Reassessing international investment patterns: a revisitation of Lane and Milesi-Ferretti's evidence.

Pierucci, Eleonora and Pericoli, Filippo and Ventura, Luigi

7 February 2014

Reassessing International Investment Patterns: A Revisitation of Lane and Milesi-Ferretti's Evidence

F.M. Pericoli [*], E. Pierucci [†], L. Ventura[‡]

February 7, 2014

Abstract

We show that recent methodological advances in econometric theory raise questions about the results obtained by Lane and Milesi-Ferretti (LMF) in relation to the determinants of international investment patterns (International Investment Patterns, *The Review of Economics and Statistics* 2008; **90**(3): 538–549). We find that LMF's estimated equations are affected by heteroscedasticity (which can lead to inconsistent estimates in log-linearized models), and that the results depend on the pattern of heteroscedasticity assumed and on the estimation method applied. Thus, LMF's findings need to be reassessed. Moreover, we extend the dataset over time to estimate the panel version of the LMF's equations (over years 2001–2009). Our panel allows for the proper accounting of unobserved heterogeneity through country-pair fixed effects and improves the cross-section analysis reconciling empirical evidence with economic theory. Irrespective of the estimation method, we identify a clear diversification motive which drives international equity purchases.

JEL Classifications: F21, G11, G15
Keywords: International investment patterns, Portfolio choices, Gravity models

[*]Department of the Treasury of the Italian Ministry of Economy and Finance.
[†]University of Basilicata.
[‡]Sapienza University of Rome.

1 Introduction

Gravity models have been widely applied to the study of international trade flows since their introduction into economic modeling by Tinbergen (1962). This approach has been recently extended to the study of the international trade of financial assets. The seminal theoretical contribution is due to Martin and Rey (2004), while a first empirical application can be found in Portes and Rey (2005) on a dataset of bilateral flows of 14 countries for the 1989–1996 period. On the basis of these studies, the gravitational approach becomes the standard empirical tool to investigate the main drivers of the international trade of financial assets.[1] Among the most influential contributions, Lane and Milesi-Ferretti (2008) (LMF) first apply a gravity model to a dataset of a very large number of source and host countries. Using the first wave (2001) of the Coordinated Portfolio Investment Survey (CPIS),[2] they estimate a gravity equation for the international trade of equity assets for a sample of 68 source countries and 218 host countries.[3] This work is recognized as a point of reference within this strand of literature.
In this contribution, we replicate the paper by LMF in the light of recent methodological advances in econometric theory that show that ordinary least-squares (OLS) estimates of log-linearized gravity equations can lead to inconsistent estimates. We find that if heteroscedasticity is taken into account, the magnitude and significance of the determinants of portfolio choices detected by LMF need to be drastically reassessed. While Manning and Mullahy (2001) stress the fact that it is appropriate to investigate the pattern of heteroscedasticity to identify the optimal Pseudo Maximum Likelihood (PML) estimator, Santos Silva and Tenreyro (2006) point out that the Poisson Pseudo Maximum Likelihood (PPML) estimator can be the most natural choice when there is no further information (or a certain degree of uncertainty exists) about the pattern of heteroscedasticity (since the PPML performs better than all of the other estimators normally applied). However, a growing number of contributions has recently highlighted that the performance of different estimators applied to gravity equations can be significantly affected by the specific characteristics of the dataset, and it is therefore preferable to effect a model selection in any application.[4] Accordingly, we perform a model selection on the LMF's dataset re-estimating the same equations as in LMF using several estimators that assume different

[1]See, amongst others, Aviat and Coeurdacier (2007), Mishra (2007), Coeurdacier and Guibaud (2011), Balli et al. (2011).

[2]The International Monetary Fund has conducted the survey yearly since 2001.

[3]For the purpose of clarification, the cross-section analysis hereinafter corresponds to what LMF call "panel OLS regressions" given that their dataset refers to just one year (2001).

[4]"[R]esults may, of course, vary across data sets, and more research is definitely needed on this topic", Burger et al. (2009), pag. 182. "The results of the empirical estimations, using three different samples containing real data indicate that the choice of estimator has to be made for each specific dataset. In general it is highly recommended to follow a model selection approach using a number of tests to select the more appropriate estimator for any application", Martinez-Zarzoso (2013), pag. 312. "Every method has advantages and disadvantages and it cannot be asserted that any one of them absolutely outperforms the others. For that reason, it has become a frequent practice in the literature to include several estimation methods for the same database", Gomez-Herrera (2013), pag. 1095.

