$. The following limits exist and are finite:

(i) $B(\alpha) = Z'VM(\alpha)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha)M(\alpha)^{\frac{1}{2}}Z$, $B_0 = \lim_{C\to\infty}B(\alpha_0)$, $B_0$ is non-singular.

(ii) The limits $\Gamma_0^c = \lim_{C\to\infty}SM(\alpha,Z^c)Q_{M^{1/2}D}(\alpha)M(\alpha,Z^c)^{\frac{1}{2}}Z^c$ and $\Gamma_0 = \lim_{C\to\infty}SM(\alpha,Z)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha)M(\alpha,Z)^{\frac{1}{2}}Z$ exist, are non-zero and have rank $s$, where $s$ is the rank of the $s\times C^2$, $s\le K$, selection matrix that is scaled by $C^2$.

(iii) The limits $\Upsilon_0^c = \lim_{C\to\infty}RM(\alpha_0,Z)^{-1}M(\alpha_0,Z^c)[I_{C^2} - D\left(D'M(\alpha_0,Z^c)D\right)^{-1} * D'M(\alpha_0,Z^c)]Z^c$ and $\Upsilon_0 = \lim_{C\to\infty}R[I_{C^2} - D\left(D'M(\alpha_0,Z)D\right)^{-1}D'M(\alpha_0,Z)]$ $Z$ exist, are non-zero and have rank $r$, where $r$ is the rank of the $r\times C^2$, $r\le K$, selection matrix.

# B The implicit solution to the system of multilateral resistances

The parameters $\beta_{i,C}(\alpha)$ and $\gamma_{j,C}(\alpha)$ are derived by solving the non-stochastic system of the multilateral resistance equations with solutions $\phi_C(\alpha) = [\beta_C(\alpha),'\gamma_C(\alpha)']'$. The system can be written as

$$r_C(\alpha,\phi_C(\alpha)) := D'm(\alpha,\phi_C(\alpha)) - \theta_C = 0,$$

where $r_C(\alpha,\phi_C(\alpha))$ is continuously differentiable. The derivative is

$$\frac{\partial r_C(\alpha,\beta_C,\gamma_C)}{\partial(\beta_C',\gamma_C')} = \begin{bmatrix} \chi_C & T_C(\alpha) \\ T_C'(\alpha) & \Theta_C \end{bmatrix},$$

where $\chi_C = diag(\kappa_{1,C}, ..., \kappa_{C-1,C})$, $\Theta_C = diag(\theta_{1,C}, ..., \theta_{C,C})$ and $T_C(\alpha)$ is a $(C-1 \times C)$ matrix with typical element $m_{ij}(\alpha)$, $i = 1, ..., C-1$ and $j = 1, ..., C$. To guarantee the existence of a unique solution (Sydsaeter et al., 2005, p. 102) it has to hold that

$$\left|\det\left(\frac{\partial r_C(\beta_C,\gamma_C)}{\partial(\beta_C',\gamma_C')'}\right)\right| \geq c_h > 0 \text{ and } \sup_{ij}\left\{\left|\frac{\partial r_{C,ij}(\beta_C,\gamma_C)}{\partial\beta_{i,C})}\right|, \left|\frac{\partial r_{C,ij}(\beta_C,\gamma_C)}{\partial\gamma_{j,C}}\right|\right\} \leq c_k$$

for some positive constants $c_h$ and $c_k$. With respect to latter observe that $0 < \frac{c_a}{C^2} \leq e^{z^{ij\prime}\alpha+\beta_{i,C}(\alpha)+\gamma_{j,C}(\alpha)} \leq \frac{1-c_a}{C^2} < 1$. The former holds as $\frac{\partial r_C(\beta_C,\gamma_C)}{\partial(\beta_C',\gamma_C')'}$ is a strictly diagonally dominant matrix with real positive diagonal entries and it is thus positive definite. Hence, one can conclude that for finite $C$ in its normalized form the system of multilateral resistances possesses a unique solution $\phi_C(\alpha) = [\beta_C(\alpha)', \gamma_C(\alpha)']'$, which is continuously differentiable in $\alpha$ at every interior point in $\Theta$.

# C Proof of Proposition 1 (Constrained PPML)

Following Falocci, Paniccià and Stanghellini (2009) assume that at iteration $r$ the estimate $\hat{\vartheta}_{C,r}$ is given and use a linearization of the score around $\hat{\vartheta}_{C,r}$. The remainders are denoted by $b_r$ and $c_r$, respectively. To simplify notation, the arguments of $\widehat{G}_r = W'VM(\widehat{\vartheta}_{C,r})VW$ and $\widehat{F}_r = D'M(\widehat{\vartheta}_{C,r})W$ are skipped.

$$\underbrace{W'V(s_C - m(\widehat{\vartheta}_{C,r+1}))}_{0} = W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) - \widehat{G}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) + \widehat{F}_r'\lambda_r + b_r$$

$$\begin{aligned}\underbrace{D'm(\widehat{\vartheta}_{C,r+1}) - \theta_C}_{0} &= D'm(\widehat{\vartheta}_{C,r}) - \theta_C + \underbrace{D'M(\widehat{\vartheta}_{C,r})W}_{\widehat{F}_r}\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) + c_r \\ \widehat{F}_r\widehat{\vartheta}_{C,r+1} &= \underbrace{\theta_C - D'm(\widehat{\vartheta}_{C,r}) + \widehat{F}_r\widehat{\vartheta}_{C,r} - c_r}_{\widehat{h}_r}\end{aligned}$$

Under the assumption that the inverses of $\widehat{G}_r$ and $\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'$ exist for finite $C$, one obtains

$$\begin{aligned}\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r} &= \widehat{G}_r^{-1}\left(W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) + \widehat{F}_r'\lambda_r + b_r\right) \\ \widehat{F}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) &= \widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\widehat{\lambda}_r + \widehat{F}_r\widehat{G}_r^{-1}\left(W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) + b_r\right)\end{aligned}$$

$$\widehat{\lambda}_r = \left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\widehat{F}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) - \widehat{F}_r\widehat{G}_r^{-1}W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) + b_r\right).$$

Inserting $\widehat{\lambda}_r$ into the score equation yields

$$\begin{aligned}
0 &= W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) - \widehat{G}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) \\
&\quad + \widehat{F}_r'\left[\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\underbrace{\widehat{F}_r\widehat{\vartheta}_{C,r+1}}_{\widehat{h}_r} - \widehat{F}_r\widehat{\vartheta}_{C,r} - \widehat{F}_r\widehat{G}_r^{-1}\left(W'V\left(s - m(\widehat{\vartheta}_{C,r})\right) + b_r\right)\right)\right] + b_r \\
&= -\widehat{G}_r\left(\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r}\right) \\
&\quad + \left(I - \widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)\left(W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) + b_r\right) \\
&\quad + \widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\widehat{h}_r - \widehat{F}_r\widehat{\vartheta}_{C,r}\right).
\end{aligned}$$

Given $\widehat{\vartheta}_{C,r}$, one can calculate $\widehat{h}_r$, $\widehat{F}_r$ and $\widehat{G}_r$ to obtain

$$\begin{aligned}
\widehat{\vartheta}_{C,r+1} - \widehat{\vartheta}_{C,r} &= \left(\widehat{G}_r^{-1} - \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)W'V\left(s_C - m(\widehat{\vartheta}_{C,r}) + b_r\right) \\
&\quad + \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\widehat{h}_r - \widehat{F}_r\widehat{\vartheta}_{C,r}\right). \\
&= \left(\widehat{G}_r^{-1} - \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)W'V\left(s_C - m(\widehat{\vartheta}_{C,r}) + b_r\right) \\
&\quad + \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\left(\theta_C - D'm(\widehat{\vartheta}_{C,r}) - c_r\right),
\end{aligned}$$

using $\widehat{h}_r = \theta_C - D'm(\widehat{\vartheta}_{C,r}) + \widehat{F}_r\widehat{\vartheta}_{C,r} - c_r$. Upon convergence $\widehat{\vartheta}_{C,r+1} = \widehat{\vartheta}_{C,r}$ and $c_r = 0$, so that

$$\begin{aligned}
\widehat{h}_r - \widehat{F}_r\widehat{\vartheta}_{C,r} &= \theta_C - D'm(\widehat{\vartheta}_{C,r}) + \widehat{F}_r\widehat{\vartheta}_{C,r} - c_r - \widehat{F}_r\widehat{\vartheta}_{C,r} \\
&= \theta_C - D'm(\widehat{\vartheta}_{C,r}) = 0.
\end{aligned}$$

$$\widehat{\lambda}_r = -\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\left(W'V\left(s_C - m(\widehat{\vartheta}_{C,r})\right) + b_r\right).$$

