The Log of Gravity*

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

October 13, 2004

*The authors are grateful to Francesco Caselli, Juan Carlos Hallak, Daniel Mota, Paulo Parente, Manuela Simarro and Kim Underhill for helpful suggestions and comments on previous versions of this paper. The usual disclaimer applies. Jiaying Huang provided excellent research assistance. Santos Silva gratefully acknowledges the partial financial support from Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER. A previous version of this paper circulated as "Gravity-Defying Trade."

†ISEG/Universidade Técnica de Lisboa, R. do Quelhas 6, 1200 Lisboa, Portugal. Fax: +351 213922781. E-mail: jmcss@iseg.utl.pt.

‡Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02106-2076, USA. Fax: +1 6179733957. E-mail: Silvana.Tenreyro@bos.frb.org.

The Log of Gravity

Abstract

Although economists have long been aware of Jensen's inequality, many econometric applications have neglected an important implication of it: estimating economic relationships in logarithms can lead to significant biases in the presence of heteroskedasticity. This paper explains why this problem arises and proposes an appropriate estimator. Our criticism to conventional practices and the solution we propose extends to a broad range of economic applications where the equation under study is log-linearized. We develop the argument using one particular illustration, the gravity equation for trade, and use the proposed technique to provide novel estimates of this equation. Three results stand out. First, contrary to general belief, income elasticities are significantly smaller than 1. Second, standard estimators greatly exaggerate the roles of distance and colonial links. Finally, trade gains associated with preferential-trade agreements are remarkably smaller than those predicted by conventional methods.

Key words: Preferential-trade agreements, Gravity equation, Heteroskedasticity, Jensen's inequality, Poisson regression.

JEL Codes: C21, F10, F11, F12, F15.

1. Introduction

Economists have long been aware that Jensen's inequality implies that $E(\ln y) \neq \ln E(y)$, i.e., the expected value of the logarithm of a random variable is different from the logarithm of its expected value. This basic fact, however, has been neglected in many econometric applications. Indeed, one important implication of Jensen's inequality is that the standard practice of interpreting the parameters of log-linearized models estimated by ordinary least squares (OLS) as elasticities can be highly misleading in the presence of heteroskedasticity.

Although many authors have addressed the problem of obtaining consistent estimates of the conditional mean of the dependent variable when the model is estimated in the log-linear form (see, for example, Goldberger, 1968; Manning and Mullahy, 2001), we were unable to find any reference in the literature to the potential bias of the elasticities estimated using the log-linear model.

In this paper we use the gravity equation for trade as a particular illustration of how the bias arises and propose an appropriate estimator. We argue that the gravity equation, and, more generally, constant-elasticity models should be estimated in their multiplicative form and propose a simple pseudo-maximum likelihood estimation technique. Besides being consistent in the presence of heteroskedasticity, this method also provides a natural way to deal with zero values of the dependent variable.

Using Monte Carlo simulations, we compare the performance of our estimator with that of OLS (in the log-linear specification). The results are striking. In the presence of heteroskedasticity, estimates obtained using log-linearized models are severely biased, distorting the interpretation of the model. These biases might be critical for the comparative assessment of competing economic theories, as well as for the evaluation of the effects of different policies. In contrast, our method is robust to the different patterns of heteroskedasticity considered in the simulations.

We next use the proposed method to provide new estimates of the gravity equation and, in particular, to reassess the impact of preferential-trade agreements on the volume of international trade. Our estimation method paints a very different picture of the determinants of international trade. The coefficients on GDP are clearly not, as generally believed, close to 1. Instead, they are significantly smaller, calling for modifications to the simple gravity models.[1] Incidentally, the smaller estimated elasticities help reconcile the gravity equation with the observation that the trade-to-GDP ratio decreases with total GDP (or, in other words, that smaller countries tend to be more open to international trade). In addition, OLS greatly exaggerates the roles of colonial ties and geographical proximity. Perhaps more interesting, the pseudo-maximum likelihood estimation indicates that, on average, bilateral trade between countries that have signed a preferential-trade agreement is 20 percent larger than trade between pairs without agreement. The trade increase predicted by OLS is remarkably larger. The striking contrast in estimates suggests that inferences drawn on the standard regressions used in the literature can produce misleading conclusions and confound policy decisions.

Despite this focus on the gravity equation, our criticism to the conventional practice and the solution we propose extends to a broad range of economic applications where the equations under study are log-linearized, or, more generally, transformed by a non-linear function. A short list of examples includes the estimation of Mincerian equations for wages, production functions, and Euler equations, which are typically estimated in logarithms.

The remainder of the paper is organized as follows. Section 2 studies the econometric problems raised by the estimation of gravity equations. Section 3 considers constant-elasticity models in general and introduces the pseudo-maximum likelihood estimator and specification tests to check the adequacy of the proposed estimator. Section 4 presents the Monte Carlo simulations. Section 5 provides new estimates of the gravity equation, revisiting the role of preferential-trade agreements in international trade. Section 6 contains concluding remarks.

[1] Note that a more complex – and complete – model of gravity, like the one proposed by Anderson and van Wincoop (2003) can rationalize our results, as their model is consistent with smaller income elasticities.

2. Gravity-Defying Trade

The pioneering work of Jan Tinbergen (1962) initiated a vast theoretical and empirical literature on the gravity equation for trade. Theories based on different foundations for trade, including endowment and technological differences, increasing returns to scale, and "Armington" demands, all predict a gravity relationship for trade flows analogous to Newton's "Law of Universal Gravitation."[2] In its simplest form, the gravity equation for trade states that exports from country i to country j, denoted by T_{ij}, are proportional to the product of the two countries' GDPs, denoted by Y_i and Y_j, and inversely proportional to their distance, D_{ij}, broadly construed to include all factors that might create trade resistance. That is,

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3}, \tag{1}$$

where α_0, α_1, α_2, and α_3 are parameters to be estimated.

The analogy between trade and the physical force of gravity, however, clashes with the observation that there is no set of parameters for which equation (1) will hold exactly for an arbitrary set of observations. To account for deviations from the theory, stochastic versions of the equation are used in empirical studies. Typically, the stochastic version of the gravity equation has the form

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \eta_{ij}, \tag{2}$$

[2] See, for example, Anderson (1979), Helpman and Krugman (1985), Bergstrand (1985), Davis (1995), Deardoff (1998), and Anderson and van Wincoop (2003). A feature common to these models is that they all assume complete specialization: each good is produced in only one country. However, Haveman and Hummels (2001), Feenstra, Markusen, and Rose (1999), and Eaton and Kortum (2001) derive the gravity equation without relying on complete specialization. Examples of empirical studies framed on the gravity equation include the evaluation of trade protection (e.g., Harrigan, 1993), regional trade agreements (e.g., Frankel, Stein, and Wei, 1995; Frankel, 1997), exchange rate variability (e.g., Frankel and Wei, 1993; Eichengreen and Irwin, 1995), and currency unions (e.g., Rose, 2000; Frankel and Rose, 2002; and Tenreyro and Barro, 2002). See also the various studies on "border-effects" influencing the patterns of intranational and international trade, including McCallum (1995), and Anderson and van Wincoop (2003), among others.

where η_{ij} is an error term with $E(\eta_{ij}|Y_i, Y_j, D_{ij}) = 1$, assumed to be statistically independent of the regressors, leading to

$$E(T_{ij}|Y_i, Y_j, D_{ij}) = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3}.$$

There is a long tradition in the trade literature of log-linearizing (2) and estimating the parameters of interest by least squares using the equation

$$\ln(T_{ij}) = \ln(\alpha_0) + \alpha_1 \ln(Y_i) + \alpha_2 \ln(Y_j) + \alpha_3 \ln(D_{ij}) + \ln(\eta_{ij}). \tag{3}$$

The validity of this procedure depends critically on the assumption that η_{ij}, and therefore $\ln(\eta_{ij})$, are statistically independent of the regressors. In (2) the parameters α_1, α_2, and α_3 are the elasticities of $E(T_{ij}|Y_i, Y_j, D_{ij})$ (not of T_{ij}) with respect to Y_i, Y_j and D_{ij}. Due to Jensen's inequality, only under very restrictive assumptions on the distribution of the error terms will it also be possible to interpret these parameters as the derivatives of $E(\ln(T_{ij})|Y_i, Y_j, D_{ij})$ with respect to the logarithms of the conditioning variables, which is how they appear in (3).

