Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically*

J.M.C. Santos Silva† Silvana Tenreyro‡

March 6, 2009

Abstract

Helpman, Melitz, and Rubinstein (2008)—HMR—present a rich theoretical model to study the determinants of bilateral trade flows across countries. The model is then empirically implemented through a two-stage estimation procedure. This note seeks to clarify some econometric aspects of the estimation approach used by HMR and explore the consequences of possible departures from the maintained distributional assumptions.

JEL Classification codes: C13; C50; F10.

Key words: Gravity equation, Heteroskedasticity, Jensen's inequality.

---

*We are grateful to Elhanan Helpman, Marc Melitz, and Yona Rubinstein for generously sharing their data. We thank especially Marc Melitz for useful conversations. Santos Silva also gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

†University of Essex and CEMAPRE. Wivenhoe Park, Colchester CO1 1ED, United Kingdom. Fax: +44 (0)1206 872769. E-mail: jmcss@essex.ac.uk.

‡London School of Economics, CEP, and CEPR. Department of Economics, s.600. St. Clement's Building. Houghton St., London WC2A 2AE, United Kingdom. Fax: +44 (0)20 78311840. E-mail: s.tenreyro@lse.ac.uk.

1. Introduction

In a highly insightful and stimulating paper, Helpman, Melitz and Rubinstein (2008), hereinafter HMR, present a theoretical framework to study bilateral trade flows across countries. The model has three appealing features that make it suitable to describe empirical patterns of bilateral trade flows. First, the model can yield asymmetric trade flows between country pairs. Second, it can generate zero trade flows between some countries, as well as zero exports from one country, say j, to a second country i, together with positive exports from country i to country j. Third, it generates a gravity equation for positive trade flows. The model has therefore the potential to explain three prevalent regularities in trade data: The asymmetry in bilateral trade flows between country pairs; the high prevalence of zeroes (in either one or both directions of bilateral trade flows); and the remarkably good fit of the gravity equation.

HMR use their conceptual framework to develop a two-stage estimation procedure that generalizes the empirical gravity equation by taking into account the extensive margin (the decision to export from j to i), and the intensive margin (the volume of exports from j to i, conditional on exporting). The first stage consists of a probit regression that models the probability that country j exports to country i. The second stage is a gravity equation estimated in logarithmic form. This two-stage procedure aims at correcting for two potential problems present in estimations of the gravity equation: The first is a standard selection bias resulting from the need to drop the observations with zero trade when estimating logged gravity models. The second is a bias due to the potential unobserved firm level heterogeneity resulting from an omitted variable that measures the impact of the number of exporting firms (the extensive margin).

Though HMR's model makes a significant step towards a better understanding of the determinants of bilateral trade flows, the proposed two-stage non-linear least squares estimation procedure has some limitations. In this paper, we seek to clarify and make progress on the two following issues.

First, the approach used by HMR to deal with the selectivity bias is only approximately correct and, consequently, the proposed estimator is not generally consistent for the parameters of interest. Although we argue that the approximation proposed by HMR is likely to be reasonably accurate in many applications, we present an alternative procedure to deal with the selectivity bias which, under the assumptions maintained by HMR, leads to a consistent estimator.

Second, HMR obtain their model under very strong distributional assumptions. Although the authors explore the consequences of relaxing some of them, all the results presented in their paper depend critically on the untested assumption that all random components of the model are homoskedastic. We explore the consequences of possible departures from this assumption and provide empirical evidence on its importance. We conclude that, given the available econometric technology, the presence of heteroskedasticity in trade data precludes the estimation of models that purport to separately identify the effects of the covariates in the intensive and extensive margins.

The remainder of the paper is organized as follows. The next section reconsiders the HMR model, focusing particularly on the sample-selection correction and on the distributional assumptions. Section 3 reappraises the empirical study presented in HMR and, finally, Section 4 concludes the paper.