Then, also $b_r = 0$ holds and

$$0 = \left(\widehat{G}_r^{-1} - \widehat{G}_r^{-1}\widehat{F}_r'\left(\widehat{F}_r\widehat{G}_r^{-1}\widehat{F}_r'\right)^{-1}\widehat{F}_r\widehat{G}_r^{-1}\right)\left(W'V\left(s - m(\widehat{\vartheta}_{C,r})\right)\right),$$

see Newey and McFadden (1994, p. 2219) and the projection estimator of Heyde and Morton (1993, p. 756).

# D Proof of Proposition 2, part I (the consistency proof)

(a) Non-stochastic counterpart to the likelihood:

$$Q_{0,C}(\vartheta_C) = \tfrac{Y_W}{C^2}\left(\sum_{j=1}^{C}\sum_{i=1}^{C}\nu_{ij}\left(m_{ij}(\vartheta_{C,0})\ln\left(m_{ij}(\vartheta_C)Y_W\right) - m_{ij}(\vartheta_C)\right)\right)$$

where $\vartheta_C = (\alpha,\ \phi_C(\alpha))$ and $\phi_C(\alpha)$ is continuous and continuously differentiable. Hence the restricted non-stochastic counterpart to the likelihood is given as

$$Q_{0,C}(\alpha, \phi_C(\alpha)) + \lambda'\left(\underbrace{D'm(\alpha, \phi_C(\alpha)) - \theta_C}_{0}\right) = Q_{0,C}(\alpha, \phi_C(\alpha))$$

and it is sufficient to consider $Q_{0,C}(\alpha, \phi_C(\alpha))$ to establish consistency.
(b) Likelihood under true DGP:

$$Q_C(\alpha, \phi_C(\alpha)) = \tfrac{Y_W}{C^2}\sum_{j=1}^{C}\sum_{i=1}^{C}\nu_{ij}((m_{ij}(\vartheta_{C,0}) + \varepsilon_{ij})\ln\left(m_{ij}(\vartheta_C)Y_W\right) - m_{ij}(\vartheta_C))$$

(c) Difference (b)-(a):

$$Q_C(\alpha, \phi_C(\alpha)) - Q_{0,C}(\alpha, \phi_C(\alpha)) = \tfrac{Y_W}{C^2}\sum_{j=1}^{C}\sum_{i=1}^{C}v_{ij}\varepsilon_{ij}\ln\left(m_{ij}(\phi_C(\alpha))Y_W\right)$$

(d) Identification follows from an argument put forward by Wooldridge (1997). For scalars $\mu_0$ and $\mu$ a function $f(\mu) = \mu_0\ln(\mu) - \mu$ is maximized at $\mu = \mu_0$ as $\frac{df(\mu)}{d\mu} = \frac{\mu_0}{\mu} - 1$ and $\frac{d^2f(\mu)}{d\mu^2} = -\frac{\mu_0}{\mu^2} < 0$. For finite $C$ it follows that $Q_{0,C}(\alpha, \phi_C(\alpha)) \leq Q_{0,C}(\alpha_0, \phi_C(\alpha_0))$ for $\alpha \neq \alpha_0$, since under Assumption Part I.6 $Z$ is of full column rank $K$ and varies at the bilateral level. Thus, $\alpha_0$ is a unique maximizer of $Q_{0,C}(\alpha, \phi_C(\alpha_0))$ at given $V$. Note $c_a/C^2 < m_{ij}(\alpha, \phi_C(\alpha)) < (1 - c_a)/C^2$ for some constant $0.5 > c_\alpha > 0$ and $\alpha \in \Theta$. Consider a summand in $Q_C(\alpha, \phi_C(\alpha))$.

$$\begin{aligned}
& Y_W\left|q_{ij,C}(\alpha)\right| \\
= \;& Y_W\left|(m_{ij}(\alpha_0, \phi_C(\alpha_0)) + \varepsilon_{ij})\ln\left(m_{ij}(\alpha, \phi_C(\alpha))Y_W\right) - m_{ij}(\alpha, \phi_C(\alpha))\right| \\
\leq \;& Y_W\left|m_{ij}(\alpha_0, \phi_C(\alpha_0))\right|\left|\ln\left(m_{ij}(\alpha, \phi_C(\alpha))Y_W\right)\right| + Y_W\left|m_{ij}(\alpha, \phi_C(\alpha))\right| \\
& + Y_W\left|\varepsilon_{ij}\right|\left|\ln\left(m_{ij}(\alpha, \phi_C(\alpha))Y_W\right)\right| \\
\leq \;& c_W C^2\tfrac{1-c_a}{C^2}\left|\ln(c_a c_W)\right| + (1 - c_a)c_W + c_W C^2\left|\varepsilon_{ij}\right|\left|\ln(c_a c_W)\right|
\end{aligned}$$

$$E[\sup_{\alpha\in\Theta'} |Y_W q_{ij,C}(\alpha)|] \leq c_W(1-c_a)\,|\ln(c_a c_W)| + c_W C^2 \tfrac{\overline{\sigma}}{C^2}\,|\ln(c_a c_W)| < \infty,$$

since $E\left[|\varepsilon_{ij}|\right] \leq E\left[|\varepsilon_{ij}|^2\right]^{\frac{1}{2}} \leq \frac{\overline{\sigma}}{C^2}$ by Lyaponov's inequality. The claim follows from the ULLN given in Pötscher and Prucha (2003), Theorem 23.

(e) $plim_{C\to\infty}\left[\sup_{\alpha\in\Theta'} |Q_C(\alpha,\phi_C(\alpha)) - Q_{0,C}(\alpha,\phi_C(\alpha))|\right] = o_p(1)$ can be proved directly.

$$\sup_{\alpha\in\Theta'} |Q_C(\alpha,\phi_C(\alpha)) - Q_{0,C}(\alpha,\phi_C(\alpha))| = \sup_{\alpha\in\Theta'}\left|\frac{1}{C^2}\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}\varepsilon_{ij}Y_W \ln\left(m_{ij}(\alpha,\phi_C(\alpha))Y_W\right)\right|$$

$$\leq \tfrac{1}{C^2}\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}\left|\varepsilon_{ij}c_W C^2\right| \left|\ln\left(\left(1-c_a\right)c_W\right)\right| = \left|\ln\left(\left(1-c_a\right)c_W\right)c_W\right| \sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}\left|\varepsilon_{ij}\right|$$

and Chebyshev's inequality implies

$$\begin{aligned}
P\left(\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}\left|\varepsilon_{ij}\right| \geq \kappa\right) &\leq \frac{1}{\kappa^2}\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij} Var(\varepsilon_{ij}) \\
&\leq \frac{1}{\kappa^2}\left(\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}\right)\frac{\overline{\sigma}^2}{C^4} \\
&= \frac{1}{\kappa^2}\frac{\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}}{C^2}\frac{\overline{\sigma}^2}{C^2} \to 0,
\end{aligned}$$

By Assumption Part I.3 $\sum_{j=1}^{C}\sum_{i=1}^{C} Var(\varepsilon_{ij}) = \sum_{j=1}^{C}\sum_{i=1}^{C}\frac{\sigma_{ij}^2}{C^4} < \frac{\overline{\sigma}^2}{C^2}$ and Assumption Part I.5 $\frac{\sum_{j=1}^{C}\sum_{i=1}^{C} v_{ij}}{C^2} = O(1)$. Therefore, consistency of $\widehat{\alpha}$ follows. Consistency of $\overline{\alpha}$ under the unrestricted model with dummies follows, since it also has $Q_{0,C}(\alpha,\phi_C(\alpha))$ as non-stochastic counterpart and the same arguments as above apply.