There is no reason to assume that the variance of η_{ij} will be independent of the countries' GDPs and of the various measures of distance between them. In other words, the error term η_{ij} will, in general, be heteroskedastic. This implies that the standard estimation method will be consistent for the derivatives of $E(\ln(T_{ij})|Y_i, Y_j, D_{ij})$ with respect to the logged regressors, but generally these parameters are different from the elasticities of $E(T_{ij}|Y_i, Y_j, D_{ij})$ with respect to Y_i, Y_j and D_{ij}.

To see why this is so, notice that the expected value of the logarithm of a random variable depends both on its mean and on higher-order moments of the distribution. Hence, whenever the variance of the error term η_{ij} in equation (1) depends on Y_i, Y_j, or D_{ij}, the expected value of $\ln(\eta_{ij})$ will also depend on the regressors, violating the condition for consistency of OLS.[3] Clearly, in this instance, homoskedasticity is critical

[3]As an illustration, consider the case in which η_{ij} follows a log-normal distribution, with $E(\eta_{ij}|Y_i, Y_j, D_{ij}) = 1$ and variance $\sigma_{ij}^2 = f(Y_i, Y_j, D_{ij})$. The error term in the log-linearized representation will then follow a normal distribution, with $E\left[\ln(\eta_{ij})|Y_i, Y_j, D_{ij}\right] = -\frac{1}{2}\ln(1 + \sigma_{ij}^2)$, which is also a function of the covariates.

not only for the efficiency of the estimator, but also for its consistency for the parameters of interest.

A related problem with the analogy between Newtonian gravity and trade is that gravitational force can be very small, but never zero, whereas trade between several pairs of countries is literally zero. In many cases, these zeros occur simply because some pairs of countries did not trade in a given period. For example, it would not be surprising to find that Tajikistan and Togo did not trade in a certain year.[4] These zero observations pose no problem at all for the estimation of gravity equations in their multiplicative form. In contrast, the existence of observations for which the dependent variable is zero creates an additional problem for the use of the log-linear form of the gravity equation. Several methods have been developed to deal with this problem (see Frankel, 1997, for a description of the various procedures). The approach followed by the large majority of empirical studies is simply to drop the pairs with zero trade from the data set and estimate the log-linear form by OLS. Rather that throwing away the observations with $y_i = 0$, some authors estimate the model using $y_i + 1$ as the dependent variable. Both these procedures, however, lead to inconsistent estimators of β. The severity of these inconsistencies will depend on the particular characteristics of the sample and model used, but there is no reason to believe that they will be negligible. The consequences of these procedures are discussed latter.

Zeroes may also be the result of rounding errors.[5] If trade is measured in thousands of dollars, it is possible that for pairs of countries for which bilateral trade did not reach a minimum value, say $500, the value of trade is registered as zero. If these rounded-down observations were partially compensated by rounded-up ones, the overall effect of these errors would be relatively minor. However, the rounding down is more likely to occur for

[4] The absence of trade between small and distant countries might be explained, among other factors, by large variable costs (e.g., bricks are too costly to transport) or large fixed costs (e.g., information on foreign markets). At the aggregate level, these costs can be best proxied by the various measures of distance and size entering the gravity equation.

[5] Trade data can suffer from many other forms of errors, as described in Feenstra, Lipsey, and Bowen (1997).

small or distant countries and, therefore, the probability of rounding down will depend on the value of the covariates, leading to the inconsistency of the estimators. Finally, the zeros can just be missing observations which are wrongly recorded as zero. This problem is more likely to occur when small countries are considered and, again, measurement error will depend on the covariates, leading to inconsistency.

3. Constant-Elasticity Models

Despite their immense popularity, there are still important econometric flaws in empirical studies involving gravity equations. These flaws are not exclusive of this literature, and extend to many areas where constant-elasticity models are used. This section examines how the deterministic multiplicative models suggested by economic theory can be used in empirical studies.

In their non-stochastic form, the relationship between the multiplicative constant-elasticity model and its log-linear additive formulation is trivial. The problem, of course, is that economic relations do not hold with the accuracy of the physical laws. All that can be expected is that they hold on average. Indeed, here we interpret economic models like the gravity equation as the expected value of the variable of interest, $y \geq 0$, for a given value of the explanatory variables, x (see Goldberger, 1991, p. 5). That is, if economic theory suggests that y and x are linked by a constant-elasticity model of the form $y_i = \exp(x_i\beta)$, the function $\exp(x_i\beta)$ is interpreted as the conditional expectation of y_i given x, denoted $E[y_i|x]$.[6] For example, using the notation in the previous section, the multiplicative gravity relationship can be written as the exponential function

[6]Notice that if $\exp(x_i\beta)$ is interpreted as describing the conditional median of y_i (or other conditional quantile) rather than the conditional expectation, estimates of the elasticities of interest can be obtained estimating the log-linear model using the appropriate quantile regression estimator (Koenker and Bassett, 1978). However, interpreting $\exp(x_i\beta)$ as a conditional median is problematic when y_i has a large mass of zero observations, like in trade data. Indeed, in this case the conditional median of y_i will be a discontinuous function of the regressors, which is generally not compatible with the standard economic theory.

$\exp\left[\ln\left(\alpha_0\right)+\alpha_1\ln\left(Y_i\right)+\alpha_2\ln\left(Y_j\right)+\alpha_3\ln\left(D_{ij}\right)\right]$, which is interpreted as the conditional expectation $E(T_{ij}|Y_i, Y_j, D_{ij})$.

Since the relation $y_i = \exp\left(x_i\beta\right)$ holds on average, but not for each i, there is an error term associated with each observation, which is defined as $\varepsilon_i = y_i - E\left[y_i|x\right]$.[7] Therefore, the stochastic model can be formulated as

$$y_i = \exp\left(x_i\beta\right) + \varepsilon_i, \tag{4}$$

with $y_i \geq 0$ and $E\left[\varepsilon_i|x\right] = 0$.

As we mentioned before, the standard practice of log-linearizing equation (4) and estimating β by OLS is inappropriate for a number of reasons. First of all, y_i can be zero, in which case log-linearization is unfeasible. Second, even if all observations of y_i are strictly positive, the expected value of the log-linearized error will in general depend on the covariates and hence OLS will be inconsistent. To see the point more clearly, notice that equation (4) can be expressed as

$$y_i = \exp\left(x_i\beta\right)\eta_i,$$

with $\eta_i = 1 + \varepsilon_i/\exp\left(x_i\beta\right)$ and $E\left[\eta_i|x\right] = 1$. Assuming for the moment that y_i is positive, the model can be made linear in the parameters by taking logarithms of both sides of the equation, leading to

$$\ln\left(y_i\right) = x_i\beta + \ln\left(\eta_i\right). \tag{5}$$

To obtain a consistent estimator of the slope parameters in equation (4) estimating (5) by OLS, it is necessary that $E\left[\ln\left(\eta_i\right)|\, x\right]$ does not depend on x_i.[8] Since $\eta_i = 1 + \varepsilon_i/\exp\left(x_i\beta\right)$, this condition is met only if ε_i can be written as $\varepsilon_i = \exp\left(x_i\beta\right)\nu_i$, where ν_i is a random variable statistically independent of x_i. In this case, $\eta_i = \nu_i$ and therefore is statistically independent of x_i, implying that $E\left[\ln\left(\eta_i\right)|\, x\right]$ is constant. Thus, the log-linear representation of the constant-elasticity model is only useful as a device to estimate the parameters of interest under very specific conditions on the error term.

[7] Whether the error term enters additively or multiplicatively is irrelevant for our purposes, as explained below.

[8] Consistent estimation of the intercept would also require $E\left[\ln\left(\eta_i\right)|\, x\right] = 0$.

When η_i is statistically independent of x_i, the conditional variance of y_i (and ε_i) is proportional to $\exp(2x_i\beta)$. Although economic theory generally does not provide any information on the variance of ε_i, we can infer some of its properties from the characteristics of data. Because y_i is non-negative, when $E[y_i|x]$ approaches zero, the probability of y_i being positive must also approach zero. This implies that $V[y_i|x]$, the conditional variance of y_i, tends to vanish as $E[y_i|x]$ passes to zero.[9] On the other hand, when the expected value y is far away from its lower bound, it is possible to observe large deviations from the conditional mean in either direction, leading to greater dispersion. Thus, in practice, ε_i will generally be heteroskedastic and its variance will depend on $\exp(x_i\beta)$, but there is no reason to assume that $V[y_i|x]$ is proportional to $\exp(2x_i\beta)$. Therefore, in general, regressing $\ln(y_i)$ on x_i by OLS will lead to inconsistent estimates of β.

