Is Gravity Linear?

Daniel J. Henderson*

Department of Economics

State University of New York at Binghamton

Daniel L. Millimet†

Department of Economics

Southern Methodist University

Abstract

Despite the solid theoretical foundation on which the gravity model of bilateral trade is based, empirical implementation requires several assumptions which do not follow directly from the underlying theory. First, unobserved trade costs are assumed to be a (log-) linear function of observables. Second, the ad-valorem tax equivalents of trade costs are predominantly assumed to be constant across space, and to a lesser extent time. Maintaining consistency with the underlying theory, but relaxing these assumptions, we estimate gravity models – in levels and logs – using two data sets via nonparametric methods. The results are striking. Despite the added flexibility of the *nonparametric* models, *parametric* models based on these assumptions offer equally or more reliable in-sample forecasts (sometimes) and out-of-sample forecasts (always), particularly in the levels model. Moreover, formal statistical tests *fail to reject* the theoretically consistent parametric functional form. Thus, concerns in the gravity literature over functional form appear unwarranted, and estimation of the gravity model in levels is recommended.

JEL: C14, C33, F14
Keywords: Gravity Model, Bilateral Trade, Border Effect, Currency Union, Generalized Kernel Estimation

*Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000. Tel: (607) 777-4480. Fax: (607) 777-2681. E-mail: djhender@binghamton.edu.

†The authors are grateful for comments on a previous version from Russell Hillberry, David Hummels, Jeff Racine, João Santos Silva, and conference participants at the Texas Camp Econometrics XI. All remaining errors are our own. Corresponding author: Daniel Millimet, Department of Economics, Box 0496, Southern Methodist University, Dallas, TX 75275-0496. Tel: (214) 768-3269. Fax: (214) 768-1821. E-mail: millimet@mail.smu.edu.

1 Introduction

There is little doubt that the gravity model of bilateral trade is of primary importance in empirical analyses of trade patterns. Not only has the model been utilized to gain understanding about trade flows in general, but also to assess the role of particular determinants of trade such as distance, borders, currency unions, WTO membership, insecurity, institutions, etc. This reliance on the gravity model is not surprising; it has a solid theoretical foundation (being derived from several underlying theories) and it has proven empirically successful (explaining much of the variation in trade volume over time and space). Leamer and Levinsohn (1995, p. 1384) state that the gravity model yields "some of the clearest and most robust empirical findings in economics." Rose (2000, p. 11) notes that the gravity model provides a "framework with a long track record of success." Anderson and van Wincoop (2003, p. 170) concur: "The gravity equation is one of the most empirically successful in economics." Others make similar claims.

Despite the solid theoretical foundation on which the gravity model is based, and the empirical success of previous empirical implementations of the gravity model, the majority of these empirical analyses require several assumptions which *do not* follow directly from the underlying theory. In this paper, we investigate the role of three of these assumptions. First, while the gravity model specifies bilateral trade as an increasing function of the income of the two trading partners, and a decreasing function of trade costs, trade costs are *unobserved* by the econometrician. For ease of estimation, trade costs are assumed to be a (log-) linear function of observables. Persson (2001, p. 438) takes perhaps an extreme view on the arbitrariness of this functional form, stating: "There is of course no reason for the relation between these proxies and true trading costs to be linear." Moreover, while it is well known that failure of this functional form restriction may yield misleading inferences, the implication is quite often simply to list this among other caveats, or perform some robustness checks within the confines of the parametric model. However, in principle, which robustness checks are actually performed is ad hoc. Persson (2001, p. 443) continues: "One obvious way of approaching these questions is to extend the regression analysis, searching for direct evidence of important non-linear terms... This is easier said than done, however, because one could imagine an infinite number of non-linear ways in which the ... variables in the basic specification of the gravity equation may affect trade." Here, rather than estimating an "infinite" number of regression models, we take a different approach.

Second, even if the linearity restriction is true, the ad-valorem tax equivalents of trade costs are typically assumed to be constant across space, and to a lesser extent across time. Failure of this restriction may mask meaningful heterogeneity in the data.

The third assumption was first brought to attention in Santos Silva and Tenreyro (2005). The authors provide evidence that the usual assumption of independence between determinants of trade flows and the error term in the *log-linear* gravity model may not hold due to the fact that unobservable determinants of trade flows in the gravity model expressed in *levels* is heteroskedastic. The authors advocate estimating gravity models in levels using a Poisson pseudo-maximum likelihood estimator.