# E Proof of Proposition 2, Part 2 (limit distribution of $\widehat{\alpha}$):

The proof uses
(a) the mean value theorem where $G^*$ and $F^*$ are evaluated at $\vartheta_C^*$, whose elements lie (element-wise) in between those of $\widehat{\vartheta}_C$ and $\vartheta_{C,0}$ (see Newey and McFadden,

1994, p. 2219) :

$$\begin{bmatrix} \left.\frac{\partial \ln L_C(\vartheta_C)}{\partial \vartheta_C}\right|_{\vartheta_C=\widehat{\vartheta}_C} \\ \left.\frac{\partial \ln L_C(\vartheta_C)}{\partial \lambda}\right|_{\vartheta_C=\widehat{\vartheta}_C} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} W'V\varepsilon \\ 0 \end{bmatrix} + \begin{bmatrix} G^* & F^{*\prime} \\ F^* & 0 \end{bmatrix} \begin{bmatrix} \vartheta_C - \vartheta_{C,0} \\ \lambda \end{bmatrix}$$

(b)

$$Q_{G^{-1/2}F'(\alpha)} = \left[I - G^{-1/2}(\alpha)F'(\alpha)\left(F(\alpha)G^{-1}(\alpha)F'(\alpha)\right)^{-1}F(\alpha)G^{-1/2}(\alpha)\right]$$

(c)

$$\begin{aligned} s_\alpha(\alpha)_{(K\times 1)} &= C^2\sum_{i=1}^{C}\sum_{j=1}^{C}\left[I_K, 0_{K\times 2C-1}\right]G^{1/2}(\alpha)Q_{G^{-1/2}F'(\alpha)}G^{-1/2}(\alpha)w_{ij}v_{ij}\varepsilon_{ij} \\ &= A(\alpha)\varepsilon. \end{aligned}$$

(d)

$$B(\alpha) = Z'VM(\alpha)^{\frac{1}{2}}\left[I - M(\alpha)^{\frac{1}{2}}D(D'M(\alpha)D)^{-1}D'M(\alpha)^{\frac{1}{2}}\right]M(\alpha)^{\frac{1}{2}}Z$$

(e)

$$C(\widehat{\alpha} - \alpha_0) \quad = \quad -B(\alpha^*)^{-1}C^{-1}A(\alpha^*)\varepsilon$$

Claims:
(i)
$E\sup_{\alpha\in\Theta'}\|s_{\alpha,ij}(\alpha)\| < \infty$
$E\sup_{\alpha\in\Theta'}\|s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'\| < \infty$
$B(\alpha)$ is continuous in $\alpha$ and $\|B(\alpha) - B(\alpha_0)\| < \infty$.

(ii)
$C^{-1}s_\alpha(\alpha_0) \xrightarrow{d} N(0, A_0\Omega_\varepsilon A_0')$
$C(\hat{\alpha} - \alpha_0) \xrightarrow{d} N\left(0, B_0^{-1}A_0\Omega_\varepsilon A_0'B_0^{-1}\right)$,
where $A_0\Omega_\varepsilon A_0 = p\lim_{C\to\infty}\frac{1}{C^2}\sum_{i=1}^{C}\sum_{j=1}^{C}s_{\alpha,ij}(\alpha_0)s_{\alpha,ij}(\alpha_0)' = p\lim_{C\to\infty}\frac{1}{C^2}A(\alpha^*)\varepsilon\varepsilon'A(\alpha^*)'$
and $B_0 = p\lim_{C\to\infty}B(\alpha^*)$.

ad (i) Consider

$$\left[I_K, 0_{K\times 2C-1}\right]G^{1/2}(\alpha)Q_{G^{-1/2}F'}(\alpha)G^{-1/2}(\alpha)w_{ij}v_{ij}\varepsilon_{ij}$$

As $Q_{G^{-1/2}F'}(\alpha)$ is a projection matrix with $Q_{G^{-1/2}F'}(\alpha) = Q_{G^{-1/2}F'}(\alpha)'$ it follows that

$$\|Q_{G^{-1/2}F'}(\alpha)v\| \leq \|v\| \text{ for any vector } v.$$

Therefore, we have that

$$\begin{aligned}
&\|s_{\alpha,ij}(\alpha)\|^2 = \left\|C^2 \left[I_K, 0_{K\times 2C-1}\right] G^{1/2}(\alpha) Q_{G^{-1/2}F'}(\alpha) G^{-1/2}(\alpha) w_{ij} v_{ij} \varepsilon_{ij}\right\|^2 \\
\leq\ & \left\|C^2 \left[I_K, 0_{K\times 2C-1}\right] G^{1/2}(\alpha)\right\|^2 \left\|G^{-1/2}(\alpha) w_{ij} v_{ij}\right\|^2 \|\varepsilon_{ij}\|^2 \\
\leq\ & C^2 K c_z^2 (1-c_a) \left(K+2C-1\right)^2 c_g c_w^2 \|\varepsilon_{ij}\|^2,
\end{aligned}$$

using

$$\begin{aligned}
\left\|C^2 \left[I_K, 0_{K\times 2C-1}\right] G^{1/2}(\alpha)\right\|^2 &= C^2 tr\left(G^{1/2}(\alpha) \begin{bmatrix} I_K \\ 0 \end{bmatrix} \left[I_K, 0\right] G^{1/2}(\alpha)\right) \\
&= C^2 tr\left(\begin{bmatrix} I_K & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Z'VM(\alpha)VZ & Z'VM(\alpha)VD \\ D'VM(\alpha)VZ & D'VM(\alpha)VD \end{bmatrix}\right) \\
&= C^2 tr\left(\begin{bmatrix} Z'VM(\alpha)VZ & Z'VM(\alpha)VD \\ 0 & 0 \end{bmatrix}\right) \\
&= C^2 tr\left(\left[Z'VM(\alpha)VZ\right]\right) = C^2 \sum_{k=1}^{K} \sum_{i=1}^{C} \sum_{j=1}^{C} z_{ij,k}^2 v_{ij} m_{ij}(\alpha) \\
&\leq C^2 K C^2 c_z^2 \frac{1-c_a}{C^2} = C^2 K c_z^2 (1-c_a).
\end{aligned}$$

Let $v_{ij} = 1$ and denote the typical element of $G^{-1}(\alpha)$ by $g_{kl}$

$$\begin{aligned}
\left\|G^{-1/2}(\alpha) w_{ij} v_{ij}\right\|^2 &= tr\left(w_{ij}' G^{-1}(\alpha) w_{ij}\right) = tr\left(G^{-1}(\alpha) w_{ij} w_{ij}'\right) \\
&= \sum_{k=1}^{K+2C-1} \sum_{l=1}^{K+2C-1} w_{ij,k} g_{kl} w_{ij,l} \leq \left(K+2C-1\right)^2 c_{g,2} c_w^2.
\end{aligned}$$

It follows that

$$\begin{aligned}
\|s_{\alpha,ij}(\alpha)\|^2 &= \left\|C^2 \left[I_K, 0_{K\times 2C-1}\right] G^{1/2}(\alpha) Q_{G^{-1/2}F'(\alpha)} G^{-1/2}(\alpha) w_{ij(K+2C-1\times 1)} v_{ij} \varepsilon_{ij}\right\| \\
&\leq C^2 K c_z^2 (1-c_a) \left(K+2C-1\right)^2 c_{g,2} c_w^2 \|\varepsilon_{ij}\|^2.
\end{aligned}$$