It may be surprising that the pattern of heteroskedasticity and, indeed, the form of all higher-order moments of the conditional distribution of the error term can affect the consistency of an estimator, rather than just its efficiency. The reason is that the non-linear transformation of the dependent variable in equation (5) changes the properties of the error term in a non-trivial way since the conditional expectation of $\ln(\eta_i)$ depends on the shape of the conditional distribution of η_i. Hence, unless very strong restrictions are imposed on the form of this distribution, it is not possible to recover information about the conditional expectation of y_i from the conditional mean of $\ln(y_i)$ simply because $\ln(\eta_i)$ is correlated with the regressors. Nevertheless, estimating (5) by OLS will produce consistent estimates of the parameters of $E[\ln(y_i)|\, x]$. The problem is that these parameters may not permit identification of the parameters of $E[y_i|x]$.

In short, even assuming that all observations on y_i are positive, it is not advisable to estimate β from the log-linear model. Instead, the non-linear model has to be estimated.

[9] In the case of trade data, when $E[y_i|x]$ is close to its lower bound (i.e., for pairs of small and distant countries), it is unlikely that large values of trade are observed since they cannot be offset by equally large deviations in the opposite direction simply because trade cannot be negative. Therefore, for these observations, dispersion around the mean tends to be small.

3.1 Estimation

Although most empirical studies use the log-linear form of the constant-elasticity model, some authors (see Frankel and Wei, 1993, for an example in the international trade literature) have estimated multiplicative models using non-linear least squares (NLS), which is an asymptotically valid estimator for (4). However, the NLS estimator can be very inefficient in this context, as it ignores the heteroskedasticity that, as discussed before, is characteristic of this type of model.

The NLS estimator of β is defined by

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} \left[y_i - \exp\left(x_i b\right)\right]^2,$$

which implies the following set of first order conditions:

$$\sum_{i=1}^{n} \left[y_i - \exp\left(x_i\hat{\beta}\right)\right] x_i \exp\left(x_i\hat{\beta}\right) = 0. \quad (6)$$

These equations give more weight to observations where $\exp\left(x_i\hat{\beta}\right)$ is large because that is where the curvature of the conditional expectation is more pronounced. However, these are generally also the observations with larger variance, which implies that NLS gives more weight to noisier observations. Thus, this estimator may be very inefficient, depending heavily on a small number of observations.

If the form of $V\left[y_i|x\right]$ was known, this problem could be eliminated using a weighted-NLS estimator. However, in practice, all we know about $V\left[y_i|x\right]$ is that, in general, it goes to zero as $E\left[y_i|x\right]$ passes to zero. Therefore, an optimal weighted-NLS estimator cannot be used without further information on the distribution of the errors. In principle, this problem can be tackled by estimating the multiplicative model using a consistent estimator, and then obtaining the appropriate weights estimating the skedastic function non-parametrically, as suggested by Delgado (1992) and Delgado and Kniesner (1997). However, this nonparametric generalized least squares estimator is rather cumbersome to implement, especially if the model has a large number of regressors. Moreover, the choice of the first round estimator is an open question as the NLS estimator may be a poor starting point due to its considerable inefficiency. Therefore, the nonparametric generalized least squares estimator is not appropriate to use as a work-horse for routine estimation of multiplicative models.[10] Indeed, what is needed is an estimator that is consistent and reasonably efficient under a wide range of heteroskedasticity patterns and is also simple to implement.

A possible way of obtaining an estimator that is more efficient than the standard NLS without the need to use nonparametric regression is to follow McCullagh and Nelder (1989) and estimate the parameters of interest using a pseudo-maximum likelihood (PML) estimator based on some assumption on the functional form of $V[y_i|x]$.[11] Among the many possible specifications, the hypothesis that the conditional variance is proportional to the conditional mean is particularly appealing. Indeed, under this assumption $E[y_i|x] = \exp(x_i\beta) \propto V[y_i|x]$, and β can be estimated by solving the following set of first order conditions:

$$\sum_{i=1}^{n} \left[y_i - \exp\left(x_i\tilde{\beta}\right)\right] x_i = 0. \tag{7}$$

Comparing equations (6) and (7), it is clear that, unlike the NLS estimator, which is a PML estimator obtained assuming that $V[y_i|x]$ is constant, the PML estimator based on (7) gives the same weight to all observations, rather than emphasizing those for which $\exp(x_i\beta)$ is large. This is because, under the assumption that $E[y_i|x] \propto V[y_i|x]$, all observations have the same information on the parameters of interest as the additional information on the curvature of the conditional mean coming from observations with large $\exp(x_i\beta)$ is offset by their larger variance. Of course, this estimator may not be optimal, but without further information on the pattern of heteroskedasticity, it seems natural to give the same weight to all observations as in this way there is no risk of giving extra

[10] A nonparametric generalized least squares estimator can also be used to estimate linear models in presence of heteroskedasticity of unknown form (Robinson, 1987). However, despite having been proposed more than 15 years ago, this estimator has never been adopted as a standard tool by empirical researchers, who generally prefer the simplicity of the inefficient OLS, with an appropriate covariance matrix.

[11] See also Manning and Mullahy (2001). A related estimator is proposed by Papke and Wooldridge (1996) for the estimation of models for fractional data.

weight to the wrong observations.[12] Even if $E[y_i|x]$ is not proportional to $V[y_i|x]$, the PML estimator based on (7) is likely to be more efficient than the NLS estimator when the heteroskedasticity increases with the conditional mean.

The estimator defined by (7) is numerically equal to the Poisson pseudo-maximum likelihood estimator which is often used for count data[13] and can be obtained maximizing

$$\tilde{\beta} = \arg\max_b \sum_{i=1}^{n} \left\{ y_i \left(x_i b \right) - \exp \left(x_i b \right) \right\}.$$

The form of (7) makes clear that all that is needed for this estimator to be consistent is the correct specification of the conditional mean, i.e., $E[y_i|x] = \exp(x_i\beta)$. Therefore, the data do not have to be Poisson at all and, what is more important, y_i does not even have to be an integer, for the estimator based on the Poisson likelihood function to be consistent. This is the well-known pseudo-maximum likelihood result first noted by Gourieroux, Monfort and Trognon (1984).

The implementation of the Poisson PML estimator is straightforward since there are standard econometric programs with commands that permit the estimation of Poisson regression, even when the dependent variables are not integers. Because the assumption $V[y_i|x] \propto E[y_i|x]$ is unlikely to hold, this estimator does not fully account for the heteroskedasticity in the model and all inference has to be based on an Eicker-White (Eicker, 1963; and White, 1980) robust covariance matrix estimator.

Of course, if it was known that $V[y_i|x]$ is a function of higher powers of $E[y_i|x]$, a more efficient estimator could be obtained down-weighing even more the observations with large conditional mean. An example of such estimator is the gamma pseudo-maximum likelihood estimator studied by Manning and Mullahy (2001) which, like the log-linearized model, assumes that $V[y_i|x]$ is proportional to $E[y_i|x]^2$. The first order conditions for

[12] The same strategy is implicitly used by Papke and Wooldridge (1996) in their pseudo-maximum estimator for fractional data models.

[13] See Cameron and Trivedi (1998) and Winkelmann (2000) for more details on the Poisson regression and on more general models for count data.

the gamma PML estimator are given by

$$\sum_{i=1}^{n}\left[y_i-\exp\left(x_i\breve{\beta}\right)\right]\frac{x_i}{\exp\left(x_i\breve{\beta}\right)}=0.$$

In the case of trade data, however, this estimator may have an important drawback. Trade data for larger countries (as gauged by GDP per capita) tend to be of higher quality (see Frankel and Wei, 1993, and Frankel 1997); hence, models assuming that $V[y_i|x]$ is a function of higher powers of $E[y_i|x]$ might give excessive weight to the observations that are more prone to measurement errors.[14] Therefore, the Poisson regression emerges as a reasonable compromise, giving less weight to the observations with larger variance than the standard NLS estimator, without giving too much weight to observations more prone to contamination by measurement error and less informative about the curvature of $E[y_i|x]$.

3.2 Testing

In this section we consider tests for the particular pattern of heteroskedasticity assumed by PML estimators, focusing on the Poisson estimator. Although PML estimators are consistent even when the variance function is misspecified, the researcher can use these tests to check if a different PML estimator would be more appropriate and to decide whether or not the use of a nonparametric estimator of the variance is warranted.