2. The HMR model

HMR specify a trade equation which can be written as (see equations (6) and (8) in HMR):

$$M_{ij} = B_0 \Lambda_j X_i \tau_{ij}^{1-\varepsilon} \max \left\{ \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1, 0 \right\},$$

where Λ_j denotes a fixed effect for exporter j, X_i is a fixed effect for importer i, a_{ij} is a measure of the productivity of the firms exporting from j to i, a_L is the lower bound for a_{ij}, and k and ε are parameters. Furthermore, the authors assume that

$$\tau_{ij}^{\varepsilon-1} = D_{ij}^{\gamma} \exp\left(-u_{ij}\right),$$

where γ is a parameter, D_{ij} is the distance (and other factors creating trade resistance) between countries i and j, and $u_{ij} \sim \mathcal{N}\left(0, \sigma_u^2\right)$. Therefore,

$$M_{ij} = B_0 \Lambda_j X_i D_{ij}^{-\gamma} \max\left\{\left(\frac{a_{ij}}{a_L}\right)^{k-\varepsilon+1} - 1, 0\right\} \exp\left(u_{ij}\right). \tag{1}$$

Direct estimation of this equation would require information about a_{ij} and a_L, which is typically not available. To overcome this problem, HMR define the latent variable (see equation (10) in HMR)

$$Z_{ij} = \Gamma_0 \Xi_i \Upsilon_j D_{ij}^{-\gamma} \Psi_{ij}^{-\kappa} \exp\left(v_{ij} + u_{ij}\right) = \left(\frac{a_{ij}}{a_L}\right)^{\varepsilon-1},$$

where κ is a parameter, Υ_j denotes a fixed effect for exporter j, Ξ_i is a fixed effect for importer i, Ψ_{ij} denotes additional country-pair specific fixed trade costs, and $v_{ij} \sim \mathcal{N}\left(0, \sigma_v^2\right)$. Furthermore, HMR assume that u_i and v_i are uncorrelated.

The new variable Z_{ij} is not observed. However, positive trade is observed when $Z_{ij} > 1$, which leads HMR to propose the following two-step estimation strategy.

Let T_{ij} be a binary indicator defined as $T_{ij} = 1\left[M_{ij} > 0\right]$. Then, defining $z_{ij} = \ln\left(Z_{ij}\right)$, $\gamma_0 = \ln\left(\Gamma_0\right)$, $\xi_i = \ln\left(\Xi_i\right)$, $y_j = \ln\left(\Upsilon_j\right)$, $d_{ij} = \ln\left(D_{ij}\right)$ and $\psi_{ij} = \ln\left(\Psi_{ij}\right)$, we have

$$\begin{aligned} \Pr\left(T_{ij} = 1\right) &= \Pr\left(z_{ij} > 0\right) \\ &= \Pr\left(\gamma_0 + \xi_i + y_j - \gamma d_{ij} - \kappa \psi_{ij} > -\left(v_{ij} + u_{ij}\right)\right). \end{aligned}$$

Under the maintained assumptions of normality and homoskedasticity, the unknown parameters can be consistently estimated up to scale using a probit. That is, under these assumptions, it is possible to consistently estimate

$$Z_{ij}^{*} = \left(\Gamma_0 \Xi_i \Upsilon_j D_{ij}^{-\gamma} \Psi_{ij}^{-\kappa}\right)^{\frac{1}{\sigma_{u+v}}},$$

where σ_{u+v} denotes the standard deviation of $\left(v_{ij} + u_{ij}\right)$. Using this result, it is possible to rewrite (1) as

$$M_{ij} = T_{ij} B_0 \Lambda_j X_i D_{ij}^{-\gamma} \left\{\left[Z_{ij}^{*} \exp\left(\varsigma_{ij}\right)\right]^{\delta} - 1\right\} \exp\left(u_{ij}\right), \tag{2}$$

where $\varsigma_{ij} = \frac{v_{ij}+u_{ij}}{\sigma_{u+v}}$ and $\delta = \sigma_{u+v} (k - \varepsilon + 1) / (\varepsilon - 1)$.

The second step in the HMR procedure is the estimation of the trade equation for the positive observations of M_{ij}. To do this, the authors take logs of both sides of (2), leading to

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln \left\{ \exp \left[\delta \left(z_{ij}^* + \varsigma_{ij} \right) \right] - 1 \right\} + u_{ij}, \qquad (3)$$

where, as usual, lower-case letters represent the log of the quantity corresponding to the same upper case letter.

2.1. The selectivity correction

Estimation of (3) has to be performed using only observations with positive values of M_{ij}, which originates a sample-selection issue. However, this is not a standard selectivity problem because the equation of interest has two random components, and one of them enters the equation within a non-linear function.

To account for the fact that $E[u_{ij}|M_{ij} > 0] \neq 0$, HMR include in the regression equation the Mills ratio from the first step, which (under normality and homoskedasticity) is proportional to $E[u_{ij}|M_{ij} > 0]$. This is the correct procedure to account for selectivity in an additive error (see, e.g., Heckman, 1979).