By Assumption Part I.3 $E\left[\|\varepsilon_{ij}\|^2\right] = E\left[\frac{\sigma_{ij}^2}{C^4}\right] \leq \frac{\overline{\sigma}^2}{C^4}$, it follows that

$$E \sup_{\alpha\in\Theta'} \|s_{\alpha,ij}(\alpha)\|^2 \leq C^2 K c_z^2 (1-c_a) \left(K+2C-1\right)^2 c_{g,2} c_w^2 \frac{\overline{\sigma}^2}{C^4} < \infty.$$

Thereby, $\Theta'$ is a closed ball around $\alpha_0$ in the interior of $\Theta$. By Assumption Part II.2 the elements of $G^{-1}(\alpha)$ are uniformly bounded by a finite constant, $c_{g,2}$, and the elements of $W$ are uniformly bounded by $c_w$. Furthermore observe, that

$$\begin{aligned}
\|s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'\| &= \left(tr\left(s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'\right)\right)^{\frac{1}{2}} \\
&= tr\left(s_{\alpha,ij}(\alpha)'s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'s_{\alpha,ij}(\alpha)\right)^{\frac{1}{2}} \\
&= tr\left(s_{\alpha,ij}(\alpha)'s_{\alpha,ij}(\alpha)\right) \\
&= \|s_{\alpha,ij}(\alpha)\|^2
\end{aligned}$$

so that

$$E\sup_{\alpha\in\Theta'}\left[\|s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'\|\right] \leq \infty.$$

Hence, it follows that by Lemma 3.2 of Pötscher and Prucha (1997) that

$$\begin{aligned}
&\sup_{\alpha\in\Theta'}\left\|C^{-2}\sum_{i=1}^{C}\sum_{j=1}^{C}\left(s_{\alpha,ij}(\alpha)-E[s_{\alpha,ij}(\alpha)]\right)\right\| \xrightarrow{P} 0 \\
&\sup_{\alpha\in\Theta'}\left\|C^{-2}\sum_{i=1}^{C}\sum_{j=1}^{C}\left(s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)'-E[s_{\alpha,ij}(\alpha)s_{\alpha,ij}(\alpha)']\right)\right\| \xrightarrow{P} 0.
\end{aligned}$$

Lastly, consider

$$\begin{aligned}
B(\alpha) &= Z'VM(\alpha)^{\frac{1}{2}}\left[I-M(\alpha)^{\frac{1}{2}}D(D'M(\alpha)D)^{-1}D'M(\alpha)^{\frac{1}{2}}\right]M(\alpha)^{\frac{1}{2}}Z \\
&= Z'VM(\alpha)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha)M(\alpha)^{\frac{1}{2}}Z
\end{aligned}$$

$$\begin{aligned}
B(\alpha)-B(\alpha_0) &= Z'VM(\alpha)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha)M(\alpha)^{\frac{1}{2}}Z \\
&= \left(Z'VM(\alpha)^{\frac{1}{2}}-Z'VM(\alpha_0)^{\frac{1}{2}}\right)Q_{M^{1/2}D}(\alpha)M(\alpha)^{\frac{1}{2}}Z \\
&\quad +Z'VM(\alpha_0)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha)M(\alpha)^{\frac{1}{2}}Z \\
&\quad +Z'VM(\alpha_0)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha_0)M(\alpha)^{\frac{1}{2}}Z \\
&\quad -Z'VM(\alpha_0)^{\frac{1}{2}}Q_{M^{1/2}D}(\alpha_0)M(\alpha)^{\frac{1}{2}}Z \\
&\quad -Z'VM(\alpha_0)^{\frac{1}{2}}\left(Q_{M^{1/2}D}(\alpha_0)\right)M(\alpha_0)^{\frac{1}{2}}Z
\end{aligned}$$

$$
\begin{aligned}
B(\alpha) - B(\alpha_0) &= Z'V\left[M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right] Q_{M^{1/2}D}(\alpha) M(\alpha)^{\frac{1}{2}} Z \\
&\quad + Z'VM(\alpha_0)^{\frac{1}{2}}\left[Q_{M^{1/2}D}(\alpha_0) - Q_{M^{1/2}D}(\alpha)\right] M(\alpha)^{\frac{1}{2}} Z \\
&\quad + Z'VM(\alpha_0)^{\frac{1}{2}} Q_{M^{1/2}D}(\alpha_0)\left[M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right] Z
\end{aligned}
$$

Observe that

$$
\begin{aligned}
0 &< \left\| Z'VM(\alpha_0)^{\frac{1}{2}}\left[Q_{M^{1/2}D}(\alpha_0) - Q_{M^{1/2}D}(\alpha)\right] M(\alpha)^{\frac{1}{2}} Z \right\| \\
&\leq \left\| Z'VM(\alpha_0)^{\frac{1}{2}} \right\| \left\| Q_{M^{1/2}D}(\alpha_0) - Q_{M^{1/2}D}(\alpha) \right\| \left\| M(\alpha)^{\frac{1}{2}} Z \right\|
\end{aligned}
$$

$$
\begin{aligned}
\left\| Q_{M^{1/2}D}(\alpha_0) - Q_{M^{1/2}D}(\alpha) \right\|^2 &= tr(Q_{M^{1/2}D}(\alpha_0)) - tr\left(Q_{M^{1/2}D}(\alpha)\right) \\
&= \left(C^2 - 2C - 1\right) - \left(C^2 - 2C - 1\right) = 0
\end{aligned}
$$

$$
\begin{aligned}
Z'VM(\alpha)^{\frac{1}{2}} &= Z'VM(\alpha_0)^{\frac{1}{2}} \\
&\quad + \tfrac{1}{2} Z'VM(\alpha^*)^{\frac{1}{2}} \left( Z + \left.\tfrac{\partial \phi(\alpha)}{\partial \alpha'}\right|_{\alpha=\alpha^*} \right) (\alpha - \alpha_0) \\
\left\| Z'V\left(M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right) \right\| &\leq \left\| Z'VM(\alpha^*)^{\frac{1}{2}} \left( Z + D \left.\tfrac{\partial \phi(\alpha)}{\partial \alpha'}\right| \right)_{\alpha=\alpha^*} \right\| \left\| \alpha - \alpha_0 \right\| \\
&\leq \tfrac{1}{2} \left( \tfrac{c_a}{C^2} K C^2 \left(c_z + c_\phi\right) \right)^{\frac{1}{2}} \left\| \alpha - \alpha_0 \right\| \\
&= L \left\| \alpha - \alpha_0 \right\|.
\end{aligned}
$$

The elements of $\left| Z + D \left.\frac{\partial \phi(\alpha)}{\partial \alpha'}\right|_{\alpha=\alpha^*} \right|$ are uniformly bounded by $c_z + c_\phi$ (Assumptions Part I.6 and Part II. 2) so that $L$ is a finite positive constant. Note the assumption of a constant share of missings. A similar conclusion holds for $\left\| \left(M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right) Z \right\|$. Hence,

$$
\begin{aligned}
\left\| B(\alpha) - B(\alpha_0) \right\| &\leq \left\| Z'V\left(M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right) \right\| \left\| M(\alpha)^{\frac{1}{2}} Z \right\| \\
&\quad + \left\| Z'VM(\alpha_0)^{\frac{1}{2}} \right\| \left\| \left(M(\alpha)^{\frac{1}{2}} - M(\alpha_0)^{\frac{1}{2}}\right) Z \right\| \\
&= \left( \left\| M(\alpha)^{\frac{1}{2}} Z \right\| + \left\| Z'VM(\alpha_0)^{\frac{1}{2}} \right\| \right) L \left\| \alpha - \alpha_0 \right\| \\
&\leq 2 \left( K C^2 c_z^2 \frac{c_a}{C^2} \right)^{\frac{1}{2}} L \left\| \alpha - \alpha_0 \right\|
\end{aligned}
$$

which proves continuity of $B(\alpha)$. The elements of $Z$ are bounded away from zero and from above. $Q_{M^{1/2}D}(\alpha)$ projects $M(\alpha, Z^c)^{\frac{1}{2}} Z^c$ onto the orthogonal complement of the hyperplane spanned by $D'M(\alpha, Z)^{\frac{1}{2}}$ (in the exporter and importer dimension), while $Z$ exhibits bilateral variation. Further, the rank of $B(\alpha)$ is $K$ and it follows by Theorem 14 of Pötscher and Prucha (2003) that $B(\alpha^*) - B(\alpha_0) = o_p(1)$.