Manning and Mullahy (2001) suggested that if

$$V[y_i|x]=\lambda_0 E[y_i|x]^{\lambda_1}, \tag{8}$$

the choice of the appropriate PML estimator can be based on a Park-type regression (Park, 1966). Their approach is based on the idea that if (8) holds and an initial consistent

[14] Frankel and Wei (1993) and Frankel (1997) suggest that larger countries should be given more weight in the estimation of gravity equations. This would be appropriate if the errors in the model were just the result of measurement errors in the dependent variable. However, if it is accepted that the gravity equation does not hold exactly, measurement errors account for only part of the dispersion of trade data around the gravity equation.

estimate of $E[y_i|x]$ is available, then λ_1 can be consistently estimated using an appropriate auxiliary regression. Specifically, following Park (1966), Manning and Mullahy (2001) suggest that λ_1 can be estimated using the auxiliary model

$$\ln(y_i - \breve{y}_i)^2 = \ln(\lambda_0) + \lambda_1 \ln(\breve{y}_i) + \upsilon_i, \tag{9}$$

where $\breve{y}_i$ denotes the estimated value of $E[y_i|x]$. Unfortunately, as the discussion in the previous sections should have made clear, this approach based on the log-linearization of (8) is valid only under very restrictive conditions on the conditional distribution of y_i. However, it is easy to see that this procedure is valid when the constant-elasticity model can be consistently estimated in the log-linear form. Therefore, using (9) a test for $H_0 : \lambda_1 = 2$ based on a non-robust covariance estimator provides a check for the adequacy of the estimator based on the log-linear model.

A more robust alternative, which is mentioned by Manning and Mullahy (2001) in a footnote, is to estimate λ_1 from

$$(y_i - \breve{y}_i)^2 = \lambda_0 (\breve{y}_i)^{\lambda_1} + \xi_i, \tag{10}$$

using an appropriate PML estimator. The approach based on (10) is asymptotically valid and inference about λ_1 can be based on the usual Eicker-White robust covariance matrix estimator. For example, the hypothesis that $V[y_i|x]$ is proportional to $E[y_i|x]$ is accepted if the appropriate confidence interval for λ_1 contains 1. However, if the purpose is to test the adequacy of a particular value of λ_1, a slightly simpler method based on the Gauss-Newton regression (see Davidson and MacKinnon, 1993) is available.

Specifically, to check the adequacy of the Poisson PML for which $\lambda_1 = 1$ and $\breve{y}_i = \exp\left(x_i\tilde{\beta}\right)$, (10) can be expanded in a Taylor series around $\lambda_1 = 1$, leading to

$$(y_i - \breve{y}_i)^2 = \lambda_0 \breve{y}_i + \lambda_0 (\lambda_1 - 1) \ln(\breve{y}_i)\, \breve{y}_i + \xi_i.$$

Now, the hypothesis that $V[y_i|x] \propto E[y_i|x]$ can be tested against (8) simply by checking the significance of the parameter $\lambda_0 (\lambda_1 - 1)$. Because the error term ξ_i is unlikely to be homoskedastic, the estimation of the Gauss-Newton regression should be performed using weighted least squares. Assuming that in (10) the variance is also proportional to the mean, the appropriate weights are given by $\exp\left(-x_i\tilde{\beta}\right)$ and therefore the test can be performed by estimating

$$\left(y_i - \breve{y}_i\right)^2 \Big/ \sqrt{\breve{y}_i} = \lambda_0 \sqrt{\breve{y}_i} + \lambda_0 \left(\lambda_1 - 1\right) \ln\left(\breve{y}_i\right) \sqrt{\breve{y}_i} + \xi_i^* \tag{11}$$

by OLS and testing the statistical significance of $\lambda_0 \left(\lambda_1 - 1\right)$ using a Eicker-White robust covariance matrix estimator.[15]

In the next section, a small simulation is used to study the Gauss-Newton regression test for the hypothesis that $V\left[y_i|x\right] \propto E\left[y_i|x\right]$, as well as the Park-type test for the hypothesis that the constant-elasticity model can be consistently estimated in the log-linear form.

4. A simulation study

This section reports the results of a small simulation study designed to assess the performance of different methods to estimate constant-elasticity models in the presence of heteroskedasticity and rounding errors. As a by-product, we also obtained some evidence on the finite sample performance of the specification tests presented above. These experiments are centered around the following multiplicative model:

$$E\left[y_i|x\right] = \mu\left(x_i\beta\right) = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right), \quad i = 1, \ldots, 1000. \tag{12}$$

Since, in practice, regression models often include a mixture of continuous and dummy variables, we replicate this feature in our experiments: x_{1i} is drawn from a standard normal and x_2 is a binary dummy variable that equals 1 with a probability of 0.4.[16] The

[15] Notice that to test $V\left[y_i|x\right] \propto E\left[y_i|x\right]$ against alternatives of the form $V\left[y_i|x\right] = \lambda_0 \exp\left(x_i\left(\beta + \lambda\right)\right)$ the appropriate auxiliary regression would be

$$\left(y_i - \breve{y}_i\right)^2 \Big/ \sqrt{\breve{y}_i} = \lambda_0 \sqrt{\breve{y}_i} + \lambda_0 \lambda x_i \sqrt{\breve{y}_i} + \xi_i^*,$$

and the test could be performed by checking the joint significance of the elements of $\lambda_0\lambda$. If the model includes a constant, one of the regressores in the auxiliary regression is redundant and should be dropped.

[16] For example, in gravity equations, continuous variables (which are all strictly positive) include income and geographical distance. In equation (12), x_1 can be interpreted as (the logarithm of) one of these

two covariates are independent and a new set of observations of all variables is generated in each replication using $\beta_0 = 0$, $\beta_1 = \beta_2 = 1$. Data on y are generated as

$$y_i = \mu(x_i\beta)\,\eta_i, \tag{13}$$

where η_i is a log-normal random variable with mean 1 and variance σ_i^2. As noted before, the slope parameters in (12) can be estimated using the log-linear form of the model only when σ_i^2 is constant. That is, when $V[y_i|x]$ is proportional to $\mu(x_i\beta)^2$.

In these experiments we analyzed PML estimators of the multiplicative model and different OLS estimators of the log-linearized model. The consistent PML estimators studied were: non-linear least squares, NLS; gamma pseudo-maximum likelihood, GPML; and the Poisson pseudo-maximum likelihood, PPML. Besides these estimators, we also considered the standard OLS estimator of the log-linear model, OLS; the OLS estimator for the model were the dependent variable is $y_i + 1$, OLS $(y + 1)$; and truncated OLS estimators to be discussed below.[17]

To assess the performance of the estimators under different patterns of heteroskedasticity, we considered the four following specifications of σ_i^2:

Case 1: $\sigma_i^2 = \mu(x_i\beta)^{-2}$; $V[y_i|x] = 1$;

Case 2: $\sigma_i^2 = \mu(x_i\beta)^{-1}$; $V[y_i|x] = \mu(x_i\beta)$;

Case 3: $\sigma_i^2 = 1$; $V[y_i|x] = \mu(x_i\beta)^2$;

Case 4: $\sigma_i^2 = \exp(x_{2i}) + \mu(x_i\beta)^{-1}$; $V[y_i|x] = \mu(x_i\beta) + \exp(x_{2i})\,\mu(x_i\beta)^2$.

In Case 1 the variance of ε_i is constant, implying that the NLS estimator is optimal. Although, as argued before, this case is unrealistic for models of bilateral trade, it is included in the simulations for completeness. In Case 2, the conditional variance of

variables. Examples of binary variables include dummies for free-trade agreements, common language, colonial ties, contiguity and access to land.

[17]We also studied the performance of Tobit models (with constant and estimated cut-off points as well as the semi-logarithmic Tobit) finding very poor results.

y_i equals its conditional mean, as in the Poisson distribution. The pseudo-likelihood estimator based on the Poisson distribution is optimal in this situation. Case 3 is the special case in which OLS estimation of the log-linear model is consistent for the slope parameters of (12). Moreover, in this case the log-linear model not only corrects the heteroskedasticity in the data, but, because η_i is log-normal, it is also the maximum likelihood estimator. The GPML is the optimal PML estimator in this case, but it should be outperformed by the true maximum likelihood estimator. Finally, Case 4 is the only one in which the conditional variance does not depend exclusively on the mean. The variance is a quadratic function of the mean, as in Case 3, but it is not proportional to the square of the mean.

