J.M.C. Santos Silva and Silvana Tenreyro
Trading partners and trading volumes: implementing the Helpman-Melitz-Rubinstein model empirically

Article (Accepted version)
(Refereed)

Trading partners and trading volumes: Implementing the Helpman-Melitz-Rubinstein model empirically[*]

J.M.C. SANTOS SILVA[†] and SILVANA TENREYRO[‡]

† University of Essex and CEMAPRE. Wivenhoe Park, Colchester CO4 3SQ, UK (*e-mail: jmcss@essex.ac.uk*)

‡ London School of Economics, CFM, CEP, and CEPR. St. Clement's Building, Houghton St., London WC2A 2AE, UK (*e-mail: s.tenreyro@lse.ac.uk*)

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. We argue that this estimation procedure is only valid under the strong distributional assumptions maintained in the paper. Statistical tests using the HMR sample, however, clearly reject such assumptions. Moreover, we perform numerical experiments which show that the HMR two-stage estimator is very sensitive to departures from the assumption of homoskedasticity. These findings cast serious doubts on any inference drawn from the empirical implementation of the HMR model.

I. 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 is especially appealing because it can potentially 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).[1] Although HMR's model makes a significant step towards a better understanding of the determinants of bilateral trade flows, the proposed two-stage estimation procedure has some limitations.

In this paper we analyze the estimation method proposed by HMR and emphasize the two following results. First, the approach used by HMR to deal with the selectivity bias caused by dropping the observations with zero trade is only approximately correct. We discuss the conditions under which the approximation tends to work better. Second, and more importantly, the HMR model and associated estimator depend critically on untested distributional assumptions. As we show in this paper, such assumptions are strongly rejected by the HMR data. Moreover, we show that the results of the two-stage estimation method proposed by HMR are very sensitive to the presence of

---

[1]An alternative two-stage procedure to estimate the extensive and intensive margins is proposed by Egger, Larch, Staub, and Winkelmann (2011).

heteroskedasticity.[2] These findings cast serious doubts on any inference drawn from the empirical implementation of the HMR model using their proposed two-stage estimator.

II. The HMR model

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

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

where M_{ij} denotes the trade flow from j to i, Λ_j denotes a fixed effect for exporter j, X_i is a fixed effect for importer i, τ_{ij} represents the usual 'melting iceberg' transport cost, a_{ij} is a measure of the minimum productivity needed for it to be profitable for a firm to export from j to i, a_L is a measure of the productivity of the most productive firm, and B_0, k, and ε are parameters. Furthermore, the authors assume that

$$\tau_{ij}^{\varepsilon-1} = D_{ij}^{\gamma} \exp\left(-u_{ij}\right), \tag{2}$$

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

$$M_{ij} = B_0 \Lambda_j X_i D_{ij}^{-\gamma} \max \left\{ \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1, 0 \right\} \exp\left(u_{ij}\right). \tag{3}$$

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 equations (10) and (11) in HMR)

$$Z_{ij} = \Gamma_0 \Xi_j \Upsilon_i D_{ij}^{-\gamma} \Psi_{ij}^{-\kappa} \exp\left(v_{ij} + u_{ij}\right), \tag{4}$$

---

[2] Throughout we understand heteroskedasticity to mean that the skedastic function is not constant; that is, that the conditional variances of the errors of the model are functions of the regressors.

where Γ_0 and κ are parameters, Ξ_j denotes a fixed effect for exporter j, Υ_i is a fixed effect for importer i, Ψ_{ij} denotes additional country-pair specific fixed trade costs, and $v_{ij} \sim \mathcal{N}(0, \sigma_v)$.

The new variable Z_{ij} can be interpreted as the ratio of the variable export profits for the most productive firm to the fixed cost of exporting from j to i, and is not observable. However, positive trade is observed only when $Z_{ij} > 1$, which leads HMR to propose the following two-stage estimation strategy.