Maintaining consistency with the underlying theory, but relaxing the first two assumptions, we estimate several gravity models in levels *and* logs using recently developed nonparametric methods. Not only does the nonparametric methodology not impose any functional form on the data, but it also yields observation-specific estimates of the income elasticities of trade and the ad-valorem tax equivalents of trade costs. Moreover, the nonparametric methodology offers two additional valuable insights. First, the bandwidths utilized in the nonparametric models provide valuable information into whether the 'true' impact of a particular variable on bilateral trade is linear or non-linear. Second, as shown in Hall et al. (2004), the bandwidths also provide insight into whether a particular variable is irrelevant for explaining bilateral trade flows. As indicated in Anderson and van Wincoop (2004), the decision to include certain variables in the trade cost function is also relatively arbitrary. While the nonparametric methodology cannot discern whether there are omitted relevant attributes, it can test whether there are any included, but irrelevant, variables. In the end, nonparametric estimation tests the usual gravity model along a host of dimensions, and offers a valuable assessment of the success of parametric gravity models relative to a completely flexible alternative.

Assessing the empirical impact of relaxing the assumptions of linearity and homogeneity is more than academic curiosity. First, as argued in Persson (2001), there are theoretical and empirical reasons for suspecting that gravity is non-linear. Second, the only two semiparametric estimations of a gravity model of which we are aware both suggest the existence of important nonlinearities. Persson (2001) utilizes propensity score matching to estimate the impact of currency unions on trade. The advantage of the matching method is that it allows one to control for observables correlated with trade volume and currency unions without specifying a functional form. The drawbacks are that: (i) the first-stage estimation of the propensity score is (typically) parametric, (ii) the method does not yield estimates of the remaining variables in the model, and (iii) the method yields an estimate of average treatment effect parameters, thus not providing any information on the heterogeneity of the effects of the variable of interest. Nonetheless, Persson (2001) finds sharp differences relative to Rose (2000), who estimates the impact of currency unions via parametric regression analysis, using the same data.[1] Ranjan and Tobias (2005) estimate a Bayesian semiparametric hierarchical threshold tobit model, where they allow contract enforcement (a determinant of trade costs) to enter nonparametrically. The results suggest a piecewise linear functional form.

The results are striking. Utilizing two data sets – one on subnational bilateral 'trade' flows across U.S. states (from Wolf (2000) and Millimet and Osang (2005)) and one on cross-country bilateral trade flows (from Glick and Rose (2002)) – and fully nonparametric specifications, we obtain a host of intriguing findings. First, the assumption of linearity *is supported* by the data. Specifically, we are unable to reject the functional form of the level *parametric* or log-linear *parametric* model using either data set (against their nonparametric counterpart). Moreover, in all cases, the distribution of *parametric* predicted values offers a 'better' approximation to the actual data – both statistically and visually – than the nonparametric models. Given our use of aggregate data on trade flows, this is even more surprising since non-linearity in aggregate data may not be synonymous with non-linearity in more disaggregate data

[1] Rose (2001) responds to Persson (2001) providing an updated data set (used here) where the parametric result appears robust to the use of the matching estimator. However, both Rose (2001) and Persson (2001) use a specification for the propensity score that does not balance the gravity variables across the matched treatment and control groups. Thus, the matching estimates are biased (see, e.g., Imbens 2004), and the more preferred estimates provided by the authors are the regression-adjusted estimates obtained using the matched data. However, this regression adjustment brings the authors back to a linear framework (although they include a quadratic for income – surprisingly the only variable that the underlying theory *does predict* should enter the gravity equation linearly), defeating the original purpose for using the matching estimator. The point here is not who is 'right' in the debate between Rose and Persson, but rather that there is evidence suggesting that gravity may be non-linear.

(e.g., trade in specific commodities or trade across more well-defined geographical units). Second, when we split each data set, re-estimate the models on a sub-sample, and predict the volume of bilateral trade for the hold-out sample, we continue to find evidence of the superiority of the *parametric* models. In particular, when estimating the gravity models in *levels*, we find that the parametric model forecasts the U.S. and international data better in-sample and out-of-sample. Using the log specification, the nonparametric model forecasts better in-sample using the U.S. data, but the *parametric model* forecasts better in-sample using the international data and out-of-sample using both data sets. Again, this finding is all the more surprising given the use of data aggregated across commodities and relatively sizeable geographical areas (i.e., states or countries). Finally, in general, the parametric model in *levels* fits the data better than the specification in *logs*, consonant with Santos Silva and Tenreyro (2005).