We carried out two sets of experiments. The first set was aimed at studying the performance of the estimators of the multiplicative and the log-linear models under different patterns of heteroskedasticity. In order to study the effect of the truncation on the performance of the OLS, and given that this data generating mechanism does not produce observations with $y_i = 0$, the log-linear model was also estimated using only the observations for which $y_i > 0.5$, OLS ($y > 0.5$). This reduces the sample size by about 25% to 35%, depending on the pattern of heteroskedasticity.

The second set of experiments studied the estimators' performance in the presence of rounding errors in the dependent variable. For that purpose, a new random variable was generated rounding to the nearest integer the values y_i obtained in the first set of simulations. This procedure mimics the rounding errors in official statistics and generates a large number of zeros, a typical feature of trade data. Because the model considered here generates a large proportion of observations close to zero, rounding down is much more frequent than rounding up. As the probability of rounding up or down depends on the covariates, this procedure will necessarily bias the estimates, as discussed before. The purpose of the study is to gauge the magnitude of these biases. Naturally, the log-linear model cannot be estimated in these conditions because the dependent variable equals zero for some observations. Following what is the usual practice in these circumstances, the truncated OLS estimation of the log-linear model was performed dropping the observations for which the dependent variable equals zero. Notice that the observations discarded with this procedure are exactly the same that are discarded by OLS ($y > 0.5$) in the first set of experiments. Therefore, this estimator is also denoted OLS ($y > 0.5$).

The results of the first set of experiments are summarized in Table 1, which displays the biases and standard errors of the different estimators of β obtained with $10,000$ replicas of the simulation procedure described above, without measurement error. Table 2 contains the results for the experiments in which y_i is rounded to the nearest integer.

As expected, OLS only performs well in Case 3.[18] In all other cases this estimator is clearly inadequate because, despite its low dispersion, it is often badly biased. Moreover, the sign and magnitude of the bias vary considerably. Therefore, even when the dependent variable is strictly positive, estimation of constant elasticity models using the log-linearized model cannot generally be recommended. As for the modifications of the least squares estimator designed to deal with the zeros of the dependent variable, OLS ($y + 1$) and OLS ($y > 0.5$), their performance is also very disappointing. These results clearly emphasize the need to use adequate methods to deal with the zeros in the data and raise serious doubts about the validity of the results obtained using the traditional estimators based on the log-linear model. Overall, except under very special circumstances, estimation based on the log-linear model cannot be recommended.

One remarkable result of this set of experiments is the extremely poor performance of the NLS estimator. Indeed, when the heteroskedasticity is more severe (cases 3 and 4) this estimator, despite being consistent, leads to very poor results because of its erratic behavior.[19] Therefore, it is clear that the loss of efficiency caused by some of the forms of heteroskedasticity considered in these experiments is strong enough to render this estimator useless in practice.

---

[18] Notice that only the results for β_1 and β_2 are of interest in this case since it is well know that the estimator of β_0 is inconsistent.

[19] Manning and Mullahy (2001) report similar results.

Table 1: Simulation results under different forms of heteroskedasticity

	β_0		β_1		β_2	
Estimator:	Bias	S. Error	Bias	S. Error	Bias	S. Error
Case 1:	$V[y_i\|x] = 1$					
PPML	−0.00018	0.035	−0.00004	0.016	0.00009	0.027
NLS	0.00009	0.021	−0.00006	0.008	−0.00003	0.017
GPML	−0.00771	0.073	0.01276	0.068	0.00754	0.082
OLS	−0.53337	0.043	0.39008	0.039	0.35568	0.054
OLS $(y > 0.5)$	0.17812	0.037	−0.16402	0.027	−0.15487	0.038
OLS $(y + 1)$	0.74497	0.015	−0.40237	0.014	−0.37683	0.022
Case 2:	$V[y_i\|x] = \mu(x_i\beta)$					
PPML	−0.00035	0.036	−0.00011	0.019	0.00009	0.039
NLS	−0.00247	0.069	0.00046	0.033	0.00066	0.057
GPML	−0.00246	0.049	0.00376	0.043	0.00211	0.062
OLS	−0.40271	0.036	0.21076	0.030	0.19960	0.049
OLS $(y > 0.5)$	0.15061	0.036	−0.17868	0.026	−0.17220	0.043
OLS $(y + 1)$	0.73925	0.016	−0.42371	0.015	−0.39931	0.025
Case 3:	$V[y_i\|x] = \mu(x_i\beta)^2$					
PPML	−0.00215	0.104	−0.00526	0.091	−0.00228	0.130
NLS	−0.79648	9.234	0.23539	3.066	0.07323	1.521
GPML	−0.00203	0.052	−0.00047	0.041	−0.00029	0.083
OLS	−0.49964	0.040	0.00015	0.032	−0.00003	0.064
OLS $(y > 0.5)$	0.17779	0.041	−0.34480	0.039	−0.34614	0.064
OLS $(y + 1)$	0.67525	0.020	−0.51804	0.021	−0.50000	0.038
Case 4:	$V[y_i\|x] = \mu(x_i\beta) + \exp(x_{2i})\,\mu(x_i\beta)^2$					
PPML	−0.00053	0.118	−0.00696	0.103	−0.00647	0.144
NLS	−1.15429	21.850	0.35139	7.516	0.08801	1.827
GPML	−0.00342	0.063	0.00322	0.057	−0.00137	0.083
OLS	−0.60041	0.044	0.13270	0.039	−0.12542	0.075
OLS $(y > 0.5)$	−0.39217	0.044	−0.41391	0.042	−0.41391	0.070
OLS $(y + 1)$	0.67743	0.020	−0.51440	0.021	−0.58087	0.041

Table 2: Simulation results under heteroskedasticity and rounding error

	β_0		β_1		β_2	
Estimator:	Bias	S. Error	Bias	S. Error	Bias	S. Error
Case 1:	$V[y_i\|x]=1$					
PPML	-0.04161	0.038	0.01886	0.017	0.02032	0.029
NLS	-0.00638	0.022	0.00195	0.008	0.00274	0.018
GPML	-0.11466	0.095	0.10946	0.096	0.09338	0.108
OLS $(y>0.5)$	0.26861	0.034	-0.22121	0.026	-0.21339	0.036
OLS $(y+1)$	0.70142	0.017	-0.37752	0.015	-0.34997	0.024
Case 2:	$V[y_i\|x]=\mu(x_i\beta)$					
PPML	-0.04800	0.040	0.02190	0.020	0.02334	0.041
NLS	-0.00933	0.069	0.00262	0.033	0.00360	0.057
GPML	-0.13724	0.072	0.13243	0.073	0.11331	0.087
OLS $(y>0.5)$	0.25779	0.033	-0.24405	0.026	-0.23889	0.040
OLS $(y+1)$	0.68960	0.018	-0.39401	0.016	-0.36806	0.028
Case 3:	$V[y_i\|x]=\mu(x_i\beta)^2$					
PPML	-0.06562	0.109	0.02332	0.091	0.02812	0.133
NLS	-0.80964	9.269	0.23959	3.082	0.07852	1.521
GPML	-0.17807	0.073	0.17134	0.068	0.14442	0.104
OLS $(y>0.5)$	0.30765	0.037	-0.41006	0.037	-0.41200	0.060
OLS $(y+1)$	0.61141	0.022	-0.48564	0.022	-0.46597	0.040
Case 4:	$V[y_i\|x]=\mu(x_i\beta)+\exp(x_{2i})\,\mu(x_i\beta)^2$					
PPML	-0.05933	0.122	0.02027	0.104	0.01856	0.146
NLS	-1.16956	21.861	0.35672	7.521	0.09239	1.829
GPML	-0.14256	0.085	0.12831	0.085	0.10245	0.129
OLS $(y>0.5)$	0.35122	0.039	-0.45188	0.040	-0.46173	0.066
OLS $(y+1)$	0.61930	0.022	-0.48627	0.022	-0.56039	0.044

In the first set of experiments, the results of the gamma PML estimator are very good. Indeed, when no measurement error is present, the biases and standard errors of the GPML estimator are always among the lowest. However, this estimator is very sensitive to the form of measurement error considered in the second set of experiments, consistently leading to sizable biases. These results, like those of the NLS, clearly illustrate the danger of using a PML estimator that gives extra weight to the noisier observations.