Let T_{ij} be a binary variable defined as $T_{ij} = 1 [M_{ij} > 0]$, where $1 [A]$ is the indicator function of the event A. Then, defining $z_{ij} = \ln(Z_{ij})$, $\gamma_0 = \ln(\Gamma_0)$, $\xi_i = \ln(\Xi_i)$, $y_j = \ln(\Upsilon_j)$, $d_{ij} = \ln(D_{ij})$, and $\psi_{ij} = \ln(\Psi_{ij})$, we have that the conditional probability that j exports to i is[3]

$$\rho_{ij} = \Pr(T_{ij} = 1) = \Pr\left(\gamma_0 + \xi_i + y_j - \gamma d_{ij} - \kappa \psi_{ij} > -(v_{ij} + u_{ij})\right). \tag{5}$$

Under the maintained assumptions of normality and homoskedasticity, the unknown parameters can be consistently estimated up to scale using a probit. Indeed, under these assumptions we have that

$$\Pr(T_{ij} = 1) = \Phi\left(\frac{\gamma_0 + \xi_i + y_j - \gamma d_{ij} - \kappa \psi_{ij}}{\sigma_{u+v}}\right), \tag{6}$$

where σ_{u+v} denotes the standard deviation of $(v_{ij} + u_{ij})$ and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Therefore, under the distributional assumptions made by HMR it is possible to consistently estimate

$$Z_{ij}^* = \left(\Gamma_0 \Xi_i \Upsilon_j D_{ij}^{-\gamma} \Psi_{ij}^{-\kappa}\right)^{\frac{1}{\sigma_{u+v}}}, \tag{7}$$

---

[3]To simplify the notation, throughout we do not make explicit that this probability is conditional on the regressors.

by taking the exponential of the linear index estimated by the probit model. Using this result and the fact that $Z_{ij} = \left(\frac{a_{ij}}{a_L}\right)^{\varepsilon-1}$ for $T_{ij} = 1$, it is possible to rewrite (3) as

$$M_{ij} = T_{ij} B_0 \Lambda_j X_i D_{ij}^{-\gamma} \left\{ \left[Z_{ij}^* \exp\left(\varsigma_{ij}\right)\right]^\delta - 1 \right\} \exp\left(u_{ij}\right), \tag{8}$$

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

The second stage 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 (8), leading to

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\left\{\exp\left[\delta\left(z_{ij}^* + \varsigma_{ij}\right)\right] - 1\right\} + u_{ij}, \tag{9}$$

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

Distributional assumptions

The definition of δ shows that it is proportional to the standard deviation of $\left(v_{ij} + u_{ij}\right)$. Therefore, δ is a parameter if v_{ij} and u_{ij} are homoskedastic but otherwise it is a function of the regressors. Hence, the homoskedasticity of the errors is critical to establish how the regressors enter both (8) and (9). Indeed, under the assumptions in HMR, the regressors enter the equation both directly and through Z_{ij}^*, which is estimated in the first stage. However, under heteroskedasticity, δ will also be a function of the regressors.

Therefore, under heteroskedasticity, the regressors will enter the model in a much more complex form than what is assumed by HMR. Ignoring that δ is a function of the regressors has the potential to introduce severe misspecification in (8) and (9), making consistent estimation of the parameters of interest generally impossible. Moreover, heteroskedasticity will also make the estimation of the first stage inconsistent, which

4

will bring an additional source of misspecification into the model.[4] Of course, one may be tempted to tackle this problem by specifying σ_{u+v} as a function of the regressors, but this approach is foiled by the fact that economic theory provides no guidance on the possible heteroskedasticity patterns.

Most of the results in HMR are also obtained under the assumption of normality. Although HMR partially relax it, normality is always assumed in the estimation of the first stage. Therefore, the twin assumptions of homoskedasticity and normality are critical for the correct specification of (8) and (9). Moreover, these assumptions are also critical for the construction of the selectivity corrections used by HMR.

Cosslett (1991), Chen and Khan (2003), and Das, Newey, and Vella (2003), among others, have studied semi-parametric estimators for linear sample selection models which are robust to non-normality and heteroskedasticity. However, the validity of these estimators depends on conditions such as particular forms of heteroskedasticity or the existence of valid exclusion restrictions in the second stage, which are unlikely to be valid in the context of trade data.[5] More importantly, (9) has two random components affected by sample selection and one of them enters the model in a non-linear form; none of the currently available semi-parametric estimators can deal with sample selection models of this form.

The selectivity correction

Estimation of (9) is performed using only observations with positive values of M_{ij}, which originates a sample-selection issue. To account for the fact that $E\left[u_{ij}|M_{ij}>0\right]\neq$

---

[4]As noted by Santos Silva and Tenreyro (2006), homoskedasticity is also critical for the validity of the log-linearization leading from (8) to (9). However, in this paper we focus on a very different issue: the additional implications that heteroskedasticity has in the context of the two-stage estimation proposed by HMR.

[5]Additionally, due to the large number of regressors typically used, implementation of some of these methods with trade data is far from trivial.