Dealing with the effect of the sample-selection on ς_{ij} is less standard. HMR approach this problem in a way that is akin to the ad-hoc method used by Greene (1994) to address the sample selection problem in count data models. In particular, HMR suggest replacing ς_{ij} with its expectation conditional on $M_{ij} > 0$, which is again the Mills ratio from the first step. That is, denoting the Mills ratio by $\eta_{ij} = \phi \left(z_{ij}^* \right) / \Phi \left(z_{ij}^* \right)$, the second step of HMR's procedure is the estimation of (see equation (14) in HMR)

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln \left\{ \exp \left[\delta \left(z_{ij}^* + \eta_{ij} \right) \right] - 1 \right\} + \beta_{u\eta} \eta_{ij} + e_{ij}, \qquad (4)$$

where $\beta_{u\eta}$ is a parameter. Of course, in practice, z_{ij}^* and η_{ij} have to be replaced by $\hat{z}_{ij}^*$ and $\hat{\eta}_{ij}$, their estimates obtained from the first-stage probit regression.

It is easy to see that this approach to correct the effect of the sample selection on ς_{ij} is generally inappropriate (see, e.g., Terza, 1998). Indeed, for any non-linear function $f(\cdot)$, Jensen's inequality implies that replacing ς_{ij} in $f(\varsigma_{ij})$ by its expectation conditional on $M_{ij} > 0$, $f[E(\varsigma_{ij}|M_{ij} > 0)] = f(\eta_{ij})$, does not lead to a consistent estimate of $E[f(\varsigma_{ij})|M_{ij} > 0]$. Therefore, $f(\hat{\eta}_{ij})$ is not a consistent estimator of $E[f(\varsigma_{ij})|M_{ij} > 0]$ and, consequently, the proposed estimation method will generally be inconsistent for all the parameters of interest.

Although estimation of (4) should be viewed as being based on an approximation to the conditional expectation of $\ln\left\{\exp\left[\delta\left(z_{ij}^{*} + \eta_{ij}\right)\right] - 1\right\}$, it is interesting to notice that this approximation is likely to be reasonably accurate in many practical situations. To see this, notice that

$$\ln\left\{\exp\left[\delta\left(z_{ij}^{*} + \varsigma_{ij}\right)\right] - 1\right\} = \ln\left\{\exp\left[\delta\left(z_{ij}^{*} + \phi\left(z_{ij}^{*}\right)/\Phi\left(z_{ij}^{*}\right) + \omega_{ij}\right)\right] - 1\right\}, \qquad (5)$$

where ω_{ij} is just the deviation of ς_{ij} form its conditional mean $\eta_{ij} = \phi\left(z_{ij}^{*}\right)/\Phi\left(z_{ij}^{*}\right)$. The approximation used by HMR consists in ignoring ω_{ij}, which would be innocuous if the function was linear in this random term because, in that case, ω_{ij} would just be added to the error of the equation. However, it is clear that for a wide range of values of z_{ij}^{*} and reasonable values of δ, (5) is approximately linear in ω_{ij}. Therefore, in practice, the approximation used by HMR is likely to be reasonable, especially because positive values of M_{ij} tend to be associated with large values of z_{ij}^{*}, which are the ones for which the approximation is better.

However accurate, under the assumptions maintained by HMR, this approximation is not needed. Indeed, and for $M_{ij} > 0$, it is possible to write

$$\begin{aligned} M_{ij} &= B_0 \Lambda_j X_i D_{ij}^{-\gamma}\left\{\left[Z_{ij}^{*\delta}\exp\left(\delta\varsigma_{ij}\right)\right] - 1\right\}\exp\left(u_{ij}\right) \\ &= B_0 \Lambda_j X_i D_{ij}^{-\gamma} Z_{ij}^{*\delta}\exp\left(\nu_{ij}\right) - B_0 \Lambda_j X_i D_{ij}^{-\gamma}\exp\left(u_{ij}\right), \end{aligned}$$

where $\nu_{ij} = \delta\varsigma_{ij} + u_{ij} = \mathcal{N}\left(0, \delta^2 + \sigma_u^2 + \frac{2\delta\sigma_u^2}{\sigma_{u+v}}\right)$. Then, using the results in van de Ven and van Praag (1981) on the moments of log-normal random variables under incidental truncation,