As for the performance of the Poisson PML estimator, the results are very encouraging. In fact, when no rounding error is present, its performance is reasonably good in all cases. Moreover, although some loss of efficiency is noticeable as one moves away from Case 2, in which it is an optimal estimator, the biases of the PPML are always small.[20] Moreover, the results obtained with rounded data suggest that the Poisson based PML estimator is relatively robust to this form of measurement error of the dependent variable. Indeed, the bias introduced by the rounding-off errors in the dependent variable is relatively small and, in some cases, it even compensates the bias found in the first set of experiments. Therefore, because it is simple to implement and reliable under a wide variety of situations, the Poisson PML estimator has the essential characteristics needed to make it the new work-horse for the estimation of constant-elasticity models.

Obviously, the sign and magnitude of the bias of the estimators studied here depend on the particular specification considered. Therefore, the results of these experiments cannot serve as an indicator of what can be expected in other situations. However, it is clear that, apart from the Poisson PML method, all estimators are potentially very misleading.

These experiments were also used to study the finite sample performance of the Gauss-Newton regression (GNR) test for the adequacy of the Poisson PML based on (11) and of the Park test advocated by Manning and Mullahy (2001), which, as explained above, is only valid to check for the adequacy of the estimator based on the log-linear model.[21] Given that the Poisson PML estimator is the only estimator with a reasonable behavior under all the cases considered, these tests were performed using residuals and estimates of $\mu(x_i\beta)$ from the Poisson regression. Table 3 contains the rejection frequencies at the 5 percent nominal level of the two tests for the four cases considered in the two sets of experiments. In this table the rejection frequencies under the null hypothesis are given in bold.

[20] These results are in line with those reported by Manning and Mullahy (2001).

[21] To illustrate the pitfalls of the procedure suggested by Manning and Mullahy (2001) we note that the means of the estimates of λ_1 obtained using (9) in cases 1, 2 and 3 (without measurement error) were 0.58955, 1.29821 and 1.98705, whereas the true values of λ_1 in these cases are, respectively, 0, 1 and 2.

Table 3: Rejection frequencies at the 5% level for the two specification tests

Without Measurement Error				
Test	Case 1	Case 2	Case 3	Case 4
GNR	0.91980	**0.05430**	0.58110	0.49100
Park	1.00000	1.00000	**0.06680**	0.40810
With Measurement Error				
Test	Case 1	Case 2	Case 3	Case 4
GNR	0.91740	0.14980	0.57170	0.47580
Park	1.00000	1.00000	1.00000	1.00000

Since both tests have adequate behavior under the null and reveal reasonable power against a wide range of alternatives, the results suggest that these tests are important tools to assess the adequacy of the standard OLS estimator of the log-linear model and of the proposed Poisson PML estimator.

5. The gravity equation

In this section, we use the Poisson PML estimator to quantitatively assess the determinants of bilateral trade flows, uncovering significant differences in the roles of various measures of size and distance from those predicted by the "logarithmic tradition." We focus particular attention on the role of trade agreements, since this policy instrument has been the object of intense debate (see, for example, Frankel, 1997, and Bhagwati and Panagariya, 1996).

In recent years, the estimation of gravity equations using panel data has become increasingly popular, and the literature has focused on the propper way to include country-specifinc effects in these regressions (see Mátyás, 1997 and 1998; Egger, 2000 and 2002; Cheng and Wall, 2002 and Egger and Pfaffermayr, 2003). However, since the vast majority of empirical applications of constant-elasticity models does not use panel data, here we illustrate the use of the proposed Poisson PML estimator using a single cross-section.[22] With this type of data, the inclusion of country-specific fixed effects is unappealing since

[22] Notice that this estimator can be easily applied with panel data.

it does not permit the estimation of some parameters of primary interest, like the income elasticities. Therefore, we estimated a model without fixed-effects which serves our main purpose, which is to highlight the large difference between the results obtained with the Poisson PML estimator and those resulting from the use of the standard log-linearized model.[23]

5.1 The data

The analysis covers a cross section of 136 countries in 1990. Hence, our data set consists of $18,360$ observations of bilateral export flows (136×135 country pairs). The list of countries is reported in Table A1 in the Appendix. Information on bilateral exports comes from Feenstra et al. (1997). Data on real GDP per capita and population come from the World Bank's *World Development Indicators* (2002). Data on location and dummies indicating contiguity, common language, colonial ties, and access to water are constructed from the CIA's *World Factbook*. Bilateral distance is computed using the great circle distance algorithm provided by Andrew Gray (2001). Remoteness – or relative distance – is calculated as the (log of) GDP-weighted average distance to all other countries (see Wei, 1996). Finally, information on preferential-trade agreements comes from Frankel (1997), complemented with data from the World Trade Organization. The list of preferential trade agreements (and stronger forms of trade agreements) considered in the analysis is displayed in Table A2 in the Appendix. Table A3 in the Appendix provides a description of the variables and displays the summary statistics.

5.2 Results

Table 4 presents the estimation outcomes resulting from OLS and Poisson regressions. The first column reports OLS estimates using the logarithm of trade as the dependent variable; as noted before, this regression leaves out pairs of countries with zero bilateral

[23] As later explained, we include the variable remoteness, or relative distance, in our estimation, in an attempt to control for third-country effects.

exports (only 9, 613 country pairs, or 52 percent of the sample, exhibit positive export flows). For comparison, the second column reports Poisson estimates using only the subsample of positive-trade pairs. Finally, the third column shows the Poisson results for the whole sample (including zero-trade pairs).

The first point to notice is that Poisson-estimated coefficients are remarkably similar using both the whole sample and the positive-trade subsample.[24] However, most coefficients differ – oftentimes significantly – using OLS. This suggests that in this case, heteroskedasticity (rather than truncation) can distort results in a material way. Poisson estimates reveal that the coefficients on importer's and exporter's GDPs are not, as generally believed, close to 1. The estimated GDP elasticities are just above 0.7 (s.e. = 0.03). OLS generates significantly larger estimates, especially on exporter's GDP (0.94, s.e. = 0.01). These findings suggest that the simpler models of gravity equation (those that predict unit-income elasticities typically as a result of specialization in production and homothetic preferences) should be modified to feature a less-than-proportional relationship between trade and GDP.[25] (See Anderson and van Wincoop (2003), who provide a model consistent with smaller elasticities.) It is worth pointing out that unit-income elasticities in the simple gravity framework are at odds with the observation that the trade-to-GDP ratio decreases with total GDP, or, in other words, that smaller countries tend to be more open to international trade.[26]

The role of geographical distance as trade deterrent is significantly larger under OLS; the estimated elasticity is -1.17 (s.e. = 0.03), whereas the Poisson estimate is -0.78 (s.e. = 0.06). Our lower estimate suggests a smaller role for transport costs in the determination of trade patterns. Furthermore, Poisson estimates indicate that, after controlling for bilateral distance, sharing a border does not influence trade flows, while OLS, instead,

[24] The reason why truncation has little effect in this case is that observations with zero trade correspond to pairs for which the estimated value of trade is close to zero. Therefore, the corresponding residuals are also close to zero and their elimination from the sample has little effect.

[25] This result holds when one looks at the subsample of OECD countries. It is also robust to the exclusion of GDP per capita from the regressions.

[26] Note also that Poisson predicts almost equal coefficients for the GDPs of exporters and importers.

generates a substantial effect: It predicts that trade between two contiguous countries is 37 percent larger than trade between countries that do not share a border.[27]

We control for remoteness to account for the hypothesis that larger distances to all other countries might increase bilateral trade between two countries.[28] Poisson regressions support this hypothesis, whereas OLS estimates suggest that only exporter's remoteness increases bilateral flows between two given countries. Access to water appears to be important for trade flows, according to Poisson regressions; the negative coefficients on the land-locked dummies can be interpreted as an indication that ocean transportation is significantly cheaper. In contrast, OLS results suggest that whether or not the exporter is landlocked does not influence trade flows, whereas a landlocked importer experiences lower trade; this asymmetry is hard to interpret. We also explore the role of colonial heritage, obtaining, as before, significant discrepancies: Poisson indicates that colonial ties play no role in determining trade flows, once a dummy variable for common language is introduced. OLS regressions, instead, generate a sizeable effect (countries with a common colonial past trade almost 45 percent more than other pairs). Language is statistically and economically significant under both estimation procedures.

Strikingly, preferential-trade agreements play a much smaller — although still substantial — role according to Poisson regressions. OLS estimates suggest that preferential-trade agreements rise expected bilateral trade by 63 percent, whereas Poisson estimates indicate an average enhancement effect below 20 percent. The contrast in estimates suggests that the biases generated by standard regressions can be substantial, leading to misleading inferences and, perhaps, erroneous policy decisions.[29]

---

[27] The formula to compute this effect is $(e^{b_i} - 1) \times 100\%$, where b_i is the estimated coefficient.