patterns of heteroscedasticity. More precisely, in order to choose the best suitable estimator, once the presence of heteroscedasticity was detected by means of a simple Breusch and Pagan (1979) test, we investigated the pattern of heteroscedasticity and the adequacy of the model specification through the use of a set of tests employed in the aforementioned reference literature.[5] As a result of this model selection, we end up in favour of the PPML estimator proposed by Santos Silva and Tenreyro (2006) obtaining results that are remarkably different from those obtained by LMF. As a further step, we construct the panel counterpart (over years 2001–2009) of the dataset used in LMF in order to conduct a panel analysis which allows for the introduction of proper fixed effects capturing country-pair unobserved heterogeneity.[6] Given this peculiarity, the panel model is able to identify clearly a significant diversification motive which drives international portfolio choices. This result is very different than the LMF's findings (see Section 3 for details) and provides a contribution to an open issue in the literature about international portfolio allocation, especially considering there is still limited and controversial evidence about the existence of a diversification motive as driver of international portfolio allocation.
The paper is organized as follows: Section 2 briefly reviews theory regarding the impact of heteroscedasticity on gravity models; Section 3 discusses the results obtained from cross-section estimates; Section 4 addresses the panel extension of the analysis; and Section 5 provides concluding remarks.

2 The impact of heteroscedasticity on gravity models

The traditional stochastic version of the gravity model is the following:

$$y_{ij} = \beta_0 x_i^{\beta_1} x_j^{\beta_2} d_{ij}^{\beta_3} \eta_{ij} \tag{1}$$

where y_{ij} is the bilateral trade flow between countries i and j; x_i and x_j are GDPs of countries i and j, respectively; d_{ij} is the distance between the two countries; and η_{ij} is the error term with $E(\eta_{ij}|x_i, x_j, d_{ij}) = 1$. Although this model can be directly estimated through the Nonlinear Least Squares (NLS) method, estimating the model in log-linearized form is the most common approach due to the presence of heteroscedasticity:

$$\ln y_{ij} = \ln \beta_0 + \beta_1 \ln x_i + \beta_2 \ln x_j + \beta_3 \ln d_{ij} + \ln \eta_{ij} \tag{2}$$

However, Santos Silva and Tenreyro (2006) stress that this approach has two main drawbacks: first it is not possible to include zero trade observations (as the logarithm of zero is not defined); and second and more importantly, the presence of heteroscedasticity in η_{ij} implies that $E(\ln \eta_{ij}|x_i, x_j, d_{ij})$ is a function of the covariates, violating the required conditions to obtain consistent estimates.

[5]Following the reference literature, we effected the modified Park's test (Park (1966); Manning and Mullahy (2001)), the Gauss-Newton Regression (GNR) test and the Ramsey RESET test.
[6]See section 3 for details on the panel model.

Equation (2) can be rewritten in more general terms as follows:

$$\ln y_p = \boldsymbol{\beta}\mathbf{X}_p + \epsilon_p \tag{3}$$

where, in order to simplify the notation, p indicates the generic ordered pair (i,j); $\mathbf{X}_p$ is a vector of the logged value of the covariates including also bilateral variables; and ϵ_p is a residual term. In the presence of heteroscedasticity, it is necessary to find consistent alternative estimators such as, for instance, the PPML estimator proposed by Santos Silva and Tenreyro (2006). Manning and Mullahy (2001) highlight that even though Pseudo Maximum Likelihood (PML) estimators are consistent in the case of a misspecified variance function, it is preferable to investigate the pattern of the conditional variance:

$$V[y_p|\mathbf{X}_p] = E[y_p|\mathbf{X}_p]^\lambda \tag{4}$$

in order to choose the appropriate estimator through Park-type regression tests (Park (1966)). Three main cases can be identified from equation (4):[7]

Case 1. $V[y_p|\mathbf{X}_p] = 1$ if $\lambda = 0$
Case 2. $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)$ if $\lambda = 1$
Case 3. $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)^2$ if $\lambda = 2$

In case 1 the NLS[8] assumptions hold, while in case 2, the PPML assumptions are met; case 3 satisfies conditions for employing OLS, Feasible Generalized Least Squares (FGLS) and Gamma Pseudo Maximum Likelihood (GPML). The GPML is the optimal PML estimator for case 3 (Santos Silva and Tenreyro (2006)).