ad (ii) The Lyapunov central limit theorem for triangular arrays (Billingsley, 1995, Theorem 27.3) and the Cramer-Wold device can be applied to derive

$$C^{-1} s_\alpha(\widehat{\alpha}) \xrightarrow{d} N(0, A_0 \Omega_\varepsilon A_0')$$

and

$$C(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, B_0^{-1} A_0 \Omega_\varepsilon A_0' B_0^{-1}).$$

For estimation one uses

$$\widehat{B} = B(\widehat{\alpha}) \xrightarrow{p} B_0,\ C^{-2} A(\widehat{\alpha}) \widehat{\Omega}_\varepsilon A(\widehat{\alpha}) \xrightarrow{p} A_0 \Omega_\varepsilon A_0'.$$

# F Proof of Proposition 3:

Part 1 (see Pollard, 2002, p. 184):

$$\begin{aligned} CSm(\widehat{\alpha}, Z) &= CSm(\alpha_0, Z^c) + C\Gamma^c_{C,0} (\widehat{\alpha} - \alpha_0) + o_p \|C(\widehat{\alpha} - \alpha_0)\| \\ &= CSm(\alpha_0, Z^c) + C\Gamma^c_{C,0} (\widehat{\alpha} - \alpha_0) + o_p(O_p(1)) \end{aligned}$$

Claims:
(i) Let

$$\Gamma_C(\alpha_0, Z^c)_{s \times K} = SM(\alpha_0, Z^c)^{\frac{1}{2}} Q_{M^{1/2}D}(\alpha) M(\alpha, Z^c)^{\frac{1}{2}} Z^c.$$

The elements of $\Gamma_0^c = \lim_{C \to \infty} \Gamma_C(\alpha_0, Z^c)$ are finite, non-zero and have rank $s$ so that

$$C\Gamma_0^c (\widehat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \Gamma_0^c V_a \Gamma_0^{c\prime}).$$

(ii) $p\lim_{C \to \infty} \Gamma_C^c(\widehat{\alpha}) = \Gamma_0^c$

ad (i) Remember that the normalized selection matrix $S$ has finite dimension so that $SM(\alpha_0, Z)$ possesses typical non-zero element $c_a \leq C^2 m_{ij}(\alpha_{0,} z_{ij}) \leq 1 - c_a$. The elements $\Gamma_C^c(\alpha_0)$ are bounded away from zero, since $Q_{M^{1/2}D}(\alpha)$ projects $M(\alpha, Z^c)^{\frac{1}{2}} Z^c$ onto the orthogonal complement of the hyperplane spanned in the exporter and importer dimension by $D'M(\alpha, Z^c)^{\frac{1}{2}}$, while $Z^c$ exhibits bilateral variation. Hence, the rank $\Gamma_C(\alpha_0, Z^c)$ is $s$ as $M(\alpha_0, Z^c)^{\frac{1}{2}}$ is a diagonal matrix with non-zero diagonal elements with lower bound $\left(C^2 \frac{c_a}{C^2}\right)^{\frac{1}{2}} > 0$ and upper bound $\left(C^2 \frac{1-c_a}{C^2}\right)^{\frac{1}{2}}$

$< \infty$ by Assumption Part I.4, respectively. By Assumption Part I.6 the elements of $Z$ are uniformly bounded and by Assumption Part II.6 $\lim_{C\to\infty} \Gamma_C(\alpha_0, Z^c)$ is assumed to exist with rank $s$.

ad (ii): The claim follows from the continuity of $\Gamma_C^c(\alpha)$, which can be proved using the same arguments as in the proof of Proposition 2, and Pötscher and Prucha (2003, Theorem 14). Therefore, it holds that $\widehat{\Gamma}^c - \Gamma_0^c = o_p(1)$ and $\widehat{\Gamma} - \Gamma_0 = o_p(1)$. The claim then follows from the limit distribution of $\widehat{\alpha}$ given in Proposition 2 and Corollary 5 in Pötscher and Prucha (2003).

(Part 2) For percent changes define the selection matrix $R$ so that $RM(\alpha_0, Z)^{-1}$ typical non-zero element $m_{ij,C}(\alpha_{0,} z_{ij}^c)^{-1}$ and observe that

$$RM(\widehat{\alpha}, Z)^{-1} m(\widehat{\alpha}, Z^c)$$

has typical non-zero element

$$e^{\left(z_{ij}^c - z_{ij}\right)'\widehat{\alpha} + \beta_{i,C}^c(\widehat{\alpha}) + \gamma_{i,C}^c(\widehat{\alpha}) - \beta_{i,C}(\widehat{\alpha}) - \gamma_{j,C}(\widehat{\alpha})}$$

Taylor series expansion leads to

$$\begin{aligned}
& C\left(RM(\widehat{\alpha}, Z)^{-1} m(\widehat{\alpha}, Z^c) - RM(\alpha_0, Z)^{-1} m(\alpha_0, Z^c)\right) \\
= \;& C\Bigg( R\,\underbrace{M(\alpha_0, Z)^{-1} M(\alpha_0, Z^c)\left(Z^c - D\frac{\partial \phi_C^c}{\partial \alpha'}\right)}_{\Upsilon_C(\alpha_0, Z^c)}(\widehat{\alpha} - \alpha_0) \\
& - R\,\underbrace{\left(Z - D\frac{\partial \phi_C}{\partial \alpha'}\right)(\widehat{\alpha} - \alpha_0)}_{\Upsilon_C(\alpha_0, Z)}\Bigg) + o_p\left(1\right) \\
= \;& CR\left(\Upsilon_C(\alpha_0, Z^c) - \Upsilon_C(\alpha_0, Z)\right)(\widehat{\alpha} - \alpha_0) + o_p\left(1\right),
\end{aligned}$$

where

$$\begin{aligned}
\Upsilon_C(\alpha_0, Z^c) &= RM(\alpha_0, Z)^{-1} M(\alpha_0, Z^c)^{\frac{1}{2}} Q_{M^{1/2}D}(\alpha) M(\alpha, Z^c)^{\frac{1}{2}} Z^c \\
\Upsilon_C(\alpha_0, Z) &= RM(\alpha_0, Z)^{-\frac{1}{2}} Q_{M^{1/2}D}(\alpha) M(\alpha, Z)^{\frac{1}{2}} Z.
\end{aligned}$$

The elements $\Upsilon_C(\alpha_0, Z^c)$ are bounded away from zero, since $Q_{M^{1/2}D}(\alpha)$ projects onto the orthogonal complement of the hyperplane spanned by $D'M(\alpha, Z^c)^{\frac{1}{2}}$. The rank of $\Upsilon_C(\alpha_0, Z^c)$ is $r < K$ as $RM(\alpha_0, Z)^{-1}$ is a diagonal matrix with non zero diagonal elements $\frac{C^2}{1-c_a} \leq m_{ij}(\alpha_0, Z)^{-1} \leq \frac{C^2}{c_a}$. Further, the rank of $\Upsilon_C(\alpha_0, Z^c)$

is $K$ and $\|\Upsilon_C(\alpha_0, Z^c)\| \leq R\,\|M(\alpha_0, Z)^{-1}\|\,\|M(\alpha_0, Z^c)Z^c\| \leq (Rc_a(1-c_a)c_z)^{\frac{1}{2}} = O(1)$. The same arguments apply for $\Upsilon_C(\alpha_0, Z)$. By Assumption Part II.4 $\Upsilon_0^c = \lim_{C\to\infty}\Upsilon_C(\alpha_0, Z^c)$ and $\Upsilon_0 = \lim_{C\to\infty}\Upsilon_C(\alpha_0, Z)$ are assumed to exist with rank $r$.