0, HMR include in the regression equation the inverse Mills ratio from the first stage, which (under normality and homoskedasticity) is proportional to $E\left[u_{ij}|M_{ij}>0\right]$. This is the correct procedure to account for selectivity in an additive error (see, e.g., Wooldridge, 2010).

However, the equation of interest has a second random component, ς_{ij}, which enters the equation within a non-linear function. HMR deal with the effect of the sample-selection on ς_{ij} in a way that is akin to the ad-hoc method used by Greene (1994). In particular, HMR replace ς_{ij} with its expectation conditional on $M_{ij}>0$, which is the inverse Mills ratio from the first stage. That is, using z_{ij}^* to denote the linear index in (6) and denoting the inverse Mills ratio by $\eta_{ij}=\phi\left(z_{ij}^*\right)/\Phi\left(z_{ij}^*\right)$, the second stage of HMR's procedure is the estimation of

$$m_{ij}=\beta_0+\lambda_j+\chi_i-\gamma d_{ij}+\ln\left\{\exp\left[\delta\left(z_{ij}^*+\hat{\eta}_{ij}\right)\right]-1\right\}+\beta_{u\eta}\hat{\eta}_{ij}+e_{ij},\qquad(10)$$

where $\hat{\eta}_{ij}$ denotes the fitted value of η_{ij} and $\beta_{u\eta}$ is a parameter (see equation (14) in HMR).

As noted for example by Terza (1998), this approach to correct the effect of the sample selection on ς_{ij} is generally inappropriate. Indeed, for any non-linear function $f\left(\cdot\right)$, Jensen's inequality implies that $f[E(\varsigma_{ij}|M_{ij}>0)]\neq E\left[f\left(\varsigma_{ij}\right)|M_{ij}>0\right]$. Therefore, $f\left(\hat{\eta}_{ij}\right)$ is not a consistent estimator of $E\left[f\left(\varsigma_{ij}\right)|M_{ij}>0\right]$ and, consequently, the proposed estimation method will generally be inconsistent for all the parameters of interest.

Nevertheless, it is interesting to notice that the approximation used by HMR is likely to be reasonably accurate in many practical situations. To see this, notice that

$$\ln\left\{\exp\left[\delta\left(z_{ij}^*+\varsigma_{ij}\right)\right]-1\right\}=\ln\left\{\exp\left[\delta\left(z_{ij}^*+\eta_{ij}+\omega_{ij}\right)\right]-1\right\},\qquad(11)$$

6

where $\omega_{ij} = \varsigma_{ij} - \eta_{ij}$. The approximation used by HMR consists of 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, (11) is approximately linear in ω_{ij}. Therefore, under their assumptions, 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.[6]

III. A reappraisal of the HMR study

In this section we reconsider the empirical study presented in HMR.[7] We start by testing whether there is evidence of violations of the distributional assumptions required for the validity of the HMR estimator and then study the sensitivity of the results to the presence of heteroskedasticity.

Testing the distributional assumptions

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 defined by (6). Tests for this purpose were introduced by Davidson and MacKinnon (1984) and are now described in textbooks such as Wooldridge's (2010, pp. 571-573). Although these tests are well known, it is perhaps useful to briefly present them here.

---

[6] Santos Silva and Tenreyro (2009) show that, under the distributional assumptions maintained by HMR, it is possible to obtain an exact selectivity correction.

[7] In order to maintain comparability with the results in HMR, we use exactly the same data, the same set of regressors, and the same estimation methods used in the original study. HMR provide details on the data and data sources.

Consider the following generalization of (6)

$$\Pr\left(T_{ij} = 1\right) = \Phi\left(\frac{\gamma_0 + \xi_i + y_j - \gamma d_{ij} - \kappa\psi_{ij}}{\sigma_{u+v}(w_{ij}\theta)}\right), \tag{12}$$

where w_{ij} is a vector of functions of ξ_i, y_j, d_{ij}, and ψ_{ij}, θ is a vector of parameters, and $\sigma_{u+v}(w_{ij}\theta)$ denotes the standard deviation of $u + v$. Moreover, let $\sigma_{u+v}(w_{ij}\theta)$ be such that $\sigma_{u+v}(0) = \sigma_{u+v}$, which is constant. In this setup the null of homoskedasticity can tested by testing $H_0 : \theta = 0$ and, as Davidson and MacKinnon (1984) show, under mild regularity conditions the test can be derived without specifying the form of $\sigma_{u+v}(w_{ij}\theta)$. Indeed, by expanding the argument of $\Phi\left(\cdot\right)$ in a Taylor series around $\theta = 0$ we obtain