In the end, then, concerns in the literature over functional form and homogeneity, while well-founded, do not appear to hinder the empirical success of the gravity model. This last finding provides some context and precision to Rose's (2001, p. 456) claim that these "nonlinearities ... are of academic interest, but unimportant in practice." However, concerns over estimation in levels versus logs, posed in Santos Silva and Tenreyro (2005), appear well-founded. The remainder of the paper is organized as follows. Section 2 discusses a theoretical foundation for the gravity model. Section 3 presents the empirical methodology. Section 4 describes the data. Section 5 discusses the results. Section 6 concludes.

2 Theoretical Foundation

As noted in Deardorff (1998), the basic gravity model arises from a large class of underlying structures. These structures have three common elements: (i) trade separability, which arises when production and consumption decisions within a location are separable from bilateral trade decisions across locations; (ii) the aggregator of differentiated products is identical across locations and is of the constant elasticity of substitution (CES) form; and (iii) ad-valorem tax equivalents of trade costs are invariant to trade volume (Anderson and van Wincoop 2004). Given this basic set-up, and considering a one-sector economy where consumers have CES preferences with a common elasticity of substitution σ among all goods, Anderson and van Wincoop (2003) derive the following theoretically consistent gravity model:

$$x_{ij} = \frac{y_i y_j}{y^W}\left[\frac{t_{ij}}{\Pi_i P_j}\right]^{(1-\sigma)} \tag{1}$$

$$\Pi_i^{1-\sigma} = \sum_j \theta_j \left[\frac{t_{ij}}{P_j}\right]^{(1-\sigma)} \qquad \forall i \tag{2}$$

$$P_j^{1-\sigma} = \sum_i \theta_i \left[\frac{t_{ij}}{\Pi_i}\right]^{(1-\sigma)} \qquad \forall j \tag{3}$$

where x_{ij} is the nominal value of exports from i to j, $y_{i(j)}$ denotes total income for i (j), y^W is total world income, $\theta_{i(j)} = y_{i(j)}/y^W$ is the income share of locations, $t_{ij} - 1$ represents the ad-valorem tax equivalent of trade costs, and Π_i and P_j are referred to as multilateral resistance variables. With symmetric trade costs, such that $t_{ij} = t_{ji}$,

$\Pi_i = P_i$, the gravity equation simplifies to

$$\begin{aligned} x_{ij} &= \frac{y_i y_j}{y^W} \left[\frac{t_{ij}}{P_i P_j} \right]^{(1-\sigma)} \\ &= k y_i y_j t_{ij}^{1-\sigma} P_i^{\sigma-1} P_j^{\sigma-1} \end{aligned} \tag{4}$$

where $k \equiv 1/y^w$.

In the literature, there presently exists two proposed methods for estimating (4): in levels or in logs. However, regardless of which approach is utilized, in both cases one must first confront the fact that t_{ij}, P_i, and P_j are unobserved. To handle the unobserved nature of P_i and P_j, four approaches have been utilized. First, observable price variables are used to proxy for the multilateral resistance variables. Second, Anderson and van Wincoop (2003) use the market-clearing conditions to solve for the multilateral resistance variables and obtain a reduced form gravity equation estimable by non-linear least squares. Third, Baier and Bergstrand (2005) propose a computationally more practical alternative to Anderson and van Wincoop's (2003) non-linear least squares method based on a first order Taylor series approximation of the multilateral resistance terms. Finally, as in Hummels (2001), Rose and van Wincoop (2001), Eaton and Kortum (2002), Redding and Venables (2004), Combes et al. (2005) and others, location-specific dummies for i and j are used to control for multilateral resistance (as well as other location-specific unobservables that do not vary across trading partner).[2] The first approach is problematic in that observed price measures do not map directly into their theoretical counterparts (Baier and Bergstrand 2001). The fixed effect approach is robust to certain types of mis-specification, whereas in the absence of mis-specification of these types, the second approach is more efficient – with the third approach offering a reasonable approximation – as it utilizes information derived from the theoretical model. Our specifications are based on the fixed effect approach.