$$\begin{aligned} E\left[M_{ij}|\Lambda_j, X_i, D_{ij}, Z_{ij}^{*}, T_{ij} = 1\right] = B_0 \Lambda_j X_i D_{ij}^{-\gamma}&\left[Z_{ij}^{*\delta}\exp\left(\frac{\sigma_\nu^2}{2}\right)\Phi\left(z_{ij}^{*} + \rho_{\varsigma\nu}\sigma_\nu\right)\right. \\ &\left.- \exp\left(\frac{\sigma_u^2}{2}\right)\Phi\left(z_{ij}^{*} + \rho_{\varsigma u}\sigma_u\right)\right]\Big/\,\Phi\left(z_{ij}^{*}\right), \end{aligned}$$

where $\rho_{\varsigma\nu}$ and $\rho_{\varsigma u}$ denote the correlations between ς_{ij} and ν_{ij} and u_{ij}, respectively.

Noting that $\sigma_{\nu}^2 = \delta^2 + \sigma_u^2 + \frac{2\delta\sigma_u^2}{\sigma_{u+v}}$, $\rho_{\varsigma\nu} = \left(\delta + \frac{\sigma_u^2}{\sigma_{u+v}}\right)/\sigma_{\nu}$, and $\rho_{\varsigma u} = \sigma_u^2/\left(\sigma_u\sigma_{u+v}\right)$, it is finally possible to obtain

$$E\left[M_{ij}|\Lambda_j, X_i, D_{ij}, Z_{ij}^*, T_{ij} = 1\right] = B_0\Lambda_j X_i D_{ij}^{-\gamma} S\left(z_{ij}^*, \delta, r\right)\exp\left(\sigma_u^2/2\right), \qquad (6)$$

where $S\left(z_{ij}^*, \delta, r\right) = \left[\exp\left(\delta^2/2 + \delta r + \delta z_{ij}^*\right)\Phi\left(z_{ij}^* + \delta + r\right) - \Phi\left(z_{ij}^* + r\right)\right]/\Phi\left(z_{ij}^*\right)$ and $r = \sigma_u^2/\sqrt{\sigma_u^2 + \sigma_v^2}$. Naturally, B_0 and σ_u^2 are not separately identified.[1]

Even if the maintained distributional assumptions are valid, the (multiplicative) errors of the model defined by (6) will have conditional expectation equal to one, but are not independent of the regressors because of the selectivity correction. Therefore, (6) should be estimated as a multiplicative model, for example using a pseudo-maximum likelihood estimator, as suggested by Santos Silva and Tenreyro (2006).[2]

2.2 Distributional assumptions

The main results in HMR are obtained under the assumption that v_{ij} and u_{ij} are independent, homoskedastic and jointly normal. To check the robustness of the results obtained with their two-stages procedure, HMR also estimate less parametric specifications, but all the results presented by HMR are based on the assumption that v_{ij} and u_{ij} are homoskedastic. Heteroskedasticity is often viewed as a minor problem in that, under very general conditions, it does not affect the consistency of the OLS estimator. However, in the model proposed by HMR the situation is very different.

[1] It is interesting to notice that, integrating out T_{ij}, it is possible to obtain

$$E\left[M_{ij}|\Lambda_j, X_i, D_{ij}, Z_{ij}^*\right] = B_0\Lambda_j X_i D_{ij}^{-\gamma} S^*\left(z_{ij}^*, \delta, r\right)\exp\left(\sigma_u^2/2\right),$$

where $S^*\left(z_{ij}^*, \delta, r\right) = \exp\left(\delta^2/2 + \delta r + \delta z_{ij}^*\right)\Phi\left(z_{ij}^* + \delta + r\right) - \Phi\left(z_{ij}^* + r\right)$. Therefore, the parameters of interest can be estimated using either this expression or (6). Here, as in HMR, we focus on the estimation of the second stage using only the observations with positive trade.

[2] Recall that, when the errors of the model are not statistically independent of the regressors, estimation of the logged form of the model does not generally lead to consistent estimates of the parameters of interest because the mean of m_{ij} depends both on the log of the mean of M_{ij} and on its higher order moments. For details, see Santos Silva and Tenreyro (2006).

Because, as pointed out by Santos Silva and Tenreyro (2006), heteroskedasticity is pervasive in trade data, it is important to explore the consequences of departures from this assumption.