3 Reassessing evidence from LMF's model

The first step of our work was to replicate the entire LMF analysis.[9] In accordance with LMF, the model for the cross-section regressions is:

$$\ln(y_p) = \phi_i + \phi_j + \boldsymbol{\beta}\mathbf{X}_p + \epsilon_p \tag{5}$$

where y_p is the portfolio equity holdings of country i in country j; and ϕ_i and ϕ_j are dummy variables for the source and host countries, respectively. This model includes a dummy variable for each source and each host country, so that the constant term is the sum of ϕ_i

[7] A fourth case, which implies $\lambda = 2$ as in case 3, is given by: $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p) + exp(d_p)\mu(\boldsymbol{\beta}\mathbf{X}_p)^2$, where d_p is a binary dummy variable. For details, see Santos Silva and Tenreyro (2006), page 647.

[8] Given the high number of dummies included in the equations (both cross-section and panel), it was not possible to apply this estimator. However, this model can be very inefficient as it ignores heteroscedasticity and assigns greater weight to noisier observations. Thus, the estimated parameters depend on a small number of observations (see Santos Silva and Tenreyro (2006) pages 644–645 for a discussion on this point).

[9] Complete results and codes available upon request.

and ϕ_j, capturing the individual heterogeneity of countries i and j. This approach allows for exploiting the bilateral dimension of the data in order to take into account national characteristics.

From the "strict sense" replication exercise, we selected the most comprehensive specifications for the full sample of countries (respectively in Columns 3 and 4 of Table 5 in LMF for OLS and instrumental variable estimates) and tested the hypothesis of homoscedasticity. Once the presence of heteroscedasticity was detected (see the Breusch and Pagan (1979) test reported in Table 1), we re-estimated the aforementioned model with different estimators following the recent literature regarding the impact of heteroscedasticity on gravity models (see Section 2). More precisely, we investigated the pattern of heteroscedasticity and the model specification to select the best suitable estimator to be applied to the dataset used by LMF. Given our results, we favour the use of the PPML estimator in line with Santos Silva and Tenreyro (2006) and we have been prompted to reassess LMF's empirical findings. We thus estimated Equation (1) (Columns 1–6, Table 1) by OLS (the exact replication of LMF's model), FGLS, PPML[10] and GPML estimators. We focused on those estimators which performed well in the simulations by Santos Silva and Tenreyro (2006) and Martinez-Zarzoso (2013) and disregarded the estimators with poor performance, such as the truncated OLS, Tobit, and OLS(y+1). Moreover, the Tobit model is not easy to justify when, as in this case, the bounded variable is the result of individual choices and is not a consequence of natural censoring (Ramalho et al. (2011) and Maddala (1991)).

In the case of the PPML and GPML estimators, we considered two samples: Sample *a*) we eliminated those cases for which the dependent variable had zero values in order to estimate the model using the same sample as the OLS and FGLS, and make them completely comparable; and Sample *b*) we included zero-value observations to evaluate the impact of their inclusion on the estimated coefficients (since the PPML and GPML models can deal with the existence of dependent variables with zero values). In Columns 7–12, Table 1, some regressors are assumed endogenous and the Instrumental Variables (IV)[11] method is applied (Column 7 replicates LMF's IV estimates). We are not able to accept any of the null hypotheses ($\lambda = 0, 1, 2$) on the pattern of heteroscedasticity from the modified Park-tests.[12] Thus, it is not possible to give a preference to any of the patterns of heteroscedasticy assumed by the OLS, FGLS, PPML and GPML. However, what we do know from Santos Silva and Tenreyro (2006) is that when there is uncertainty about the pattern of heteroscedasticity, the most reliable estimates are those obtained from the PPML inasmuch as it gives the same weight to all observations, thus avoiding the overweighting of the

[10]See Santos Silva and Tenreyro (2006) for details. Pericoli et al. (2013) recently investigated the diversification motive in international investment patterns estimating a gravity model where the dependent variable is represented by equity shares and the gravity equation is estimated through a PML estimator for fractional data due to Papke and Wooldridge (1996) and Papke and Wooldridge (2008), which follows, in the case of fractional data, the same strategy proposed by Santos Silva and Tenreyro (2006).

[11]To the best of our knowledge, this is the first contribution where the IV PPML and IV GPML estimators are applied to a gravity equation. For details on endogenous regressors and instruments, see appendix A and table notes.

[12]Santos Silva and Tenreyro (2006) highlighted the weaknesses of this test, thus we conducted other tests to check our model selection.

noisier observations (as it is in the case of OLS, FGLS and GPML which assume the conditional variance to be a quadratic function of the conditional mean). Indeed, because trade data for larger countries are usually of better quality, the models which assume that the conditional variance is a function of the conditional mean of an order higher than one (OLS, FGLS and GPML) might actually overweight observations more affected by measurement errors. Even though the modified Park-tests do not accept the null of proportionality of the conditional variance to the conditional mean (which is the pattern of heteroscedasticity assumed by the PPML), we can rely on two more tests which suggest the adoption of the PPML. Indeed, the GNR test, as shown by Santos Silva and Tenreyro (2006), is a valid tool for verifying the adequacy of the PPML estimator and its result (see Table 1) is in favour of the hypothesis of proportionality between the conditional variance and the conditional mean ($\lambda = 1$). The result of the RESET test (see Table 1) also confirms the validity of the PPML estimates.