To address the unobservable nature of bilateral trade costs, t_{ij}, we assume that trade costs may be expressed as a function of observables, location-specific unobservables, as well as a multiplicative, pair-specific unobservable, η_{ij}. As a result, we can express trade costs as

$$t_{ij} = g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij})\eta_{ij} \tag{5}$$

where b_{ij} and c_{ij} represent M_1- and M_2-dimensional vectors of binary and continuous observable determinants of trade costs, respectively, and $\widetilde{\alpha}_i$ and $\widetilde{\alpha}_j$ capture unobserved location-specific attributes that affect trade costs with all trading partners. Substituting (5) into (4) yields

$$x_{ij} = k y_i y_j \left[g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij}) \right]^{1-\sigma} P_i^{\sigma-1} P_j^{\sigma-1} \eta_{ij} \tag{6}$$

Equation (6) forms the basis for all the estimations we present below. Our goal is to assess the validity of various methods of estimating (6).

[2] Eaton and Kortum (2002, p. 1760) use location-specific effects as well to control for structural terms in their model which they refer to as "competitiveness."

3 Empirical Methodology

3.1 Parametric Estimation

As stated previously, (6) may be estimated either in levels or in logs. In either case, to estimate the model parametrically, we need to assume a functional form for t_{ij}. Thus, assume that $g(\cdot)$ may be expressed as

$$g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij}) = \widetilde{\alpha}_i \widetilde{\alpha}_j \prod_{m_1=1}^{M_1} \delta_{m_1}^{b_{ij}^{m_1}} \prod_{m_2=1}^{M_2} (c_{ij}^{m_2})^{\rho_{m_2}} \quad (7)$$

where $\delta_{m_1}$is one plus the tariff equivalent of trade barriers associated with b^{m_1} and ρ_{m_2} is the trade elasticity with respect to c^{m_2}(see, e.g., Hummels 2001; Eaton and Kortum 2002; Anderson and van Wincoop 2003, 2004). Substituting (7) into (6), (6) becomes

$$\begin{aligned} x_{ij} &= \exp\{\ln(k) + \ln(y_i) + \ln(y_j) \\ &\quad + (1-\sigma)\left[\ln(\widetilde{\alpha}_i) + \ln(\widetilde{\alpha}_j) + \sum_{m_1} \ln(\delta_{m_1}) b_{ij}^{m_1} + \sum_{m_1} \rho_{m_2} \ln(c_{ij}^{m_2})\right] + (\sigma - 1)\left[\ln(P_i) + \ln(P_j)\right]\}\eta_{ij} \\ &= \exp\left\{\ln(k) + (1-\sigma)\left[\sum_{m_1} \ln(\delta_{m_1}) b_{ij}^{m_1} + \sum_{m_1} \rho_{m_2} \ln(c_{ij}^{m_2})\right] + \alpha_i + \alpha_j\right\}\eta_{ij} \end{aligned} \quad (8)$$

where $\alpha_i \equiv \ln(y_i) + (1-\sigma)\left[\ln(\widetilde{\alpha}_i) - \ln(P_i)\right]$ and α_j is defined analogously. Assuming $\mathrm{E}[\eta_{ij}|b_{ij}, c_{ij}, \alpha_i, \alpha_j] = 1$, then (8) may be estimated consistently using an estimator that is numerically equivalent to the Poisson pseudo-maximum likelihood (PPML) estimator (Santos Silva and Tenreyro 2005). We refer to (8) as the *parametric level specification*.

Alternatively, and the conventional approach prior to Santos Silva and Tenreyro (2005), one may take logs of (8), yielding

$$\ln(x_{ij}) = \ln(k) + (1-\sigma)\left[\sum_{m_1} \ln(\delta_{m_1}) b_{ij}^{m_1} + \sum_{m_1} \rho_{m_2} \ln(c_{ij}^{m_2})\right] + \alpha_i + \alpha_j + \ln(\eta_{ij}) \quad (9)$$