[28] To illustrate the role of remoteness, consider two pairs of countries, (i, j) and (k, l), and assume that the distance between the countries in each pair is the same $D_{ij} = D_{kl}$, however, i and j are closer to other countries. In this case, the most remote countries, k and l, will tend to trade more between each other because they do not have alternative trading partners. See Deardoff (1998).

[29] It is interesting to remark that there is a pattern in the direction of the bias generated by OLS. The bias tends to be positive for the coefficients on variables that relate to larger volumes of trade and, presumably, to larger variance. It tends to be negative for variables that deter trade and, possibly, reduce the variance.

Table 4. The Gravity Equation OLS and Poisson Estimations.

	OLS	Poisson Trade > 0	Poisson
Log of exporter's GDP	0.938**	0.721**	0.733**
	(0.012)	(0.027)	(0.027)
Log of importer's GDP	0.798**	0.732**	0.741**
	(0.012)	(0.028)	(0.027)
Log of exporter's per capita GDP	0.207**	0.154**	0.157**
	(0.017)	(0.053)	(0.053)
Log of importer's per capita GDP	0.106**	0.133**	0.135**
	(0.018)	(0.044)	(0.045)
Log of distance	−1.166**	−0.776**	−0.784**
	(0.034)	(0.055)	(0.055)
Contiguity dummy	0.314*	0.202	0.193
	(0.127)	(0.105)	(0.104)
Common-language dummy	0.678**	0.752**	0.746**
	(0.067)	(0.134)	(0.135)
Colonial-tie dummy	0.397**	0.019	0.024
	(0.070)	(0.150)	(0.150)
Landlocked-exporter dummy	−0.062	−0.873**	−0.864**
	(0.062)	(0.157)	(0.157)
Landlocked-importer dummy	−0.665**	−0.704**	−0.697**
	(0.060)	(0.141)	(0.141)
Exporter's remoteness	0.467**	0.647**	0.660**
	(0.079)	(0.135)	(0.134)
Importer's remoteness	−0.205*	0.549**	0.561**
	(0.085)	(0.120)	(0.118)
Free-trade agreement dummy	0.491**	0.179*	0.181*
	(0.097)	(0.090)	(0.088)
Openness dummy	−0.170**	−0.139	−0.107
	(0.053)	(0.133)	(0.131)
Constant	−28.187**	−31.527**	−32.327**
	(1.101)	(2.161)	(2.059)
Observations	9613	9613	18360
RESET test, p-values	0.000	0.941	0.331

Note: In the OLS regression, the dependent variable is ln(trade). In the Poisson estimation, the dependent variable is trade (the gravity equation is estimated in its multiplicative form). Results for the restricted sample (with positive trade) and the whole sample are reported. The equations use data for 1990. Robust standard errors in parentheses. * significant at 5%; ** significant at 1%

Preferential-trade agreements might also cause trade diversion; if this is the case, the coefficient on the trade-agreement dummy will not reflect the net effect of trade agreements. To account for the possibility of diversion, we include an additional dummy, "openness," similar to that used by Frankel (1997). This dummy takes the value 1 whenever one (or both) of the countries in the pair is part of a preferential-trade agreement and, thus, it captures the extent of trade between members and non-members of a preferential-trade agreement. The sum of the coefficients on the trade agreement and the openness dummies gives the net creation effect of trade agreements. OLS suggests that there is trade destruction coming from trade agreements. Still, the net creation effect is around 40 percent. In contrast, Poisson regressions provide no significant evidence of trade diversion, although the point estimates are of the same order of magnitude under both methods. Hence, even when accounting for trade diversion effects, on average, the Poisson method estimates a smaller effect of preferential-trade agreements on trade, approximately half of that indicated by OLS.

To check the adequacy of the estimated models, we performed a heteroskedasticity-robust RESET test (Ramsey, 1969). This is essentially a test for the correct specification of the conditional expectation, which is performed by checking the significance of an additional regressor constructed as $(x'b)^2$, where b denotes the vector of estimated parameters. The corresponding p-values are reported at the bottom of table 4. In the OLS regression, the test rejects the hypothesis that the coefficient on the test variable is zero. This means that the model estimated using the logarithmic specification is inappropriate. In contrast, the models estimated using the Poisson regressions pass the RESET test, i.e., the RESET test provides no evidence of misspecification of the gravity equations estimated using the Poisson method.

Finally, we also check whether the particular pattern of heteroskedasticity assumed by the models is appropriate. As explained in section 3.2, the adequacy of the log-linear model was checked using the Park-type test, whereas the hypothesis $V[y_i|x] \propto \mu(x_i\beta)$ was tested by evaluating the significance of the coefficient on the term $\ln(\breve{y}_i)\sqrt{\breve{y}_i}$ in the Gauss-Newton regression indicated in expression (11). The p-values of the tests are reported in table 5. Again, the log-linear specification is unequivocally rejected. On the other hand, these results indicate that the estimated coefficient on $\ln(\breve{y}_i)\sqrt{\breve{y}_i}$ is insignificantly different from zero at the usual 5 percent level. This implies that the Poisson PML assumption, $V[y_i|x] = \lambda_0 E[y_i|x]$ cannot be rejected at this significance level.

Table 5: Results of the tests for type of heteroskedasticity (p-values).

Test (Null hypothesis)	Trade > 0	Full sample
GNR ($V[y_i\|x] \propto \mu(x_i\beta)$)	0.063	0.065
Park (OLS is valid)	0.000	0.000

6. Conclusions

In this paper, we argue that the standard empirical methods used to estimate gravity equations are inappropriate. The basic problem is that log-linearization (or, indeed, any non-linear transformation) of the empirical model in the presence of heteroskedasticity leads to inconsistent estimates. This is because the expected value of the logarithm of a random variable depends on higher-order moments of its distribution. Therefore, if the errors are heteroskedastic, the transformed errors will be generally correlated with the covariates. An additional problem of log-linearization is that it is incompatible with the existence of zeroes in trade data, which led to several unsatisfactory solutions, including truncation of the sample (i.e., elimination of zero-trade pairs) and further non-linear transformations of the dependent variable.

To address the various estimation problems, we propose a simple Poisson pseudo-maximum likelihood method and assess its performance using Monte Carlo simulations. We find that in the presence of heteroskedasticity the standard methods can severely bias the estimated coefficients, casting doubt on previous empirical findings. Our method, instead, is robust to different patterns of heteroskedasticity and, in addition, provides a natural way to deal with zeroes in trade data.

We use our method to re-estimate the gravity equation and document significant differences from the results obtained using the log-linear method. Among other differences, income elasticities are systematically smaller than those obtained with log-linearized OLS regressions. In addition, OLS estimation exaggerates the role of geographical proximity and colonial ties. Finally, and perhaps more interesting, the pseudo-maximum likelihood results indicate that bilateral trade between countries that have signed a preferential-trade agreements is, on average, 20 percent larger than that between pairs of countries without agreement, which contrasts with the substantially larger estimate obtained by OLS. Our results suggest that heteroskedasticity (rather than truncation of the data) is responsible for the main differences.

Log-linearized equations estimated by OLS are of course used in many other areas of empirical economics and econometrics. Our Monte Carlo simulations and the regression outcomes indicate that in the presence of heteroskedasticity this practice can lead to significant biases. These results suggest that, at least when there is evidence of heteroskedasticity, the Poisson pseudo-maximum likelihood estimator should be used as a substitute for the standard log-linear model.

References

Anderson, J. (1979). "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69, 106-116.

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

Bhagwati, J. and A. Panagariya, eds. (1996). *Free Trade Areas or Free Trade? The Economics of Preferential Trading.* Washington, DC: The AEI Press.

Bergstrand, J. (1985). "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics*, 69, 474-481.

Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data*, Cambridge, MA: Cambridge University Press.

Cheng, I-H. and H.J. Wall (2002). "Controlling for Heterogeneity in Gravity Models of Trade Integration," Federal Reserve Bank of St. Louis, Working Paper No. 1999-010C.

Davidson, R. and MacKinnon, J.G. (1993). "Estimation and Inference in Econometrics". Oxford: Oxford University Press.

Davis, D. (1995). "Intra-industry Trade: A Hecksher-Ohlin-Ricardo Approach," *Journal of International Economics*, 39, 201-226.