Comparing the PPML estimates (both in the IV and non-IV version) with all other estimates, it can be clearly seen how the magnitude of the coefficient attached to trade is the lowest and in the IV case, remarkably (about one-third) lower than that estimated by LMF. This means that the impact of trade on equity holdings is considerably reduced. Moreover, the PPML estimates are totally unaffected by the inclusion of zero values as there is no change in the magnitude of coefficients when zero values are included (compare Columns 5 and 6 and Columns 11 and 12 in Table 1). In addition to noting the reduction of the weight of trade in shaping equity holding choices, it is also important to highlight that the existence of a currency union becomes an important determinant, while a common language between source and host countries remains significant, but its magnitude is reduced. Surprisingly, the diversification-motive variables[13] are not statistically significant. To sum up, looking at these results, the most relevant factors shaping portfolio choices are: trade links, financial market integration implied by the existence of a currency union, and a common language. These results differ from LMF's in several ways. For example, in the LMF analyses (see Columns 1 and 7 in Table 1), bilateral trade has a more important role and two out of three diversification variables are sizable in magnitude and statistically significant but with the "wrong" positive sign; such a result is at odds with economic theory since it implies that agents tend to concentrate the purchase of equities in countries whose business cycles and stock-market returns are highly correlated with those of the agents' home country. This would imply irrational behavior by the agents, who would not be willing to exploit the opportunity of diversifying the risk associated with their portfolios.

4 A panel extension

One of the innovative aspects of our work is the extension of the LMF's dataset to the panel case. Using the same sample of source and host countries as in LMF, we constructed

[13] Correlation in GDP growth rates, correlation in stock-market returns and correlation between growth and stock-market returns.

all variables and instruments employed in our reference work for nine years (from 2001 to 2009). Thus, our panel dataset is the exact counterpart of the LMF cross-section dataset. While some other works have used CPIS data to conduct panel estimation analyses,[14] none of them is an exact extension of the LMF case since a different number of source and/or host countries is considered. Moreover, these works only adopt the "double fixed effects" structure used by LMF in the cross-section[15] case without accounting for country-pair unobserved heterogeneity. The time dimension can be exploited to include standard individual fixed effects, i.e. source-host pair dummies rather than a dummy variable for each source and each host country as in the standard practice. The inclusion of "country-pairs fixed effects" is a novelty introduced in our panel extension of the LMF dataset and it allows for capturing the unobserved heterogeneity characterizing any bilateral portfolio equity allocation. This is more general than the "double fixed effects" model (employed in LMF, see Equation 5 in the previous section), which constraints each country's fixed effect to be identical, regardless of the partner country. In terms of the number of dummy variables, a total of $i + j$ individual dummies is estimated in the more restrictive model (LMF), whereas our panel model includes $i \cdot j$ individual fixed effects.
Therefore, for the panel analysis, we adopt the following true fixed effects model specification:

$$\ln(y_{pt}) = \phi_p + \nu_t + \boldsymbol{\beta}\mathbf{X}_{pt} + \epsilon_{pt} \tag{6}$$

where ϕ_p is the intercept of the generic pair of countries (i,j) and ν_t are time fixed effects. As in the cross-section case, we estimated all possible models assuming the exogeneity (Columns 1–6 in Table 2) and the endogeneity (Columns from 7 to 12 in Table 2) of the regressors.[16] One of the most interesting results from the panel case regards the diversification motive. Indeed, in all the estimated models (IV and non-IV) at least one diversification-motive variable is highly statistically significant with a sizable magnitude, and above all, with the "right" negative sign (precisely correlation in stock-market returns).[17] Thus, accounting for country-pair unobserved heterogeneity allows for recognizing the significance of the diversification motive in shaping portfolio choices. Considering the PPML estimator as the best choice according to the results on the pattern of the conditional variance obtained in the cross-section analysis (estimates in Columns 11 and 12), two main conclusions can be drawn: i) trade flows and the diversification motive play a major role; ii) a "tax motive" exists since tax agreements between source and host countries can favour the purchase of equity assets. The result regarding the diversification motive is deemed to be an important contribution since it is still an open question and there are only a limited number of contributions that have found weak evidence of a diversification motive in the allocation of international investments.[18]

[14]Such as Aviat and Coeurdacier (2007).