A first consequence of the presence of heteroskedasticity is that the selectivity corrections used in (4) and (6) are no longer valid. Estimation of linear sample selection models which are robust to non-normality and heteroskedasticity has been studied by Chen and Khan (2003).[3] However, due to the large number of regressors typically used, implementation of these methods with trade data is far from trivial. In the case of the HMR model, these difficulties are compounded by the fact that ς_{ij} enters the model non-linearly. Therefore, an estimator of the HMR model that is robust to incidental distributional assumptions is not currently available, and may even not be possible to obtain at all.

Another, perhaps more important, consequence of heteroskedasticity in v_{ij} and u_{ij} is that the functional form of (2) directly depends on the homoskedasticity assumption. Indeed, δ is a function of σ_v and σ_u and it is clear that if either of the random components of the model is heteroskedastic, δ will be a function of the regressors, which will then enter the model in a much more complex form. This sensitivity to the homoskedasticity assumption, which is independent of the particular estimation method used, has the potential to introduce severe misspecification in both (4) and (6), making consistent estimation of the parameters of interest not generally possible. Of course, one may be tempted to tackle this problem by specifying σ_v and σ_u as functions of the regressors, but this approach is foiled by the fact that economic theory provides no guidance on the possible heteroskedasticity patterns.

Given the potential impact of heteroskedasticity on the estimation results, it is important to check for its presence and to gauge its potential impact. Since δ is proportional to the standard deviation of the error in the first-stage, the assumption that δ is independent of the regressors can be tested by testing for heteroskedasticity in the probit.

[3] The Chen and Khan (2003) estimator allows for the presence of unspecified heteroskedasticity, but imposes restrictions on the way higher order moments depend on the regressors.

As described in Godfrey (1988), tests for heteroskedasticity in the probit model can be performed as simple tests for omitted variables of the form $\hat{z}^*_{ij} w_{ij}$, where w_{ij} denotes the set of variables suspected of causing heteroskedasticity. Therefore, by analogy with the popular two-degrees-of-freedom special case of White's test for heteroskedasticity (see Wooldridge, 2002, p. 127), one can test for heteroskedasticity in the first-stage probit by checking for the joint significance of the additional regressors $\hat{z}^{*2}_{ij}$ and $\hat{z}^{*3}_{ij}$.

In the specific case of the probit, this particular heteroskedasticity test, which is analogous to a two-degrees-of-freedom RESET test (Ramsey, 1969), is also particularly interesting in that it can be interpreted as a normality test (see Newey, 1985). Therefore, this simple test provides a direct check for the validity of the main distributional assumptions required for consistent estimation of the model of interest.

Because heteroskedasticity also impacts on the functional form of (4) and (6), it is important to check whether the specification of these models is reasonably adequate. Although it is certainly possible to develop more specific tests, a simple way to gauge the adequacy of these models is the following.

In the spirit of Cosslett (1991), HMR partially relax the distributional assumptions used to obtain (4) by estimating models of the form

$$m_{ij} = \lambda_j + \chi_i - \gamma d_{ij} + \sum_{s=1}^{Q} \alpha_s 1 \left[q_{s-1} < z^*_{ij} < q_s \right] + e^*_{ij}, \qquad (7)$$

where $\alpha_1, \ldots, \alpha_Q$ are parameters and $q_0 = -\infty$ and $q_Q = \infty$. A similar generalization of (6) is possible, leading to

$$M_{ij} = \exp \left(\lambda_j + \chi_i - \gamma d_{ij} + \sum_{s=1}^{Q} \alpha_s 1 \left[q_{s-1} < z^*_{ij} < q_s \right] \right) + \zeta_{ij}. \qquad (8)$$

Although these models are more flexible than their fully parametric counterparts, they still assume that $\left\{ \left[Z^*_{ij} \exp\left(\varsigma_{ij}\right) \right]^{\delta} - 1 \right\}$ depends on the regressors only through z^*_{ij}; that is, both (7) and (8) assume that δ is constant. To check for departures from this assumption one can check for the significance of interactions between the indicator variables and functions of the other regressors.

A simple way of doing this is again to perform a RESET-type test for the significance of additional variables constructed as powers of the estimated linear indexes $\lambda_j +$

$\chi_i - \gamma d_{ij} + \sum_{s=1}^{Q} \alpha_s 1 \left[q_{s-1} < \rho_{ij} < q_s \right]$. Although in this case there is no particularly competing reason to check for the significance of squares and cubes of the estimated indexes, by analogy with what is done for the probit, we will also use two-degrees-of-freedom RESET tests to check the validity of (7) and (8).