We refer to (9) as the *parametric log specification*. Equation (9) may be estimated using standard fixed effects methods, and yields consistent estimates of the parameters if $\mathrm{E}[\ln(\eta_{ij})|b_{ij}, c_{ij}, \alpha_i, \alpha_j] = 0$. However, as pointed out in Santos Silva and Tenreyro (2005), the assumption that $\mathrm{E}[\ln(\eta_{ij})|b_{ij}, c_{ij}, \alpha_i, \alpha_j] = 0$ *does not follow* from the assumption $\mathrm{E}[\eta_{ij}|b_{ij}, c_{ij}, \alpha_i, \alpha_j] = 1$ since Jensen's Inequality implies that $\mathrm{E}[\ln(z)] \neq \ln(\mathrm{E}[z])$ for some random variable z. Moreover, if η_{ij} is heteroskedastic and its *variance* depends on b_{ij} and/or c_{ij}, then the *expectation* of $\ln(\eta_{ij})$ will in general also depend on b_{ij} and/or c_{ij}, implying that fixed effects estimates based on the log-linear model will be biased.

Aside from issues concerning the error term, as stated in Anderson and van Wincoop (2004), two biases may arise in the estimation of (9) or (8).[3] First, the functional form in (7) may be mis-specified.[4] In general, the gravity

[3]A third bias, due to the potential endogeneity of determinants of trade costs, is not discussed here.

[4]For example, Hummels (2001) argues that trade costs should be an additive, rather than a multiplicative, function of the proxies. However, that may be a misspecification as well.

model in levels is given by

$$
\begin{aligned}
x_{ij} &= \exp\left\{\ln(k) + (1-\sigma)\ln[g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij})] + (\sigma-1)\left[\ln(P_i) + \ln(P_j)\right]\right\}\eta_{ij} \\
&= \exp\{\ln(k) + (1-\sigma)\{\sum_{m_1}\ln(\delta_{m_1})b_{ij}^{m_1} + \sum_{m_1}\rho_{m_2}\ln(c_{ij}^{m_2}) \\
&\qquad + \ln[g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij})] - \sum_{m_1}\ln(\delta_{m_1})b_{ij}^{m_1} - \sum_{m_1}\rho_{m_2}\ln(c_{ij}^{m_2}) - \ln(\widetilde{\alpha}_i) - \ln(\widetilde{\alpha}_j)\} + \alpha_i + \alpha_j\}\eta_{ij} \\
&= \exp\left\{\ln(k) + (1-\sigma)\left[\sum_{m_1}\ln(\delta_{m_1})b_{ij}^{m_1} + \sum_{m_1}\rho_{m_2}\ln(c_{ij}^{m_2}) + \widetilde{\mu}_{ij}\right] + \alpha_i + \alpha_j\right\}\eta_{ij} \\
&= \exp\left\{\ln(k) + (1-\sigma)\left[\sum_{m_1}\ln(\delta_{m_1})b_{ij}^{m_1} + \sum_{m_1}\rho_{m_2}\ln(c_{ij}^{m_2})\right] + \alpha_i + \alpha_j\right\}(\mu_{ij} + \eta_{ij}) \qquad (10)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{ij} &\equiv \exp\left\{(1-\sigma)\left\{\ln[g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij})] - \sum_{m_1}\ln(\delta_{m_1})b_{ij}^{m_1} - \sum_{m_1}\rho_{m_2}\ln(c_{ij}^{m_2}) - \ln(\widetilde{\alpha}_i) - \ln(\widetilde{\alpha}_j)\right\}\right\} \\
&= \left[g(\widetilde{\alpha}_i, \widetilde{\alpha}_j, b_{ij}, c_{ij}) - \widetilde{\alpha}_i\widetilde{\alpha}_j \prod_{m_1=1}^{M_1}\delta_{m_1}^{b_{ij}^{m_1}} \prod_{m_2=1}^{M_2}(c_{ij}^{m_2})^{\rho_{m_2}}\right]^{1-\sigma} \qquad (11)
\end{aligned}
$$

is the deviation between actual trade costs and those obtained by (7). If (7) is mis-specified, it is likely that μ_{ij} will be correlated with b_{ij} and c_{ij}, and biased estimates will result from both the level model (8) and the log-linear model (9).