Similar to Part 1 $\Upsilon_C(\widehat{\alpha}, Z^c) - \Upsilon_C(\alpha_0, Z^c) = o_p(1)$ and $\Upsilon_C(\widehat{\alpha}, Z) - \Upsilon_C(\alpha_0, Z) = o_p(1)$ by the continuity of $\Upsilon_C(\alpha, Z)$ and the claim follows.

# G Remarks on Proposition 2:

**Remark 2**: The comparison of dummy PPML and constrained PPML

(a) In order to derive the limit distribution of the dummy PPML, we define $\overline{G}^* = W'V\overline{M}^*VW$ with $\overline{M}^* = M(\overline{\alpha}^*, \overline{\phi}_C^*)$, where $\overline{\vartheta}^* = (\overline{\alpha}^*, \overline{\phi}_C^*)$ lies elementwise between $\overline{\vartheta}$ and $\vartheta_0$. Applying the mean-value theorem, the score of the unconstrained likelihood yields

$$0 = W'V\varepsilon - \overline{G}^*\begin{bmatrix}\overline{\alpha} - \alpha_0 \\ \overline{\phi}_C - \phi_C(\alpha_0)\end{bmatrix}.$$

The inverse of $\overline{G}^*$ has blocks

$$\begin{aligned}
\overline{G}^{*11} &= \left(Z'V\overline{M}^{*\frac{1}{2}}Q_{\overline{M}^{*\frac{1}{2}}VD}\overline{M}^{*\frac{1}{2}}VZ\right)^{-1} \\
\overline{G}^{*12} &= -G_0^{11}Z'V\overline{M}^{*\frac{1}{2}}VD\left(D'V\overline{M}^*VD\right)^{-1} \\
\overline{G}^{*22} &= \left[D'V\left(\overline{M}^* - \overline{M}^*VZ\left(Z'V\overline{M}^*VZ\right)^{-1}Z'V\overline{M}^*\right)VD\right]^{-1}
\end{aligned}$$

and it holds that

$$\begin{aligned}
\begin{bmatrix}\overline{\alpha} - \alpha_0 \\ \overline{\phi}_C - \phi_C(\alpha_0)\end{bmatrix} &= \begin{bmatrix}\overline{G}^{*11} & \overline{G}^{*12} \\ \overline{G}^{*21} & \overline{G}^{*22}\end{bmatrix} W'V\varepsilon \\
\overline{\alpha} - \alpha_0 &= \overline{G}^{*11}(Z' - \overline{G}_{12}^*\overline{G}_{22}^{*-1}D')V\varepsilon.
\end{aligned}$$

In terms of the results of Proposition 2 it is easily seen that

$$\begin{aligned}
\overline{B}^{*-1} &= \overline{G}^{*11} \\
\overline{A}^* &= C^2Z'\overline{M}^{*\frac{1}{2}}\left(I - \overline{M}^{*\frac{1}{2}}VD\left(D'V\overline{M}^*VD\right)^{-1}D'\overline{M}^{*\frac{1}{2}}\right)\overline{M}^{*-\frac{1}{2}}V \\
&= C^2Z'\overline{M}^{*\frac{1}{2}}Q_{\overline{M}^{*\frac{1}{2}}VD}\overline{M}^{*-\frac{1}{2}}V.
\end{aligned}$$

The comparison of dummy PPML and constrained PPML is straight forward under fully observed trade flows with $V = I_{C^2}$ . In this case we have $B^{*-1} = \overline{B}^{*-1} = \overline{G}^{*11}$ and (dropping arguments)

$$
\begin{aligned}
\left(FG^{-1}F'\right)^{-1} &= \left( \left[\begin{array}{cc} G_{21} & G_{22} \end{array}\right] \left[\begin{array}{cc} G^{11} & -G^{11}G_{12}G_{22}^{-1} \\ -G_{22}^{-1}G_{21}G^{11} & G^{22} \end{array}\right] \begin{array}{c} G_{21} \\ G_{22} \end{array} \right)^{-1} \\
&= \left( \left[\begin{array}{cc} G_{21}G^{11} - G_{21}G^{11}, & -G_{21}G^{11}G_{12}G_{22}^{-1} + G_{22}G^{22} \end{array}\right] \begin{array}{c} G_{21} \\ G_{22} \end{array} \right)^{-1} \\
&= -\left(G_{21}G^{11}G_{12} + G_{22}G^{22}G_{22}\right)^{-1}
\end{aligned}
$$

and

$$
\begin{aligned}
FG^{-1}W' &= \left( \left[\begin{array}{cc} G_{21} & G_{22} \end{array}\right] \left[\begin{array}{cc} G^{11} & -G^{11}G_{12}G_{22}^{-1} \\ -G_{22}^{-1}G_{21}G^{11} & G^{22} \end{array}\right] \left[\begin{array}{c} Z' \\ D' \end{array}\right] \right) \\
&= \left[\begin{array}{cc} G_{21}G^{11} - G_{21}G^{11}, & -G_{21}G^{11}G_{12}G_{22}^{-1} + G_{22}G^{22} \end{array}\right] \left[\begin{array}{c} Z' \\ D' \end{array}\right] \\
&= \left(-G_{21}G^{11}G_{12} + G_{22}G^{22}G_{22}\right) G_{22}^{-1}D'
\end{aligned}
$$

Therefore, the term $\left(FG^{-1}F'\right)^{-1} FG^{-1}W'$ reduces to $G_{22}^{-1}D'$. It follows that

$$
\begin{aligned}
\widehat{\alpha} - \alpha_0 &= G^{*11}[I_K, 0] \left( \left[\begin{array}{c} Z' \\ D' \end{array}\right] - \left[\begin{array}{c} G_{12}^{*}G_{22}^{*-1}D' \\ G_{22}^{*}G_{22}^{*-1}D' \end{array}\right] \right) \varepsilon \\
&= G^{*11}\left(Z' - G_{12}^{*}G_{22}^{*-1}D'\right) \varepsilon
\end{aligned}
$$

Thus constrained PPML and dummy PPML estimators of $\alpha_0$ have the same limit distribution in this case.

**Remark 3**: The estimation of $Var\left(\hat{\alpha}\right)$ and $Var\left(\overline{\alpha}\right)$ under fully observed trade flows

The difference in the variance estimation between the two estimators is best illustrated for the case of fully observed trade flows. Defining $\widetilde{Z} = M_0^{1/2}Z$, $\widetilde{D} = M_0^{1/2}D$ and $Q_{\widetilde{D}} = I_{C^2} - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}'$ the residuals of constrained PPML are based on

$$
\begin{aligned}
A(\alpha_0) &= C^2Z'\left[I_{C^2} - M_0D(D'M_0D)^{-1}D'\right] \\
&= C^2Z'M_0^{1/2}\left[I_{C^2} - M_0^{1/2}D(D'M_0D)^{-1}D'M_0^{1/2}\right]M_0^{-1/2} \\
&= C^2\widetilde{Z}'Q_{\widetilde{D}}M_0^{-1/2}
\end{aligned}
$$

and

$$\left(\widetilde{Z}'Q_{\widetilde{D}}\widetilde{Z}\right)(\widehat{\alpha}-\alpha_0) = \widetilde{Z}'Q_{\widetilde{D}}M_0^{-1/2}\varepsilon + o_p(1)$$