[15]See equation 5.

[16]In the panel FE case all time-invariant variables considered by LMF in the cross-section case are automatically excluded (precisely, time difference, log of distance, colony dummy, currency union dummy and common legal origin).

[17]Coherently with the economic theory of portfolio diversification.

[18]See Pericoli et al. (2013) for a brief review of the most relevant works.

5 Concluding remarks

The replication of the LMF's influential contribution indicates that its estimated log linearized gravity equations suffer from heteroscedasticity that might bias the coefficients. Therefore, after effecting an estimator selection, we re-estimated the LMF gravity equation by applying the PPML estimator proposed by Santos Silva and Tenreyro (2006) and obtained significantly different results. More specifically: the role of trade is remarkably reduced; the existence of a currency union (which can be seen as a proxy of complete financial integration) plays a significant role; and, the perverse effect of the diversification variables vanishes. In further expanding the analysis, we proposed the panel extension of the LMF's dataset to correctly model for country-pair unobserved heterogeneity, applying panel fixed effects estimators. The panel analysis confirms that the choice of the PPML estimator has appreciable consequences on the magnitude of the estimated coefficients. The most relevant finding is the identification of the significant role that diversification strategies play with respect to the international trade of equity assets. This result, which is very different with respect to LMF's findings, reconciles empirical evidence with economic theory, and provides a contribution to a still heavily debated issue in economic literature.

Appendix A - The Dataset

For a detailed description of variables and instruments used in the cross-section analysis (which refers to year 2001), see appendix B in LMF. The sample of source and host countries considered is the same as in LMF for both the cross-section and panel analyses. For the complete list of countries refer to appendix A in LMF. Variables constructed for the panel extension of the LMF gravity equation are the following:[19]

Bilateral portfolio equity holdings: millions of U.S. dollars of portfolio equity holdings issued by host countries and held by source country. Source: 2001-2009 Coordinated Portfolio Investment Survey.

Bilateral trade: five-year backward looking moving average of imports plus exports over the period 2001-2009. Source: United Nations Commodity Trade Statistics Database.

Tax treaty: dummy variable taking the value of 1 if source and host countries enacted a double-taxation agreement prior to 1999. Agreements considered are: Capital, Income

[19]In the panel extension of the analysis (Table 2), the instruments used by LMF were not enough in number to identify the model (given that just three out of six were time variant and could be used in a panel FE model). We attempted to enlarge this set of instruments, however the resulting set failed to pass the Hansen test (Hansen (1982)). Thus we selected alternative instruments which could identify the model and pass the overidentifying restrictions test. These instruments are one- and two-period lagged values of the following variables: correlation in GDP growth rates; correlation in idiosyncratic GDP growth rates; logged bilateral trade. Correlation between growth and stock-market returns and correlation in stock-market returns are considered only at lag one.

and Capital, Income and Inheritance. Double-taxation agreements on Air, Land and Sea Transport have been excluded. Source: Authors' calculations on the DTT (Double Taxation Treaties) database from www.unctad.org.

Correlation in stock-market returns: eleven-year backward looking moving average correlation between the monthly stock-market returns of the host and source country, expressed in U.S. dollars over the period 2001-2009. Source: authors' calculations based on returns data from Morgan Stanley Capital International (Datastream).

Correlation between growth and stock-market returns: 21 one-year backward looking moving average correlation between annual GDP growth rates in the source country and real stock returns in the host country over the period 2001-2009. Source: authors' calculation based on Morgan Stanley Capital International (Datastream) and World Bank (on-line database World Development Indicators).

Correlation in GDP growth rates: 21-year backward-looking moving average correlation between the annual GDP growth rate of source and host countries over the period 2001-2009. Source: authors' calculation based on World Bank (on-line database World Development Indicators).