Second, even if the functional form assumption in (7) is a reasonable approximation, the parameters, δ_{m_1} and ρ_{m_2}, may vary across locations. Recent evidence suggests this is likely to be the case. Specifically, Fratianni and Kang (2006) find that the trade elasticity with respect to distance varies depending on the number of OECD countries in the i, j pair, as well as the religious composition of the trading partners. While suggestive of important heterogeneity, their model only allows heterogeneity in the effect of one variable, distance, and only along certain dimensions, OECD and religious status.[5] Other restricted forms of heterogeneity have also been allowed in the literature. For instance, Eaton and Kortum (2002) estimate a spline in distance, allowing the marginal effect of distance to vary. Rose (2000) allows the impact of a currency union to vary by income, income per capita, distance, and other variables. Persson (2001) includes a few interaction terms between determinants of trade costs. Finally, some researchers have included quadratic terms for income or continuous determinants of trade costs (e.g., Rose 2000; Persson 2001; Rose and van Wincoop 2001).

To assess the magnitude of these two biases in both the level and log specifications, we relax the functional form assumption in (7) that is commonplace in the literature and compare parametric estimates of (8) and (9) with estimates obtained from more flexible, nonparametric models. Moreover, to see if the extent of the biases differs across data sets, we utilize data on subnational 'trade' flows across U.S. states, as well as international data on trade flows across 132 countries and territories. We now turn to the nonparametric models.

[5] Moreover, the authors do not estimate a theoretically consistent gravity model as they omit controls for the multilateral resistance terms. Thus, it is not clear whether the heterogeneity is 'real' or due to the mis-specification of the model.

3.2 Generalized Kernel Estimation

To relax the functional form assumption in (7), we estimate the nonparametric equivalents to (8) and (9) utilizing Li-Racine Generalized Kernel Estimation (Li and Racine 2004; Racine and Li 2004). The nonparametric counterpart to the PPML model follows from (6), and may be written as

$$x_i = m(z_i) + \varepsilon_i, \qquad i = 1, ..., N \tag{12}$$

where $m(\cdot)$ is the unknown smooth *level* gravity function, z_i is a vector of covariates (in levels), ε_i is a mean zero additive error, i indexes observations for ease of exposition (i.e., i refers to a particular exporter-importer-time combination), and N is the total sample size. The argument in m is $z_i = [z_i^c, z_i^u, z_i^o]$, where z_i^c is a vector of continuous regressors, z_i^u is a vector of regressors that assume unordered discrete values, and z_i^o is a vector of regressors that assume ordered discrete values. Specifically, the variables in b and c are placed into the appropriate vector (z^c, z^u, or z^o) and z^u also includes importer and exporter fixed effects.[6] To obtain the additive error form in (12), define $\widetilde{\varepsilon}_i \equiv \eta_i - 1$, substitute $\widetilde{\varepsilon}_i$ into (6), and define $\varepsilon_i \equiv \widetilde{\varepsilon}_i m(z_i)$, which is mean zero and independent of z if η_i has unit mean and is independent of z. We refer to (12) as the *nonparametric level specification.*

The nonparametric counterpart to the log-linear model follows directly from (9), and may be written as

$$\ln(x_i) = \widetilde{m}(\widetilde{z}_i) + \ln(\eta_i), \qquad i = 1, ..., N \tag{13}$$

where $\widetilde{m}(\cdot)$ is the unknown smooth *log* gravity function and $\widetilde{z}_i$ is a vector of covariates (where the continuous variables are in logs). As in the parametric log-linear model, consistency requires $\mathrm{E}[\ln(\eta_i)|z_i] = 0$, which does not follow from the assumption that $\mathrm{E}[\eta_i|z_i] = 1$, and hence $\mathrm{E}[\varepsilon_i|z_i] = 0$. We refer to (13) as the *nonparametric log specification.*

To implement the nonparametric estimation of (12), we take a first-order Taylor expansion of (12) with respect to z_j, yielding

$$x_i \approx m(z_j) + (z_i^c - z_j^c)\beta(z_j) + \varepsilon_i \tag{14}$$

where $\beta(z_j)$ is defined as the partial derivative of $m(z_j)$ with respect to z^c.[7] The estimator of $\Gamma(z_j) \equiv \binom{m(z_j)}{\beta(z_j)}$ is given by