Ignoring the remainder, the residuals estimated by constrained PPML can be written as

$$\begin{aligned}
M_0^{-1/2}\widehat{\varepsilon} &= M_0^{-1/2}\varepsilon - Q_{\widetilde{D}}\widetilde{Z}(\widehat{\alpha}-\alpha_0) \\
&= M_0^{-1/2}\varepsilon - Q_{\widetilde{D}}\widetilde{Z}\left(\widetilde{Z}'Q_{\widetilde{D}}\widetilde{Z}\right)^{-1}\widetilde{Z}'Q_{\widetilde{D}}M_0^{-1/2}\varepsilon \\
M_0^{-1/2}\widehat{\varepsilon} &= H_{Q_{\widetilde{D}}\widetilde{Z}}M_0^{-1/2}\varepsilon
\end{aligned}$$

with $H_{Q_{\widetilde{D}}\widetilde{Z}} = I - P_{Q_{\widetilde{D}}\widetilde{Z}}$, $P_{Q_{\widetilde{D}}\widetilde{Z}} = Q_{\widetilde{D}}\widetilde{Z}\left(\widetilde{Z}'Q_{\widetilde{D}}\widetilde{Z}\right)^{-1}\widetilde{Z}'Q_{\widetilde{D}}$ being symmetric and idempodent. In contrast, dummy PPML implies that (again ignoring the remainder)

$$\overline{\varepsilon} = \varepsilon - WG_0^{-1}W'\varepsilon = M_0^{1/2}H_{\widetilde{W}}M_0^{-1/2}\varepsilon,$$

where $H_{\widetilde{W}} = I_{C^2} - \widetilde{W}\left(\widetilde{W}'\widetilde{W}\right)^{-1}\widetilde{W}$ is symmetric and idempotent. Let $P_{\widetilde{D}} = \widetilde{D}\left(\widetilde{D}'\widetilde{D}\right)^{-1}\widetilde{D}'$. Since $H_{\widetilde{W}}$ can be factored as $H_{\widetilde{W}} = \left(I_{C^2} - P_{Q_{\widetilde{D}}\widetilde{Z}}\right)\left(I_{C^2} - P_{\widetilde{D}}\right)$ it follows that $H_{\widetilde{W}} - I_{C^2} = P_{Q_{\widetilde{D}}\widetilde{Z}} + P_{\widetilde{D}}$. Following Chesher and Jewitt (1987) the proportionate bias of $\widehat{V}_\alpha$ is defined as $pb(\widehat{V}_\alpha) = E\left[\frac{v'\widehat{V}_\alpha v}{v'V_\alpha v}\right]$ for some vector $v \neq 0$ and similarly for $\overline{V}_\alpha$. For the presten estimators it follows that

$$\begin{aligned}
v'\widehat{V}_\alpha v &= v'B(\alpha_0)^{-1}\widetilde{Z}Q_{\widetilde{D}}diag(M_0^{-1}\widehat{\varepsilon}\widehat{\varepsilon})Q_{\widetilde{D}}\widetilde{Z}'B(\alpha_0)^{-1}v \\
&= z'diag(\widehat{\varepsilon}\widehat{\varepsilon})z
\end{aligned}$$

with $z = M_0^{-1/2}Q_{\widetilde{D}}\widetilde{Z}'B(\alpha_0)^{-1}v$. Chesher and Jewitt (1987) demonstrate that the proportionate bias in general depends on the degree of heteroskedasticity and on the features of the data as represented by the main diagonal elements $h^c_{ij,ij}$ of $P_{Q_{\widetilde{D}}\widetilde{Z}}$ and $h^d_{ij,ij}$ of $H_{\widetilde{W}} - I_{C^2}$, respectively. Thereby, $1 - h^l_{ij,ij}$, $l = c, d$, are used as a measures of leverage. Further, Chesher and Jewitt (1987) show that an upper bound of $pb(\widehat{V}_\alpha)$ is given as

$$sup_z\left(pb(\widehat{V}_\alpha)\right) \leq \max_{ij}\sum_{l=1,l\neq i}^{C}\sum_{k=1,l\neq j}^{C}\frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}}\left(h^c_{ij,lk}\right)^2 + h^c_{ij,ij}(h^c_{ij,ij}-2).$$

This uses the multiplicative error model with $\varepsilon = M_0(\eta - \iota_{C^2})$ and sets $z = M_0^{1/2}Q_{\widetilde{D}}\widetilde{Z}'B(\alpha_0)^{-1}v$. Since $0 \leq \sum_{l=1,l\neq i}^{C}\sum_{k=1,l\neq j}^{C}\left(h^c_{ij,lk}\right)^2 = h^c_{ij,ij}\left(1 - h^c_{ij,ij}\right) \leq \frac{1}{2}$,

$\frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}} \leq \frac{\overline{\sigma}^2_\eta}{\underline{\sigma}^2_\eta}$, the elements of the main diagonal of $P_{\widetilde{D}}$ lie in $[0,1]$ and its trace is $2C-1$, it follows that $h^d_{ij,lk} = h^c_{ij,lk} + O(C^{-1})$ and

$$\begin{aligned} pb(\widehat{V}_\alpha) &\leq \tfrac{1}{2}\left(\frac{\overline{\sigma}^2_\eta}{\underline{\sigma}^2_\eta} - 1\right) - \min(h^c_{ij,ij}) \\ pb(\overline{V}_\alpha) &\leq \tfrac{1}{2}\left(\frac{\overline{\sigma}^2_\eta}{\underline{\sigma}^2_\eta} - 1\right) - \min(h^c_{ij,ij}) - O(C^{-1}), \end{aligned}$$

showing that the upper bound of the proportionate bias of $\overline{V}_\alpha$ is lower than that of $\widehat{V}_\alpha$. The difference is $O(C^{-1})$.

**Remark 4**: The restriction $D'\varepsilon = 0$

Often data are constructed such that $D's_C = \theta_C$ and $V = I_{C^2}$. For example, domestic trade flows can be derived as $s_{ii} = \kappa_{C,i} - \sum_{j=1,j\neq i}^{C} s_{ij}$ and $s_{Cj} = \theta_{C,j} - \sum_{i=1,i\neq C}^{C} s_{Cj}$. This poses restrictions on the disturbances such that $D'\varepsilon = 0$ and the score of both dummy and constrained PPML reduces to

$$C\left(\overline{\alpha} - \alpha_0\right) = C\overline{G}^{*11} Z'\varepsilon.$$

A simple possible specification of the disturbances would be to assume that $\varepsilon_{ij}$ is independently distributed as $(0, \sigma^2_{ij})$ for $i \neq j$ and $i \neq C$, while $\varepsilon_{ii} = -\sum_{j=1,j\neq i}^{C} \varepsilon_{ij}$ and $\varepsilon_{Cj} = -\sum_{i=1,i\neq C}^{C} \varepsilon_{ij}$. One can partition the data so that $\varepsilon = (\varepsilon'_R, \varepsilon'_U)'$, where $\varepsilon_R$ includes $\varepsilon_{ii}$, $i = 1,..,C-1$ and $\varepsilon_{Cj}$, $j = 1,..,C$, while $\varepsilon_U$ in comprises the remaining disturbances that are assumed to be distributed independently $(0, \sigma^2_{ij})$. The matrices $D$ and $Z$ are be partitioned in the same way. Then one can write

$$D'\varepsilon = D'_R\varepsilon_R + D'_U\varepsilon_U = 0.$$

Note, $D_R$ has dimension $2C-1\times\ 2C-1$ and it is invertible so that

$$\varepsilon_R = -D'^{-1}_R D'_U \varepsilon_U.$$

Inserting yields

$$C\left(\overline{\alpha} - \alpha_0\right) = C\overline{G}^{*11}\left[Z'_R, Z'_U\right]\begin{bmatrix} -D'^{-1}_R D'_U \\ I_R \end{bmatrix}\varepsilon_U = C\overline{G}^{*11}\left[-Z'_R D'^{-1}_R D'_U + Z'_U\right]\varepsilon_U$$

$$A(\alpha) = C^2\left[-Z'_R D'^{-1}_R D'_U + Z'_U\right]$$

with

$$\overline{G}^{*11} \xrightarrow{p} B_0^{-1}$$
$$C^{-2} A(\widehat{\alpha}) \widehat{\varepsilon}_U' \widehat{\varepsilon}_U A(\widehat{\alpha})' \xrightarrow{p} \lim_{C\to\infty} C^{-2} \left(-Z_R' D_R'^{-1} D_U' + Z_U'\right) \Omega_U \left(-Z_R' D_R'^{-1} D_U' + Z_U'\right)'.$$