$$\Pr\left(T_{ij} = 1\right) \simeq \Phi\left(z_{ij}^* - \theta\sigma'_{u+v}z_{ij}^* w_{ij}/\sigma_{u+v}\right), \tag{13}$$

where z_{ij}^* is defined as before and σ'_{u+v} denotes the derivative of $\sigma_{u+v}(\cdot)$ evaluated at zero. Therefore, if $\sigma'_{u+v} \neq 0$, (12) is locally equivalent to a homoskedastic probit where the linear index z_{ij}^* is augmented by the inclusion of variables of the form $z_{ij}^* w_{ij}$ and $H_0 : \theta = 0$ can be tested by testing the significance of the parameters associated with the additional regressors.

The test can be implemented in three simple steps: first, we estimate the model under the null, i.e., (6); then we construct variables of the form $\hat{z}_{ij}^* w_{ij}$, where $\hat{z}_{ij}^*$ is the fitted value of z_{ij}^* obtained under the null; finally, we estimate the augmented probit model defined by (13) and check the significance of the parameters associated with the additional regressors. That is, in the context of a probit model, homoskedasticity can be tested by testing the exclusion of a particular type of regressors.

To implement the test it is necessary to choose the variables in w_{ij}. Although w_{ij} can contain any function of the regressors, in what follows we will consider only the

case where $w_{ij} = \left(\hat{z}_{ij}^{*}, \hat{z}_{ij}^{*2}\right)$;[8] that is, the test is performed by checking for the joint significance of the parameters associated with the additional regressors $\hat{z}_{ij}^{*2}$ and $\hat{z}_{ij}^{*3}$.

This form of the heteroskedasticity test, which is analogous to a two-degrees-of-freedom RESET test (Ramsey, 1969), is particularly interesting in the case of a probit model because it can also be interpreted as a normality test. Heuristically, the intuition for this can be provided as follows (see Cramer and Ridder, 1988, pp. 307-308). Suppose that $\Pr\left(T_{ij} = 1\right) = F\left(z_{ij}^{*}\right)$, where $F\left(\cdot\right)$ is a cumulative distribution function, and rewrite $F\left(z_{ij}^{*}\right)$ as $\Phi\left(\Phi^{-1}\left(F\left(z_{ij}^{*}\right)\right)\right)$. Then, approximating $\Phi^{-1}\left(F\left(z_{ij}^{*}\right)\right)$ with a third order polynomial we obtain $F\left(z_{ij}^{*}\right) = \Phi\left(\upsilon_1 z_{ij}^{*} + \upsilon_2 z_{ij}^{*2} + \upsilon_3 z_{ij}^{*3}\right)$. If $F\left(\cdot\right) = \Phi\left(\cdot\right)$, then $\Phi^{-1}\left(F\left(z_{ij}^{*}\right)\right) = z_{ij}^{*}$ and therefore $\upsilon_2 = \upsilon_3 = 0$. That is, in a probit the assumption of normality can be tested using a RESET test; see Newey (1985) for an alternative derivation and more details.

Ramalho and Ramalho (2012) perform an extensive simulation study on the properties of RESET tests in the context of binary choice models and conclude that the particular version of the RESET we use has very good performance under the null and has good power against a range of alternatives, including heteroskedasticity and non-normality. Therefore, this simple test provides a direct check for the validity of the main distributional assumptions required for consistent estimation of the model developed by HMR.

Additionally, because heteroskedasticity also impacts on the functional form of (10) it is important to check whether the specification of this model is reasonably adequate. In the spirit of Cosslett (1991), HMR partially relax the distributional assumptions used to obtain (10) by estimating models of the form

$$m_{ij} = \lambda_j + \chi_i - \gamma d_{ij} + \sum_{s=1}^{Q} \alpha_s 1\left[q_{s-1} < \rho_{ij} \leq q_s\right] + e_{ij}^{*}, \tag{14}$$

---

[8]This choice is motivated by analogy with the popular two-degrees-of-freedom special case of White's test for heteroskedasticity (see Wooldridge, 2010, p. 140).

where, as before, $\rho_{ij} = \Pr(T_{ij} = 1)$ and $1[A]$ is the indicator function of the event A, $\alpha_1, \ldots, \alpha_Q$ are parameters, $q_0, \ldots, q_Q$ are constants defining quantiles of ρ_{ij}, and $q_0 = -\infty$ and $q_Q = \infty$. Although this model is more flexible than (10) it still assumes that the selectivity correction depends on the regressors only through z_{ij}^*; that is, like (8), (9), and (10), (14) assumes that δ is constant. To check for departures from this assumption one can test 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 test for the significance of additional variables constructed as powers of the estimated linear indexes. By analogy with what is done for the probit, in what follows we will also use two-degrees-of-freedom RESET tests to check the validity of (14). The performance of the RESET in linear models has been studied, among others, by Godfrey and Orme (1994), who conclude that the test has good behaviour under the null and good power against alternatives of the type considered here.