$$\begin{aligned} \widehat{\Gamma}(z_j) &= \begin{pmatrix} \widehat{m}(z_j) \\ \widehat{\beta}(z_j) \end{pmatrix} = \left[\sum_i K_{\widehat{h}} \begin{pmatrix} 1 & (z_i^c - z_j^c) \\ (z_i^c - z_j^c) & (z_i^c - z_j^c)(z_i^c - z_j^c)' \end{pmatrix} \right]^{-1} \\ &\quad \times \sum_i K_{\widehat{h}} \begin{pmatrix} 1 \\ (z_i^c - z_j^c) \end{pmatrix} \ln(x_i), \end{aligned} \tag{15}$$

where

$$K_{\widehat{h}} = \prod_{s=1}^{q} \left(\widehat{\lambda}_s^c\right)^{-1} l^c \left(\frac{(z_{si}^c - z_{sj}^c)}{\widehat{\lambda}_s^c} \right) \prod_{s=1}^{r} l^u \left(z_{si}^u, z_{sj}^u, \widehat{\lambda}_s^u \right) \prod_{s=1}^{p} l^o \left(z_{si}^o, z_{sj}^o, \widehat{\lambda}_s^o \right) \tag{16}$$

is the commonly used product kernel (see Pagan and Ullah 1999). In (16) l^c is the standard normal kernel function with window width λ_s^c associated with the s^{th} component of z^c. For unordered categorical variables, l^u is a variation

[6] In models estimated using data from multiple time periods, we include importer by time and exporter by time fixed effects since the multilateral resistance variables are not time invariant.

[7] The estimation of (13) is completely analogous; thus, for brevity we simply explain the estimation in the context of equation (12).

of Aitchison and Aitken's (1976) kernel function with window width λ_s^u associated with the s^{th} component of z^u. Formally,

$$l^u\left(z_{si}^u, z_{sj}^u, \widehat{\lambda}_s^u\right) = \begin{cases} 1-\lambda_s^u & \text{if } z_{si}^u = z_{sj}^u \\ \frac{\lambda_s^u}{d_s-1} & \text{otherwise} \end{cases} \quad (17)$$

where d_s is the number of unique values z_s can take (e.g., if z_s is binary, $d_s = 2$). For ordered categorical variables, l^o is the Wang and Van Ryzin (1981) kernel function with window width λ_s^o associated with the s^{th} component of z^o. Specifically,

$$l^o\left(z_{si}^o, z_{sj}^o, \widehat{\lambda}_s^o\right) = \begin{cases} 1-\lambda_s^o & \text{if } z_{si}^o = z_{sj}^o \\ \frac{1-\lambda_s^o}{2}(\lambda_s^o)^{|z_{si}^o - z_{sj}^o|} & \text{otherwise} \end{cases} \quad (18)$$

Casual observation of (15) shows that estimates of $\beta(z_j)$ are obtained only for the continuous regressors. The returns to the categorical variables are obtained separately. For example, the coefficient on a dummy variable in the trade cost function (in terms of notation, b from Section 2) is calculated as the counterfactual change in expected bilateral trade if a particular exporter-importer combination switches from one to zero, *ceteris paribus*. Consequently, the returns to the categorical variables also vary across observations. This type of analysis is not common in parametric and semiparametric procedures. See Li and Racine (2004) and Racine and Li (2004) for further details.

Estimation of the bandwidths $h = (\lambda^c, \lambda^u, \lambda^o)$ is typically the most salient factor when performing nonparametric estimation. Although there exist many selection methods, we utilize Hurvich et al.'s (1998) Expected Kullback Leibler (AIC_c) criteria. This method – which chooses smoothing parameters using an improved version of a criterion based on the Akaike Information Criteria – has been shown to perform well in small samples and avoids the tendency to undersmooth as often happens under other approaches such as Least-Squares Cross-Validation. Specifically, the bandwidths are chosen to minimize

$$AIC_c = \log\left(\widehat{\sigma}^2\right) + \frac{1+\mathrm{tr}(H)/N}{1-\left[\mathrm{tr}(H)+2\right]/N} \quad (19)$$

where

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{N}\sum_{j=1}^{N}\left(x_j - \widehat{m}(z_j)\right)^2 \\ &= \left(\frac{1}{N}\right) x'(I-H)'(I-H)x, \end{aligned} \quad (20)$$

and $\widehat{m}(z_j) = Hx_j$.