Correlation in idiosyncratic GDP growth rates: 21-year backward-looking moving average correlation between the annual idiosyncratic GDP growth rate of source and host countries over the period 2001-2009. As in Pierucci and Ventura (2010), the idiosyncratic component of GDP growth is computed as the estimated residuals of the following regression $\Delta \log(GDP_{it}) = \beta \Delta \log(GDP_{at}) + \epsilon_{it}$, where $\Delta \log(GDP_{it})$ is the country i's GDP growth rate and $\Delta \log(GDP_{at})$ represents the average growth rate. The GDP growth rate of a given country is therefore decomposed in two orthogonal components. Indeed, $\Delta \log(GDP_{it}) = \widehat{\beta} \Delta \log(GDP_{at}) + e_{it}$, thus the idiosyncratic GDP growth will be orthogonal to the aggregate (group average) GDP growth by construction, $e_{it} \perp \widehat{\beta} \Delta \log(GDP_{at})$. The more standard practice (e.g. Asdrubali et al. (1996)) consists of simply subtracting the group average GDP growth from each country's GDP growth rate. However, this practice does not guarantee orthogonality between aggregate and idiosyncratic GDP growth and it may generate a serious omitted variable bias if one of the regressors strongly correlates with the aggregate GDP growth. Moreover the standard decomposition restricts the coefficient attached to aggregate GDP to 1, which is often at odds with the empirical evidence.

References

Asdrubali, P., B. E. Sørensen, and O. Yosha (1996). Channels of interstate risk sharing: United States 1963-1990. *The Quarterly Journal of Economics 111*(4), 1081–1110.

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

Balli, F., R. Louis, and M. Osman (2011). The patterns of cross-border portfolio investments in the GCC region: Do institutional quality and the number of expatriates play a role? *Journal of Economics and Finance 35*(4), 434–455.

Breusch, T. S. and A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica 47*(5), 1287–94.

Burger, M., F. van Oort, and G.-J. Linders (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis 4*(2), 167–190.

Coeurdacier, N. and S. Guibaud (2011). International portfolio diversification is better than you think. *Journal of International Money and Finance 30*(2), 289–308.

Gomez-Herrera, E. (2013). Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics 44*(3), 1087–1111.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–54.

Lane, P. R. and G. M. Milesi-Ferretti (2008). International investment patterns. *The Review of Economics and Statistics 90*(3), 538–549.

Maddala, G. (1991). A perspective on the use of limited-dependent and qualitative variables models in accounting research. *The Accounting Review 66*(4), 788–807.

Manning, W. G. and J. Mullahy (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics 20*, 461–494.

Martin, P. and H. Rey (2004). Financial super-markets: Size matters for asset trade. *Journal of International Economics 64*(2), 335 – 361.

Martinez-Zarzoso, I. (2013). The log of gravity revisited. *Applied Economics 45*(3), 311–327.

Mishra, A. V. (2007). International investment patterns: Evidence using a new dataset. *Research in International Business and Finance 21*(2), 342 – 360.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics 11*(6), 619–32.

Papke, L. E. and J. M. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*(1-2), 121–133.

Park, R. E. (1966). Estimation. *Econometrica 34*(4), 888.

Pericoli, F., E. Pierucci, and L. Ventura (2013). Cross-border equity portfolio choices and the diversification motive: A fractional regression approach. *Economics Letters 121*(2), 282 – 286.

Pierucci, E. and L. Ventura (2010). Risk sharing: A long run issue? *Open Economies Review 21*(5), 705–730.

Portes, R. and H. Rey (2005). The determinants of cross-border equity flows. *Journal of International Economics 65*(2), 269–296.

Ramalho, E. A., J. J. Ramalho, and J. M. Murteira (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys 25*(1), 19–68.

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

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy.* The Twentieth Century Fund, New York.