# H Data Appendix

The data on trade flows, $x_{ii,C}$, production, $Y_i$, and expenditure, $E_i$, are corrected for trade with the rest of the world as well as for trade imbalances. The total production value of country $i$ is given as $x_{i.,C} = \sum_j x_{ij,C} + x_{i,ROW,C}$ and total expenditures by $x_{.i,C} = \sum_j x_{ji,C} + x_{ROW,i,C}$ so that the trade balance is $d_{i,C} = x_{i.,C} - x_{.i,C}$. Since data are available for 59 countries, exports to the rest of the world (ROW) and imports from ROW of country $i$ have been aggregated in $s_{i,ROW,C}$ and $s_{ROW,i,C}$. Domestic shipments are implicitly defined as

$$\begin{aligned}
\kappa_{i,C} &= \frac{x_{i.,C} - x_{i,ROW,C}}{Y_W} = s_{ii,C} + \sum_{j\neq i}^{C} s_{ij,C} \\
\theta_{i,C} &= \frac{x_{.i,C} - d_{i,C} - x_{ROW,i,C}}{Y_W} = s_{ii,C} + \sum_{h\neq i}^{C} s_{hi,C}.
\end{aligned}$$

where $Y_W$ denotes overall (world) production or expenditure for the 59 countries. Note that $\sum_{i=1}^{C} d_{i,C} = 0$ per definition and that $\sum_{i=1}^{C} \kappa_{i,C} = \sum_{j=1}^{C} \theta_{i,C} = 1$. Aggregate exports and imports from OECD-Stan allow the calculation of $x_{i,ROW,C}$, $x_{ROW,i,C}$ and $d_{i,C}$ and in turn the remaining figures. The data then fulfil the above mentioned aggregation restrictions, $\kappa_{i,C} = \sum_{i=1}^{C} s_{ij,C}$ and $\theta_{i,C} = \sum_{j=1}^{C} s_{ij,C}$, and thus imply the restrictions of the disturbances as discussed in the text.

## References


Abadir, K. M. and J. R. Magnus (2005), *Matrix Algebra* Vol. 1, Cambridge University Press, Cambridge UK.

Anderson, J.E. and E. van Wincoop (2003), Gravity with Gravitas: A Solution to the Border Puzzle, *American Economic Review* **93**(1), 170-192.

Anderson, J. E. and Y. V. Yotov (2016), The Changing Incidence of Geography, *American Economic Review*, **100**(5), 2157–2186.

Arkolakis, C., A. Costinot and A. Rodriguez-Clare (2012), New Trade Models, Same Old Gains?, *American Economic Review* **102**(1), 94-130.

Allen, T., C. Arkolakis and Y. Takahashi (2014), Universal Gravity, NBER Working Paper (w20787).

Arvis, J.-F. and B. Shepherd (2013), The Poisson Quasi-Maximum Likelihood Estimator: A Solution to the 'Adding up' Problem in Gravity Models, *Applied Economic Letters* **20**(6), 515–519.

Baier, S. L. and J.H. Bergstrand (2009), Estimating the Effects of Free Trade Agreements on International Trade Flows Using Matching Econometrics, *Journal of International Economics* **77**(1), 63–76.

Bergstrand, J.H., P.H. Egger and M. Larch (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.

Berry, St., O.B. Linton and A. Pakes (2007), Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems, *Review of Economic Studies* **71** (3), 613-654.

Billingsley, P. (1995), Probability and Measure, Wiley-Interscience Publication, John Wiley & Sons, New York.

Cameron, C.A, J. B. Gelbach and D. L. Miller (2011), Robust Inference With Multiway Clustering, *Journal of Business and Economic Statistics* **29**(2), 238-249.

Costinot, A. and A. Rodríguez-Clare (2014), Trade Theory with Numbers: Quantifying the Consequences of Globalization. Chapter 4 in Gopinath, G, E. Helpman and K. S. Rogoff (eds), Handbook of International Economics Vol. 4, Elsevier, Oxford, 197–261.

Davidson, R. and J.G. MacKinnon (1998), Graphical Methods for Investigating the Size and Power of Hypothesis Tests, *Manchester School* **66**(1), 1–26.

Chesher A. and I. Jewitt (1987), The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator, *Econometrica* **55**(5), 1217–1222.

Cribari–Neto, F., S.L. Ferrari and G.M. Cordeiro (2000), Improved Heteroscedasticity–Consistent Covariance Matrix Estimators, *Biometrika* **87**(4), 907-918.

Eaton, J. and S. Kortum (2002), Technology, Geography, and Trade, *Econometrica* **70**(5), 1741–1779.

Egger, P.H., M. Larch, K.E. Staub and R. Winkelmann (2011), The Trade Effects of Endogenous Preferential Trade Agreements, *American Economic Journal: Economic Policy* **3**(3), 113–143.

Egger P. H. and S. Nigai (2015), Structural Gravity with Dummies Only: Constrained ANOVA-Type Estimation of Gravity Models, *Journal of International Economics* **97**(1), 86–99.

Egger P. H. and K.E. Staub (2016), GLM Estimation of Trade Gravity Models with Fixed Effects, *Empirical Economics* **50**(1), 137–175.

Egger, P.H. and F. Tarlea (2015), Multi-way Clustering Estimation of Standard Errors in Gravity Models, *Economic Letters* 134, 144–147.

Fally, T. (2015), Structural Gravity and Fixed Effects, *Journal of International Economics* **97**(1), 76–85.

Falocci, N.,R. Paniccià and E. Stanghellini (2009), Regression Modelling of the Flows in an Input–Output Table with Accounting Constraints, *Statistical Papers* **50**(2), 373–382.

Fernandez-Val, I. and M. Weidner (2016), Individual and Time Effects in Nonlinear Panel Models with large N, T, *Journal of Econometrics* **192(1),** 291–312.

Heyde, C.C. and R. Morton (1993), On Constrained Quasi-Likelihood Estimation, *Biometrika* **80**(4), 755-761.

Imbens G. W. and M. Kolesar (2016), Robust Standard Errors in Small Samples: Some Practical Advice, *Review of Economics and Statistics* **98**(4), 701–712.

Krugman, P.R. (1979), Increasing Returns, Monopolistic Competition, and International Trade, *Journal of International Economics* **9**(4), 469–479.

Larch, M. and Y. V. Yotov (2016), General Equilibrium Trade Policy Analysis with Structural Gravity, *CESifo Working Paper 6020.*

Mayer, T. and Zignago, S. (2011), Notes on CEPII's distances measures : the GeoDist Database, *CEPII Working Paper* 2011-25.

Newey, W. K. and D. L. McFadden (1994), Large Sample Estimation and Hypothesis Testing, in Engle R.F. and D. McFadden (eds.), Handbook of Econometrics. Vol. 4., Elsevier Science, Amsterdam.

Pollard, D. (2002), A User's Guide to Measure Theoretic Probability (Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge UK.

Pötscher, B.M. and I.R Prucha (1997), Dynamic Nonlinear Econometric Models: Asymptotic Theory, Springer, Berlin Heidelberg.

Pötscher, B.M. and I.R Prucha (2003), Basic Elements of Asymptotic Theory in Baltagi, B. ed. (2003), A Companion to Theoretical Econometrics. Blackwell, Malden, Mass. and Oxford.

Santos Silva, J.M.C. and S. Tenreyro (2006), The Log of Gravity, *Review of Economics and Statistics* **88**(4), 641–658.

Santos Silva, J.M.C. and F.A.G Windmeijer (1997), Endogeneity in Count Data Models: An Application to Demand for Health Care, *Journal of Applied Econometrics* **12**(3), 281–294.

Su, Ch. and K.L. Judd (2012), Constrained Optimization Approaches to Estimation of Structural Models, *Econometrica* **80**(5), 2213–2230.

Sydsaeter, K. , P. Hammond, A. Seierstad and A. Strøm (2005), Further Mathematics for Economic Analysis, Pearson, Edinburgh.

Wooldridge, J. (1997), Quasi-Likelihood Methods for Count Data, in Pesaran H. and P. Schmidt (eds.), Handbook of Applied Econometrics Volume II: Micro-economics, Blackwell, Oxford UK.

Yotov, Y. V. (2012). A Simple Solution to the Distance Puzzle in International Trade. *Economics Letters*, 117(3): 794–798.