Deardoff, A. (1998). "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in Jeffrey Frankel (ed.), *The Regionalization of the World Economy.* Chicago, IL: University of Chicago Press.

Delgado, M. (1992). "Semiparametric Generalized Least Squares Estimation in the Multivariate Nonlinear Regression Model," *Econometric Theory*, 8, 203-222.

Delgado, M. and Kniesner, T.J. (1997). "Semiparametric Versus Parametric Count-data Models: Econometric Considerations and Estimates of a Hedonic-Equilibrium Model of Workers Absentism," *Review of Economics and Statistics*, 79, 41-49.

Eaton, J. and S. Kortum (2001). "Technology, Geography and Trade," NBER Working Paper No. 6253.

Eicker, F. (1963). "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *The Annals of Mathematical Statistics*, 34, 447-456.

Egger, P. (2000). "A Note on the Proper Econometric Specification of the Gravity Equation," *Economic Letters*, 66, 25-31.

Egger, P. (2002). "An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials," *World Economy*, 25, 297-312.

Egger, P. and Pfaffermayr, M. (2003). "The Proper Panel Econometric Specification of the Gravity Equation: A Three-Way Model With Bilateral Interaction Effects" *Empirical Economics*, 28, 571-580.

Feenstra, R. C., R. E. Lipsey and H. P. Bowen (1997), "World Trade Flows, 1970-1992, with Production and Tariff Data," NBER Working Paper No. 5910.

Feenstra, R., J. Markusen and A. Rose (1999). "Using the Gravity Equation to Differentiate Among Alternative Theories of Trade," *http://www.econ.ucdavis.edu/~feenstra.*

Frankel, J. (1997), *Regional Trading Blocs in the World Economic System.* Washington, DC: Institute for International Economics.

Frankel, J. and A. Rose (2002). "An Estimate of the Effect of Common Currencies on Trade and Income," *Quarterly Journal of Economics*, 117, 409-466.

Frankel, J. and S. Wei (1993). "Trade Blocs and Currency Blocs," NBER Working Paper No. 4335.

Gourieroux, C., A. Monfort and A. Trognon (1984). "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701-720.

Goldberger, A. (1968). "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica*, 36, 464-472.

Goldberger, A. (1991). *A Course in Econometrics.* Cambridge, MA: Harvard University Press.

Gray, A. (2001). http://argray.fateback.com/dist/formula.html.

Harrigan, J. (1993). "OECD Imports and Trade Barriers in 1983," *Journal of International Economics,* 35, 95-111.

Haveman, J. and D. Hummels (2001). "Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization," mimeo Purdue University.

Helpman, E. and P. Krugman (1985). *Market Structure and Foreign Trade.* Cambridge, MA: MIT Press

Koenker, R. and Bassett Jr., G.S. (1978). "Regression Quantiles", *Econometrica*, 46, 33-50.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*, 2nd ed., London: Chapman and Hall.

Manning, W.G. and J. Mullahy (2001). "Estimating Log Models: To Transform or Not to Transform?," *Journal of Health Economics* 20, 461-494.

Mátyás, L. (1997). "Proper Econometric Specification of the Gravity Model," *World Economy*, 20, 363-368.

Mátyás, L. (1998). "The Gravity Model: Some Econometric Considerations," *World Economy*, 21, 397-401.

McCallum, J. (1995). "National Borders Matter: Canada-US Regional Trade Patterns," *American Economic Review*, 85, 615-623.

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, 619-632.

Park, R. (1966). "Estimation with Heteroskedastic Error Terms," *Econometrica*, 34, 888.

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.

Robinson, P. M. (1987). "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875-891.

Rose, A. (2000). "One Money One Market: Estimating the Effect of Common Currencies on Trade," *Economic Policy*, 15, 7-46.

Tenreyro, S. and R. Barro (2002). "Economic Effects of Currency Unions," FRB Boston Series, Working Paper No. 02-4.

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

Wei, S. (1996), "Intra-National versus International Trade: How Stubborn Are Nation States in Globalization?," NBER Working Paper No. 5331.

White, H., (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*, 3rd ed. Berlin: Springer-Verlag.

World Development Indicators CD-rom (2002). The World Bank.

World Factbook (2002). http://www.cia.gov/cia/publications/factbook/

Appendix

Table A1: List of countries.

Albania	Denmark	Kenya	Romania
Algeria	Djibouti	Kiribati	Russian Federation
Angola	Dominican Rp.	Korea Rp.	Rwanda
Argentina	Ecuador	Laos P. Dem. Rp.	Saudi Arabia
Australia	Egypt	Lebanon	Senegal
Austria	El Salvador	Madagascar	Seychelles
Bahamas	Eq. Guinea	Malawi	Sierra Leone
Bahrain	Ethiopia	Malaysia	Singapore
Bangladesh	Fiji	Maldives	Solomon Islands
Barbados	Finland	Mali	South Africa
Belgium-Luxemburg	France	Malta	Spain
Belize	Gabon	Mauritania	Sri Lanka
Benin	Gambia	Mauritius	St. Kitts and Nevis
Bhutan	Germany	Mexico	Sudan
Bolivia	Ghana	Mongolia	Suriname
Brazil	Greece	Morocco	Sweden
Brunei	Guatemala	Mozambique	Switzerland
Bulgaria	Guinea	Nepal	Syrian Arab Rp.
Burkina Faso	Guinea-Bissau	Netherlands	Tanzania
Burundi	Guyana	New Caledonia	Thailand
Cambodia	Haiti	New Zealand	Togo
Cameroon	Honduras	Nicaragua	Trinidad and Tobago
Canada	Hong Kong	Niger	Tunisia
Central African Rp.	Hungary	Nigeria	Turkey
Chad	Iceland	Norway	Uganda
Chile	India	Oman	United Arab Em.
China	Indonesia	Pakistan	U.K.
Colombia	Iran	Panama	U.S.A.
Comoros	Ireland	Papua New Guinea	Uruguay
Congo Dem. Rp.	Israel	Paraguay	Venezuela
Congo Rp.	Italy	Peru	Vietnam
Costa Rica	Jamaica	Philippines	Yemen
Cote D'Ivoire	Japan	Poland	Zambia
Cyprus	Jordan	Portugal	Zimbabwe

Table A2. Preferential Trade Agreements in 1990.

EEC/EC	**CARICOM**	**CACM**
Belgium	Bahamas	Costa Rica
Denmark	Barbados	El Salvador
France	Belize	Guatemala
Germany	Dominican Rp.	Honduras
Greece	Guyana	Nicaragua
Ireland	Haiti	
Italy	Jamaica	**Bilateral Agreements**
Luxembourg	Trinidad and Tobago	EC-Cyprus
Netherlands	St Kitts and Nevis	EC-Malta
Portugal	Suriname	EC-Egypt
Spain		EC-Syria
United Kingdom	**SPARTECA**	EC-Algeria
	Australia	EC-Norway
EFTA	New Zealand	EC-Iceland
Iceland	Fiji	EC-Switzerland
Norway	Kiribati	Canada-US
Switzerland	Papua New Guinea	Israel-US
Liechtenstein	Solomon Islands	
CER	**PATCRA**	
Australia	Australia	
New Zealand	Papua New Guinea	

Table A3. Summary Statistics.

Variable	Full Sample		Trade > 0	
	Mean	Std. Dev.	Mean	Std. Dev.
Trade	172132.2	1828720	328757.7	2517139
Log of trade	—	—	8.43383	3.26819
Log of exporter's GDP	23.24975	2.39727	24.42503	2.29748
Log of importer's GDP	23.24975	2.39727	24.13243	2.43148
Log of exporter's per capita GDP	7.50538	1.63986	8.09600	1.65986
Log of importer's per capita GDP	7.50538	1.63986	7.98602	1.68649
Log of distance	8.78551	0.74168	8.69497	0.77283
Contiguity dummy	0.01961	0.13865	0.02361	0.15185
Common-language dummy	0.20970	0.40710	0.21284	0.40933
Colonial-tie dummy	0.17048	0.37606	0.16894	0.37472
Landlocked exporter dummy	0.15441	0.36135	0.10767	0.30998
Landlocked importer dummy	0.15441	0.36135	0.11401	0.31784
Exporter's remoteness	8.94654	0.26389	8.90383	0.29313
Emporter's remoteness	8.94654	0.26389	8.90787	0.28412
Preferential-trade agreement dummy	0.02505	0.15629	0.04452	0.20626
Openness dummy	0.56373	0.49594	0.65796	0.47442