Table 1: Bilateral portfolio equity holdings, cross section estimates, year 2001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	†$\lambda = 2$				$\lambda = 1$		$\lambda = 2$				$\lambda = 1$	
VARIABLES	OLS‡ $\ln(y_p)$	FGLS $\ln(y_p)$	GPML $y_p > 0$	GPML $y_p \geq 0$	PPML $y_p > 0$	PPML $y_p \geq 0$	IV OLS‡ $\ln(y_p)$	IV FGLS $\ln(y_p)$	IV GPML $y_p > 0$	IV GPML $y_p \geq 0$	IV PPML $y_p > 0$	IV PPML $y_p \geq 0$
Log bilateral trade	0.42***	0.95***	0.49***	0.60***	0.35***	0.35***	0.63***	0.83***	0.50***	0.59***	0.25***	0.25***
	(0.09)	(0.04)	(0.09)	(0.11)	(0.08)	(0.08)	(0.11)	(0.07)	(0.09)	(0.11)	(0.07)	(0.07)
Log distance	−0.02	−0.11***	−0.01	0.18	−0.02	−0.02						
	(0.13)	(0.04)	(0.14)	(0.16)	(0.10)	(0.10)						
Time difference	−0.09***	0.16***	−0.04	−0.10***	0.02	0.02						
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)						
Common language	0.46***	0.12	0.24*	0.53***	0.25**	0.26**	0.34**	0.03	0.11	0.28*	0.25**	0.25**
	(0.15)	(0.11)	(0.12)	(0.15)	(0.11)	(0.11)	(0.16)	(0.13)	(0.13)	(0.15)	(0.12)	(0.12)
Colony dummy	0.34	0.12	0.54**	0.59**	−0.33**	−0.33**	0.25	0.27	0.46*	0.52*	−0.22	−0.22
	(0.25)	(0.20)	(0.25)	(0.27)	(0.15)	(0.15)	(0.25)	(0.22)	(0.24)	(0.27)	(0.17)	(0.17)
Tax treaty	−0.03	0.36***	−0.05	0.26*	−0.42**	−0.42**	−0.08	0.15	−0.14	0.01	−0.43**	−0.43**
	(0.14)	(0.11)	(0.13)	(0.15)	(0.18)	(0.18)	(0.15)	(0.13)	(0.13)	(0.14)	(0.18)	(0.18)
Currency union dummy	0.05	−0.36***	0.04	−0.26	0.59***	0.59***	0.22	−0.30*	0.17	−0.26	0.60***	0.60***
	(0.21)	(0.13)	(0.16)	(0.18)	(0.14)	(0.14)	(0.26)	(0.16)	(0.20)	(0.24)	(0.20)	(0.20)
Correl. growth rates	0.42**	1.42***	0.64***	0.77***	0.17	0.17	0.62*	0.70***	0.61**	0.52	0.43	0.43
	(0.20)	(0.15)	(0.19)	(0.21)	(0.21)	(0.21)	(0.33)	(0.22)	(0.30)	(0.33)	(0.34)	(0.34)
Correl. in stock returns	2.09***	2.62***	2.26***	1.09*	0.27	0.26	2.47**	3.34***	3.39***	3.27***	0.48	0.48
	(0.56)	(0.38)	(0.53)	(0.64)	(0.51)	(0.51)	(1.03)	(0.85)	(0.95)	(1.10)	(0.73)	(0.73)
Correl. growth-stock ret.	0.17	−0.10	0.26	0.09	−0.18	−0.19	0.51	−0.41	0.36	−1.15	0.49	0.48
	(0.20)	(0.16)	(0.19)	(0.22)	(0.21)	(0.21)	(0.69)	(0.45)	(0.70)	(0.80)	(0.84)	(0.84)
Common legal origin	0.20*	−0.03	0.15	0.15	−0.09	−0.09	0.01	−0.04	0.02	−0.02	−0.10	−0.10
	(0.11)	(0.09)	(0.10)	(0.12)	(0.09)	(0.09)	(0.12)	(0.10)	(0.10)	(0.11)	(0.09)	(0.09)
Constant	8.67***	−4.59***	7.34***	5.51**	9.50***	9.47***	4.68***	2.28	5.76***	5.14***	10.42***	10.41***
	(2.11)	(1.36)	(2.23)	(2.62)	(1.72)	(1.72)	(1.09)	(1.83)	(0.84)	(1.00)	(0.78)	(0.78)
Observations	1,011	1,011	1,011	1,127	1,011	1,127	854	854	854	925	854	925
R-squared	0.878	0.965	–	–	–	–	0.888	0.962	–	–	–	–
Breush-Pagan Test (H0: homoscedasticity)	[0.000]											
Park Test												
$\hat{\lambda}$	1.86		1.87	1.84	1.57	1.61						
	(0.03)		(0.22)	(0.15)	(0.03)	(0.21)						
Conf. Int.	1.82-1.90		1.83-1.91	1.81-1.87	1.51-1.63	1.56-1.66						
H0: $\lambda = 0$	[0.000]		[0.000]	[0.000]	[0.000]	[0.000]						
H0: $\lambda = 1$	[0.000]		[0.000]	[0.000]	[0.000]	[0.000]						
H0: $\lambda = 2$	[0.000]		[0.000]	[0.000]	[0.000]	[0.000]						
GNR test (Ho: $V[y_p \vert \mathbf{X}_p] \propto E[y_p \vert \mathbf{X}_p]$)					[0.952]	[0.942]						
RESET test	[0.000]	[0.005]	[0.000]	[0.000]	[0.843]	[0.831]						

*** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in brackets and p-values in squared brackets. Estimated equation: $\ln(y_p) = \phi_i + \phi_j + \beta \mathbf{X_p} + \epsilon_p$. Instruments used in the IV estimations are the same variables as in LMF: log of distance; time difference; existence of a common border; correlation between the annual GDP growth rates in the source and host country over the period 1980-1989; correlation between annual GDP growth in the source country and real stock-market returns in the host country over the period 1980-1989; correlation between the monthly stock-market returns of the host and source country, expressed in U.S. dollars, over the period January 1990-December 1994. See appendix see appendix B in LMF for further details.