Table 1 presents the test statistics and p-values for the two-degrees-of-freedom RESET tests for some of the models estimated by HMR. In particular, for the two sets of exclusion restrictions considered by HMR, we test the specification of the probit used in the first stage and the two most flexible specifications of the second stage, which are defined by (14) with $Q \in \{50, 100\}$. Although the simulation studies referred above suggest that the RESET tests have good size properties, it is important to make sure that in this particular application the tests do not lead to spurious rejections of the null. Therefore, we computed the p-values of the test in 3 different ways: 1) using the usual clustered standard errors, 2) using standard errors obtained with a bootstrap by clusters, and 3) bootstrap p-values. Notice that because the test statistic is asymptotically pivotal the bootstrap p-values benefit from the well-know asymptotic refinements and therefore lead to much more reliable inference than the p-values obtained by the first two methods (see, e.g., Godfrey, 2009, pp. 69-72). In the three cases the results are the same.

TABLE 1: *Specification test results*

	Costs excluded			Religion excluded		
	Probit	*50 Bins*	*100 Bins*	*Probit*	*50 Bins*	*100 Bins*
RESET statistic	105.45	24.99	24.00	323.04	58.58	58.87
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sample size	$12,198$	$6,602$	$6,602$	$24,649$	$11,146$	$11,146$

The p-values of the RESET tests for the probit models reveal clear signs of misspecification. As noted before, this particular version of the RESET test can be interpreted as a test for the assumptions of homoskedasticity and normality of u_i and v_i and the results in Table 1 strongly suggest that these assumptions are not valid in this context. This impression is reinforced by the results for the second stage models, which also clearly fail the RESET tests. Therefore, there are reasons to suspect that all the models considered by HMR are misspecified; we next evaluate how sensitive the results are to possible misspecification.

Gauging the consequences of heteroskedasticity

One way to assess the sensitivity of the HMR estimation procedure to the presence of heteroskedasticity is to estimate (14) using fitted values of ρ_{ij} obtained from an heteroskedastic probit as in (12) rather than from a standard probit. To do that we need to specify the functional form of $\sigma_{u+v}(w_{ij}\theta)$ and to define the set of variables to include in w_{ij}. Here we follow Harvey (1976) and specify $\sigma_{u+v}(w_{ij}\theta) = \exp(w_{ij}\theta)$. As for the choice of w_{ij} we consider two cases: a) $w_{ij} = (d_{ij}, \psi_{ij})$, that is, w_{ij} contains all the regressors in the original probit regression except the importer and exporter dummies; and b) w_{ij} is a set of indicator variables obtained by partitioning the fitted values of (6) into bins as done in (14); in this case we use 35 bins.[9] It is important

---

[9]Estimation of the heteroskedastic probit is relatively difficult and therefore it is necessary to be somewhat parsimonious in the definition of w_{ij}.

to notice that because economic theory provides no guidance on the choice of the functional form of $\sigma_{u+v}(w_{ij}\theta)$ or of the variables in w_{ij} our decisions on these are somewhat arbitrary and we do not in any way presume that they lead to adequate specifications. Consequently, the results presented below should be interpreted merely as illustrating the sensitivity of the HMR estimation procedure to alternative forms of accounting for heteroskedasticity in the first stage.

Table 2 presents the results obtained by estimating (14) using three different specifications for the probit model in the fist stage:[10] columns (A) and (D) correspond to the results reported by HMR using the standard probit that ignores the possible presence of heteroskedasticity, columns (B) and (E) correspond to the results obtained using the heteroskedastic probit with $w_{ij} = \left(d_{ij}, \psi_{ij}\right)$, and finally columns (C) and (F) are obtained again by using an heteroskedastic probit where now w_{ij} is a set of indicator variables obtained as described above.