3. A reappraisal of the HMR study

In this section we reconsider the study presented in HMR, comparing and contrasting the results obtained with different estimators of their theoretical model for bilateral trade flows and exploring the consequences of possible violations of the homoskedasticity assumption. In order to maintain comparability with the results in HMR, we use exactly the same data and the same set of regressors used in the original study, and do not attempt to improve upon the specification of the factors creating trade resistance.[4] HMR provide details on the data and its sources.

3.1 Baseline results

The baseline results presented by HMR use data on trade flows in 1986, for a subsample of countries for which information on regulation costs of firm entry is available. These cost variables provide the exclusion restrictions used to help in the identification of the models that allow the regressors to have different effects on the extensive and intensive margins.

Table 1 displays some of the results presented in HMR, together with new results obtained using different estimators. The estimators considered are as follows: **Probit** is the first stage estimator used for all the estimators that distinguish between the effects on the two margins, the results are identical to those reported in HMR; **OLS** corresponds to the benchmark results reported in HMR; **NLS** is the HMR second stage non-linear least squares estimator based on (4);[5] **A-Bins** is the more general

[4] See Baranga (2008) for a detailed investigation on the quality of the data set used by HMR.

[5] The results presented here differ from the ones reported in the original paper because, unlike HMR, we do not right-censor $\hat{z}^*_{ij}$ at $\Phi^{-1}(0.9999999) \approx 5.2$ (see Helpman, Melitz and Rubinstein,

semi-parametric estimator used by HMR and is defined by (7); **GPML** corresponds to the estimation of (6), in the multiplicative form, using gamma-pseudo maximum likelihood;[6] and, finally, **M-Bins** are the results of estimating (8), also in the multiplicative form using gamma-pseudo maximum likelihood.

As noted in Section 2, the NLS estimator proposed by HMR is based on a selectivity correction which is only approximately valid. Therefore, it is interesting to compare its results with those obtained using (6), which are labeled GPML in Table 1. At least for this particular example, the results in Table 1 suggest that, as argued before, the approximation implicitly used by HMR is reasonably accurate. Indeed, the estimates obtained with NLS and GPML are generally reasonably close. The major difference is, perhaps, that with GPML, FTA has a sizable and significant coefficient, which is not the case with NLS. A second difference is that GPML leads to an estimate of the distance elasticity that is about ten percent smaller that the one obtained with the NLS estimator based on the approximate selectivity correction.

It is important to notice that the results presented here under the label GPML correspond to the estimation of (6) under the assumption that the variance of the errors is quadratic in the mean, which is comparable to the assumption made when estimating models in the logged form, as is the case with NLS and A-Bins. Estimation of GPML using different methods, like the Poisson pseudo-maximum likelihood (which assumes that the variance of the errors is proportional to the mean), leads to substantially different results.[7] Given that both estimators are consistent under the same set of assumptions, this instability is worrisome and may indicate misspecification of the model. This point is pursued next.

As it was pointed out in Section 2, all the estimators based on the HMR model for trade flows assume that both u_{ij} and v_{ij} are homoskedastic. This assumption is quite

2008, p. 462, fn. 31). In the data set used by HMR, $\hat{z}^*_{ij}$ reaches values above 11. Baranga (2008) independently noted the importance of this censoring.

[6] The objective function that is minimized in the GPML estimation has the form $\left(-M_{ij}/\mu_{ij}\right) - \ln\left(\mu_{ij}\right)$, where μ_{ij} denotes the conditional mean specified by the model.

[7] These results are not reported here in the interest of brevity, but are available on request.

Table 1: Baseline results (Costs excluded)