† $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)$ if $\lambda = 1$; $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)^2$ if $\lambda = 2$.

‡ Columns (1) and (7) are exact replication of the estimated models reported in table 5 in LMF respectively in columns 3 and 4. All other results obtained in LMF have been replicated and are available from the authors along with Stata codes.

Table 2: Bilateral portfolio equity holdings, panel estimates, years 2001-2009

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	†$\lambda = 2$				$\lambda = 1$		$\lambda = 2$				$\lambda = 1$	
VARIABLES	FE OLS‡ $\ln(y_{pt})$	FE FGLS $\ln(y_{pt})$	FE GPML $y_{pt} > 0$	FE GPML $y_{pt} \geq 0$	FE PPML $y_{pt} > 0$	FE PPML $y_{pt} \geq 0$	IV FE‡ $\ln(y_{pt})$	IV FE FGLS $\ln(y_{pt})$	IV FE GPML $y_{pt} > 0$	IV FE GPML $y_{pt} \geq 0$	IV FE PPML $y_{pt} > 0$	IV FE PPML $y_{pt} \geq 0$
Log bilateral trade	0.37***	0.23***	0.34***	0.39***	0.63***	0.63***	0.96***	0.61***	1.00***	1.11***	1.55***	1.55***
	(0.06)	(0.04)	(0.03)	(0.04)	(0.08)	(0.08)	(0.26)	(0.32)	(0.27)	(0.27)	(0.43)	(0.43)
Tax treaty	0.03	0.00	0.04	−0.03	0.14*	0.14*	0.04	−0.04	0.12	0.01	0.08*	0.08*
	(0.11)	(0.08)	(0.06)	(0.07)	(0.08)	(0.08)	(0.12)	(0.47)	(0.09)	(0.09)	(0.04)	(0.04)
Correl. growth rates	−0.01	−0.13	0.03	−0.09	0.29*	0.29*	−0.13	−0.32	−0.04	−0.06	0.35	0.35
	(0.18)	(0.11)	(0.09)	(0.10)	(0.17)	(0.17)	(0.25)	(0.22)	(0.14)	(0.17)	(0.27)	(0.27)
Correl. in stock returns	−0.85**	−2.81***	−0.46**	−0.79***	−1.45***	−1.45***	−3.33***	−5.57***	−4.90***	−5.39***	−4.94***	−4.94***
	(0.34)	(0.18)	(0.18)	(0.18)	(0.39)	(0.39)	(0.65)	(0.65)	(0.56)	(0.65)	(0.94)	(0.94)
Correl. growth-stock ret.	0.05	1.22***	−0.03	−0.24***	0.06	0.05	−0.08	0.42**	0.17	−0.16	−0.42	−0.42
	(0.13)	(0.08)	(0.08)	(0.06)	(0.18)	(0.18)	(0.39)	(0.47)	(0.25)	(0.30)	(0.41)	(0.41)
Constant	−5.60***	−3.76***					−5.79***	−2.42***				
	(0.25)	(0.08)					(0.80)	(0.56)				
Observations	12,458	12,458	12,458	15,537	12,458	15,537	8,114	8,114	8,114	9,609	8,114	9,609
R-squared within	0.10	0.14	–	–	–	–	0.06	0.10	–	–	–	–
Hansen J Test							[0.218]					

*** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in brackets and p-values in squared brackets. Estimated model: $\ln(y_{pt}) = \phi_p + \nu_t + \beta \mathbf{X}_{pt} + \epsilon_{pt}$. Instruments used in the IV estimations are one and two periods lagged values of the following variables: correlation in GDP growth rates; correlation in idiosyncratic GDP growth rates and log bilateral trade and one period lagged values of correlation between growth and stock-market returns and correlation in stock-market returns. The type of instruments used by LMF were not enough to identify the model in the panel case (given that just three out of six were time variant and could be used in a panel FE model). We attempted to enlarge this set of instrument, however the resulting set of instruments failed to pass the Hansen test (Hansen (1982)), thus we selected alternative instruments which could identify the model and pass the overidentifying restrictions test.

† $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)$ if $\lambda = 1$; $V[y_p|\mathbf{X}_p] = \mu(\boldsymbol{\beta}\mathbf{X}_p)^2$ if $\lambda = 2$.

‡ Columns (1) and (7) are extension to the panel case over the period 2001-2009 of the estimated models reported respectively in columns 3 and 4 of table 5 in LMF for the year 2001.