The results presented in Table 2 show that the way heteroskedasticity is accounted for in the first stage has dramatic consequences for the estimated elasticities of the firm's trade with respect to the different trade barriers. Indeed, changing the specification of $\sigma_{u+v}(w_{ij}\theta)$ leads to important changes in the magnitude of the estimated elasticities, sometimes changing their statistical significance (e.g., Colonial ties and FTA). For example, the estimated distance elasticity varies between -0.623 and -1.073, and the estimate for the coefficient on the currency union dummy varies between 0.782 and 1.376, implying that the effect of this variable varies by a factor of almost 2.5. Additionally, we note that in (C) the coefficient on Religion has a p-value of 0.029, which calls into question the use of this variable as the excluded instrument in the second set of estimates. Finally, we note that although the models based on the heteroskedastic probit are more general than the ones used by HMR, the results of the RESET test

---

[10]We only present results for the case where the indicators are obtained as percentiles of ρ_{ij}, results with fewer indicator dummies are essentially the same.

TABLE 2: *Estimation Results*

w_{ij}	Costs excluded			Religion excluded		
	(A)	(B)	(C)	(D)	(E)	(F)
	Const.	d_{ij}, ψ_{ij}	Bins	Const.	d_{ij}, ψ_{ij}	Bins
Log distance	−0.789	−0.890	−1.062	−0.623	−0.965	−1.073
	(0.088)	(0.048)	(0.041)	(0.076)	(0.044)	(0.034)
Land border	0.863	1.166	0.705	0.924	0.808	0.636
	(0.170)	(0.177)	(0.165)	(0.150)	(0.153)	(0.146)
Island	0.197	−0.099	0.478	0.074	0.243	0.287
	(0.258)	(0.279)	(0.266)	(0.121)	(0.119)	(0.118)
Landlock	0.353	0.334	0.352	0.439	0.379	0.569
	(0.187)	(0.188)	(0.190)	(0.186)	(0.192)	(0.187)
Legal	0.418	0.339	0.489	0.345	0.354	0.442
	(0.065)	(0.068)	(0.063)	(0.050)	(0.051)	(0.049)
Language	−0.036	−0.025	0.110	−0.062	0.081	0.135
	(0.083)	(0.076)	(0.075)	(0.068)	(0.062)	(0.060)
Colonial ties	0.838	−0.386	0.901	0.929	0.894	1.123
	(0.153)	(0.264)	(0.152)	(0.119)	(0.339)	(0.118)
Currency union	1.107	0.782	1.376	0.960	1.052	1.233
	(0.346)	(0.339)	(0.334)	(0.270)	(0.267)	(0.262)
FTA	0.065	0.203	0.751	−0.091	0.494	0.669
	(0.348)	(0.257)	(0.224)	(0.210)	(0.311)	(0.189)
Religion	0.100	−0.255	0.259	—	—	—
	(0.128)	(0.146)	(0.119)	—	—	—
RESET statistic	24.00	22.13	33.89	58.85	68.22	87.11
p-value	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.706	0.706	0.703	0.723	0.722	0.720
Sample size	6,602	6,602	6,602	11,146	11,146	11,146

Note: Clustered standard errors in parentheses. All models include importer and exporter dummies.

13

suggest that they no not alleviate the misspecification of the second stage. This is not surprising because, due to the log transformation, the consistency of the second stage is also heavily dependent on the assumption of homoskedastic errors (see Santos Silva and Tenreyro, 2006) and that is not at all taken into account in the estimation procedure suggested by HMR.

The results in Table 2 clearly illustrate the sensitivity of the HMR estimation procedure to the presence of heteroskedasticity. However, because all of the models are potentially severely misspecified, these results are not informative about the magnitude of the biases caused by heteroskedasticity in the first stage.

To investigate this issue we performed a small simulation study in which data are generated as follows. First, using (4), Z_{ij} is generated as

$$Z_{ij} = \exp\left(\gamma_0 + \xi_i + y_j - \gamma d_{ij} - \kappa \psi_{ij} + v_{ij} + u_{ij}\right). \tag{15}$$

Then, for $Z_{ij} > 1$, m_{ij} is generated according to (9), that is

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\left\{Z_{ij}^{\frac{k-\varepsilon+1}{\varepsilon-1}} - 1\right\} + u_{ij}. \tag{16}$$

For the instruments ψ_{ij}, we used either Costs or Religion as in HMR. We also performed a set of experiments similar to those where Religion is the instrument but replacing this variable by random draws from a normal distribution with mean 0 and standard deviation equal to 3. When Costs and Religion are the instruments the sample sizes are as in the HMR paper and when the random instrument is used the sample size is the same as when Religion is used as the instrument.