	Probit	**OLS**	**NLS**	**A-Bins**	**GPML**	**M-Bins**
Variables	T_{ij}	m_{ij}	m_{ij}	m_{ij}	M_{ij}	M_{ij}
Log distance	−0.584	−1.167	−0.990	−0.789	−0.888	−0.786
	(0.043)	(0.040)	(0.039)	(0.088)	(0.037)	(0.076)
Land border	−0.230	0.627	0.723	0.863	0.837	0.896
	(0.183)	(0.165)	(0.161)	(0.170)	(0.137)	(0.140)
Island	−0.454	−0.553	−0.402	−0.197	−0.513	−0.402
	(0.200)	(0.269)	(0.259)	(0.258)	(0.215)	(0.215)
Landlock	−0.145	−0.432	−0.393	−0.353	−0.282	−0.228
	(0.135)	(0.189)	(0.185)	(0.187)	(0.170)	(0.167)
Legal	0.135	0.535	0.482	0.418	0.501	0.447
	(0.052)	(0.064)	(0.063)	(0.065)	(0.050)	(0.052)
Language	0.287	0.147	0.068	−0.036	0.011	−0.047
	(0.061)	(0.075)	(0.074)	(0.083)	(0.050)	(0.071)
Colonial ties	−0.026	0.909	0.892	0.838	0.835	0.846
	(0.353)	(0.158)	(0.150)	(0.153)	(0.150)	(0.155)
Currency union	0.743	1.534	1.308	1.107	1.347	1.313
	(0.182)	(0.334)	(0.324)	(0.346)	(0.303)	(0.309)
FTA	2.681	0.976	0.385	0.065	0.599	0.363
	(0.524)	(0.247)	(0.224)	(0.348)	(0.205)	(0.306)
Religion	0.385	0.281	0.191	0.100	−0.016	−0.073
	(0.093)	(0.120)	(0.118)	(0.128)	(0.096)	(0.102)
R. costs	−0.291	−0.146	—	—	—	—
	(0.095)	(0.100)	—	—	—	—
R. costs (D&P)	−0.163	−0.216	—	—	—	—
	(0.080)	(0.124)	—	—	—	—
RESET (p-value)	0.000	0.000	—	0.000	—	0.000
Sample size	12198	6602	6602	6602	6602	6602

important in that, if it does not hold, at least δ will be a function of the regressors and therefore all the models considered (and whose results are displayed in Table 1) willmisrepresent the effects of the covariates, leading to inconsistent estimators. Therefore, it is important to use appropriate tests to check for the correct specification of the functional form of all the models whose results are reported in Table 1.

The penultimate row of Table 1 presents the p-values for the two-degrees-of-freedom RESET-type tests described in Section 2 for all the models based on a linear index.[8] The p-value of the RESET test for the probit model reveals clear signs of misspecification, casting doubts over the validity of the results obtained with the two-stage estimators. These doubts are confirmed by the results of the specification test for the A-Bins and M-Bins models.[9] Therefore, there are reasons to suspect that all the models considered are misspecified, which precludes any meaningful interpretation of the displayed estimates.

To give some idea of the magnitude and relevance of this misspecification, we note that when the powers of the linear index are added to the specification of A-Bins and M-Bins, the coefficient on log distance changes by a factor of 2. Of course, no particular significance can be attributed to the coefficient on log distance in these auxiliary regressions, but these results illustrate the sensitivity of the results to the particular functional form that is assumed.

3.2. Alternative excluded variables

In order to check the robustness of their findings, HMR also present estimation results when the variable Religion provides the exclusion restriction used to help in the identification of the two-stage estimators. This also permits the use of a much larger sample

[8]Recall that low p-values mean rejection of the null hypothesis that the models are correctly specified.

[9]RESET-type tests can also certainly be performed for the NLS and HMR estimators, although that is non-standard because they are not based on linear indices. In any case, RESET tests for NLS and HMR would be somewhat redundant because their generalizations, A-Bins and M-Bins, are clearly rejected.

as now it is possible to use the observations for which information on costs are not available.

Table 2 reports the estimation results using the larger sample and using Religion as the excluded variable. The estimators whose results are displayed in Table 2 are the same as in Table 1.

As for the baseline case, we again find that the results for the NLS and GPML estimators are reasonably close, which supports our conjecture that the approximation used by HMR is reasonably accurate in typical applications. Notwithstanding, we notice again some sizable differences for the effects of log distance and FTA.

More importantly, we again find that all models badly fail the RESET test, which again casts serious doubts on the possibility of obtaining any meaningful insight from the reported results.

Overall, as in HMR, we find that the results with this alternative exclusion restriction and larger sample fully confirm the baseline estimates. Unfortunately, in both cases, the results are not particularly encouraging about the possibility of consistently estimating the parameters in the model for trade flows proposed by HMR.

4. Concluding remarks

In this note we discuss some econometric aspects of the implementation of the model for bilateral trade flows between countries proposed by Helpman, Melitz and Rubinstein (2008).

In particular, we argue that while the selectivity correction used by HMR is only approximately valid, the approximation is likely to be reasonably accurate in many empirical studies. The results reported in Section 3 support this conjecture in that the estimates obtained with the two-step method proposed by HMR are generally close to the ones obtained with the appropriate selectivity correction developed in Section 2.