The variables in d_{ij} are a subset of the trade barriers included in the original study. Specifically, to speed-up the simulation experiments we considered only Log distance,

14

Land border, Landlock, and Legal. As in HMR, when Costs are the excluded instrument Religion is also included in d_{ij}.[11]

We performed experiments with homoskedastic and heteroskedastic errors. In the homoskedastic case v_{ij} and u_{ij} are obtained as random draws from a normal distribution with mean 0 and standard deviation equal to $\sqrt{0.5}$; therefore $\sigma_{u+v} = 1$. In the heteroskedastic case u_{ij} is obtained as random draws from a normal distribution with mean 0 and standard deviation equal to 0.25 and v_{ij} is obtained as random draws from a normal distribution with mean 0 and standard deviation equal to $\sqrt{\exp\left(\theta_1 d_{ij} + \theta_2 \psi_{ij}\right)^2 - 0.25^2}$, implying $\sigma_{u+v}(w_{ij}\theta) = \exp\left(\theta_1 d_{ij} + \theta_2 \psi_{ij}\right)$.[12]

To complete the specification of the data generation process it is necessary to set the parameters of the model.[13] In the homoskedastic case, the parameters in (15) are set to match the estimates obtained when the first stage is a probit with the relevant regressors. In the heteroskedastic case, the parameters in (15) and θ_1 and θ_2 are set to match the estimates obtained when the first stage is a heteroskedastic probit. In both cases the coefficients on the regressors in (16) were set to match the estimates obtained from (14) and constructing the indicator functions as the percentiles of the probabilities estimated in the first stage.[14] Finally, we set $k = 1.85$ and $\varepsilon = 2$ (see Broda and Weinstein, 2006).

After generating the data, the model was estimated using a probit in the first stage and then estimating the second stage using (14) with indicator functions constructed as the percentiles of the probabilities estimated in the first stage; i.e., estimation is per-

---

[11] We preformed some limited experiments with the full set of regressors and the results are qualitatively similar.

[12] Notice that in both cases u_{ij}, the error of the second stage, is homoskedastic.

[13] Values for the main parameters used in the simulations are given in Appendix A.

[14] Notice, however, that the biases to be reported below are invariant to the value of these parameters.

formed exactly as in HMR and therefore ignores the possible presence of heteroskedasticity.

Table 3 reports the biases of the estimates of the coefficients of d_{ij} obtained with $10,000$ replicas of the procedure described above. These results show that when Costs or Religion are used as instruments there are substantial biases even if the errors are homoskedastic; the bias on the coefficient of Log distance is particularly noteworthy. This suggests that, at least with samples of this size, the instruments used by HMR do not have enough influence in the first stage to lead to estimates of the parameters of interest with reasonably small biases. However, when $\psi_{ij} \sim \mathcal{N}(0,3)$ and the errors are homoskedastic the biases are generally reasonably small.[15]

With heteroskedastic errors the biases again depend on how the data are generated and on the instrument used, and can be very substantial. Moreover, the results show that the sign of the biases depends on how the data is generated.[16]

TABLE 3: *Simulation Results (estimated biases)*

	Homoskedastic errors			*Heteroskedastic errors*		
Instrument	Costs	Religion	$\mathcal{N}(0,3)$	Costs	Religion	$\mathcal{N}(0,3)$
Log distance	-0.3558	-0.4607	-0.0086	-0.7928	-1.4192	-0.3872
Land border	-0.0871	-0.1744	0.0190	0.8098	-0.0727	0.2943
Landlock	0.0772	0.1262	0.0040	0.0225	0.2468	-0.2038
Legal	0.1144	0.1311	0.0002	0.1649	0.3418	0.0274
Religion	0.2094	—	—	0.1326	—	—
Sample size	$6,602$	$11,146$	$11,146$	$6,602$	$11,146$	$11,146$

The results in Table 3 confirm that the two-stage estimation procedure suggested by HMR is heavily reliant on the assumption that the errors of the model are homoskedas-

---

[15] Under homoskedasticity, smaller biases are also obtained with the other instruments if κ, the coefficient of the instrument in the generation of Z_{ij}, is increased substantially.

[16] The results of additional experiments show that, as expected, the magnitude of the biases increase with the value of $(k - \varepsilon + 1)/(\varepsilon - 1)$.

tic. Moreover, these results show that even under homoskedasticity the ability of the estimator to identify the parameters of interest depends on the availability of an instrument that not only can be validly excluded form the second stage regression but also has a determinant role in the first stage. The results obtained using Costs and Religion as instruments suggest that, even if they are valid, their role in the first stage is not significant enough to allow the estimation of the parameters of interest with reasonably small biases.