More importantly, he have emphasized that consistent estimation of the parameters in the model proposed by HMR is only possible under the assumption that all random components of the model are homoskedastic. This dependence on the homoskedasticity

Table 2: Alternative excluded variables (Religion excluded)

	Probit	**OLS**	**NLS**	**A-Bins**	**GPML**	**M-Bins**
Variables	T_{ij}	m_{ij}	m_{ij}	m_{ij}	M_{ij}	M_{ij}
Log distance	−0.660	−1.176	−1.026	−0.623	−0.919	−0.642
	(0.029)	(0.031)	(0.030)	(0.076)	(0.027)	(0.063)
Land border	−0.382	0.458	0.580	0.924	0.698	0.879
	(0.129)	(0.147)	(0.142)	(0.150)	(0.124)	(0.122)
Island	0.345	−0.391	−0.318	−0.074	−0.115	0.068
	(0.082)	(0.121)	(0.117)	(0.121)	(0.101)	(0.104)
Landlock	0.181	−0.561	−0.522	−0.439	−0.524	−0.449
	(0.114)	(0.188)	(0.183)	(0.186)	(0.177)	(0.175)
Legal	0.096	0.486	0.445	0.345	0.452	0.388
	(0.034)	(0.050)	(0.049)	(0.050)	(0.040)	(0.041)
Language	0.284	0.176	0.132	−0.062	0.105	−0.038
	(0.042)	(0.061)	(0.059)	(0.068)	(0.053)	(0.059)
Colonial ties	0.325	1.299	1.192	0.929	1.240	1.042
	(0.305)	(0.120)	(0.117)	(0.119)	(0.111)	(0.111)
Currency union	0.492	1.364	1.250	0.960	1.349	1.167
	(0.143)	(0.255)	(0.253)	(0.270)	(0.252)	(0.254)
FTA	1.985	0.759	0.398	−0.091	0.844	0.327
	(0.315)	(0.222)	(0.206)	(0.210)	(0.172)	(0.163)
Religion	0.261	0.102	—	—	—	—
	(0.063)	(0.096)	—	—	—	—
RESET (p-value)	0.000	0.000	—	0.000	—	0.000
Sample size	24649	24649	11146	11146	11146	11146

assumption is the most important drawback of the HMR model, and contrasts with more standard models for trade (e.g., Anderson and van Wincoop, 2003), which can be made robust to the presence of heteroskedasticity.

Considering its importance, we discuss possible tests to check for departures from this assumption and use them to assess the specification of the models estimated by HMR. The results reported in Section 3 provide overwhelming evidence that, no matter how we account for selectivity, all estimators based on the HMR model for bilateral trade flows are misspecified. This, of course, casts doubts on the validity of any inference drawn upon these results.

In conclusion, it seems that the assumptions needed for the estimation of the HMR model for trade flows are too strong to make it practical. In particular, the presence of heteroskedasticity in trade data seems to preclude the estimation of any model that purports to identify the effects of the covariates in the intensive and extensive margins, at least with the current econometric technology.

References

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

Baranga, T. (2008). "How Empirically Important are Zero Trade Flows and Firm-level Heterogeneity when Estimating Gravity Equations?" mimeo, Harvard University: `http://www.people.fas.harvard.edu/~baranga/HMRcrit040908.pdf`.

Chen, S. and Khan, S. (2003), "Root-n consistent estimation of heteroskedastic sample selection models," *Econometric Theory*, 19, 1040-1064.

Cosslett, S.R. (1991), "Semiparametric estimation of a regression model with sample selectivity," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. Barnett, W.A., Powell, J. and Tauchen, G., 175–197, Cambridge: Cambridge University Press.

Godfrey, L.G. (1988). *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

Greene, W. (1994), *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*, Working Paper No. EC-94-10, Department of Economics, Stern School of Business, New York University.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

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

Newey, W. (1985), "Maximum Likelihood Specification Testing and Conditional Moments Tests," *Econometrica*, 53, 1047-1070.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society B*, 31, 350-371.

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

Terza, J. (1998), "Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects," *Journal of Econometrics*, 84, 129–154.

van de Ven, W.P. and van Praag, B.M. (1981), "Risk aversion of deductibles in private health insurance: Application of an adjusted Tobit model to family health care expenditures," in *Health, Economics and Health Economics*, ed. van der Gaag, J. and Perlman, M., 125–148, Amsterdam: North Holland.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.