IV. Concluding remarks

In this paper 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 have emphasized that consistent estimation of the structural 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 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. Additionally, our simulation results reveal that estimation of the HMR model using Costs or Religion as instruments may lead to substantial biases, even if the errors are homoskedastic and the instruments valid. This happens because Costs or Religion play only a minor role in the first stage and in that sense they are weak instruments. Finally, we notice that while it is likely to be reasonably accurate in many empirical studies, the selectivity correction used by HMR is only approximately valid.

To gauge the severity of these problems we revisited the empirical illustration presented by HMR and found overwhelming evidence that all models used in their study

are misspecified. Additionally, we illustrated that their proposed estimator is very sensitive to the presence of heteroskedasticity. These findings cast doubts on the validity of any inference drawn upon the results obtained using the two-stage estimator of the model for bilateral trade flows proposed by HMR.

References

Anderson, J. and E. van Wincoop (2003), 'Gravity with gravitas: A solution to the border puzzle,' *American Economic Review*, 93, 170-192.

Broda, C. and Weinstein, D.E. (2006), 'Globalization and the gains from variety,' *Quarterly Journal of Economics*, 121, 541-585.

Chen, S. and Khan, S. (2003), 'Semiparametric estimation of a heteroskedastic sample selection model,' *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.

Cramer, J.S. and Ridder, G. (1988). 'The logit model in economics,' *Statistica Neerlandica*, 42, 291-314.

Das, M., Newey, W.K., and Vella, F. (2003). 'Nonparametric estimation of sample selection models,' *Review of Economic Studies*, 70, 33-58.

Davidson, R. and MacKinnon, J.G. (1984). 'Convenient specification tests for logit and probit models,' Journal of Econometrics, 25, 241-262.

Egger, P., Larch, M., Staub, K.E., and Winkelmann, R. (2011). 'The trade effects of endogenous preferential trade agreements,' *American Economic Journal: Economic Policy*, 3, 113-143.

Godfrey, L.G. and Orme, C.D. (1994). 'The sensitivity of some general checks to omitted variables in the linear model,' *International Economic Review*, 35, 489–506.

Godfrey, L.G. (2009). *Bootstrap tests for regression models*, Houndmills: Palgrave Macmillan.

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.

Harvey, A. C. (1976). 'Estimating regression models with multiplicative heteroscedasticity,' *Econometrica* 44, 461-465.

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.

Ramalho, E.A. and Ramalho, J.J.S. (2012), 'Alternative versions of the RESET test for binary response index models: a comparative study,' *Oxford Bulletin of Economics and Statistics*, 74, 107-130.

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, Silvana (2006), 'The log of gravity,' *The Review of Economics and Statistics*, 88, 641-658.

Santos Silva, J.M.C. and Tenreyro, Silvana (2009), 'Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically,' CEP Discussion Papers dp0935, Centre for Economic Performance, LSE.

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

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed, Cambridge, MA: MIT Press.

Appendix A

TABLE A1: *Main simulation parameters**

Instrument	Homoskedastic case		Heteroskedastic case	
	Costs	Religion	Costs	Religion
First Stage				
Log distance	−0.6801115	−0.7145263	−0.4630633	−0.6757597
Land border	−0.2053883	−0.3268029	0.0353037	−0.2673540
Landlock	0.1500615	0.2062646	0.0874460	0.2586772
Legal	0.2199632	0.1902099	0.1652583	0.2026328
Religion	0.4301583	0.3176966	0.2002376	0.2962447
Reg. costs	−0.2756534	—	−0.1174301	—
Days & proc.	−0.1513593	—	−0.0756938	—
$\sigma_{u+v}(w_{ij}\theta)$				
Log distance	—	—	−0.0665514	−0.0037485
Land border	—	—	0.4566915	0.2378667
Landlock	—	—	−0.0991268	−0.2010872
Legal	—	—	−0.1022340	−0.0682039
Religion	—	—	−0.4798412	−0.1183847
Reg. costs	—	—	−0.0624817	—
Days & proc.	—	—	−0.0926002	—
Second Stage†				
Log distance	−0.9783899	−0.6726285	−0.8552656	−0.7001846
Land border	0.8397781	0.9048378	1.3515540	1.1654240
Landlock	0.3628381	0.4477573	0.2162446	0.1051300
Legal	0.4992202	0.4170539	0.3069886	0.3025640
Religion	0.2023377	—	−0.6370270	—

∗ All models include importer and exporter dummies.

† The results are invariant to the choice of the second stage parameters.