The bandwidths, by affecting the degree of smoothing, are not just a means to an end; they also provide some indication of how the dependent variable is affected by the regressors. For instance, as the bandwidth on a continuous regressor becomes large, the weight given to each observation becomes equal. In other words, as $\lambda^c \longrightarrow \infty$, the implication is that the regressor enters linearly. However, this linearity does not mean that one should necessarily switch to a semiparametric model for the sake of efficiency. It may be the case that there are important interactions between the 'linear' regressor and the remaining variables in the model, implying that the coefficient on the 'linear' regressor will still vary across z. Moreover, linearity should be more formally assessed, when feasible, using statistical tests, such as the Hsiao et al. (2003) test of functional form discussed below.

For the discrete variables, the bandwidths indicate which variables are relevant, as well as the extent of smoothing in the estimation. From (17) and (18), it follows that if the bandwidth for a particular unordered or ordered discrete variable equals zero, then the kernel reduces to an indicator function and no weight is given to observations for which $z_i \neq z_j$. On the other hand, if the bandwidth for a particular unordered or ordered discrete variable reaches its upper bound, then equal weight is given to observations with $z_i = z_j$ and $z_i \neq z_j$. In this case, the variable is completely smoothed out (and thus does not impact the estimation results). For unordered discrete variables, the upper bound is given by $(d_s - 1)/d_s$. For ordered discrete variables, the upper bound is unity. See Hsiao et al. (2003) and Hall et al. (2004) for further details.

3.3 Model Selection Criteria

To evaluate the various models, we employ four strategies. First, we simply compare the resulting estimates. These comparisons shed light, albeit in an informal manner, on the 'costs' of the restrictions imposed in the parametric models. Second, we conduct Hsiao et al. (2003) specification tests for correct functional form. Third, we obtain the predicted value of the dependent variable from the parametric and nonparametric estimation of each model, plot kernel density estimates of predicted values as well as the actual data, and then conduct Li (1996) tests for the equality of the various distributions. Finally, we randomly split the sample, re-estimate each model on a sub-sample of the original data, and forecast out-of-sample using the estimation results. We then plot kernel density estimates of the forecasts and again use Li (1996) tests, as well as other measures, to evaluate the out-of-sample forecasts.

3.3.1 Statistical Tests

The first formal test employed to evaluate the models is the Hsiao et al. (2003) specification test for mixed categorical and continuous data. The null hypothesis is that the parametric model is correctly specified ($\text{H}_o : \Pr\left[\text{E}(x|z) = f(z,\beta)\right] = 1$) against the alternative that it is not ($\text{H}_1 : \Pr\left[\text{E}(x|z) = f(z,\beta)\right] < 1$). The test statistic is based on $I_N \equiv \text{E}\left(\text{E}\left(\varepsilon|z\right)^2 f(z)\right)$, where $\varepsilon = y - f(z,\beta)$. I_N is non-negative and equals zero if and only if the null is true. The resulting test statistic is

$$J_N = \frac{N\left(\widehat{\lambda}^c\right)^{\frac{q}{2}}\widehat{I}_N}{\widehat{\sigma}_N} \sim N(0,1) \tag{21}$$

where

$$\begin{aligned}
\widehat{I}_N &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\widehat{\varepsilon}_i\widehat{\varepsilon}_j K_{\widehat{h}} \\
\widehat{\sigma}_N^2 &= \frac{2\left(\widehat{\lambda}^c\right)^q}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\widehat{\varepsilon}_i^2\widehat{\varepsilon}_j^2 K_{\widehat{h}}^2
\end{aligned}$$

K is the product kernel, $\widehat{\lambda}^c$ $(\widehat{h})$ is the optimally chosen bandwidth for the continuous (complete set of) covariates, and q is the number of continuous regressors. If the null is false, J_N diverges to positive infinity. As the asymptotic normal approximation performs poorly in finite samples, we utilize a bootstrap procedure to approximate the finite sample null distribution of the test statistic.

As an alternative to the Hsiao et al. (2003) test, we also evaluate the various models in terms of their ability to fit the observed data. If predictions from a restricted model (e.g., the parametric model) closely resemble the observed data, the model may be viewed favorably even if technically one rejects the restrictions. To this end, we compare the predicted values from the parametric and nonparametric models against the observed values.[8] To formalize the 'closeness' of the empirical distributions, we use the Li (1996) test for the equality between two unknown distributions. In this test, the null hypothesis is $\mathrm{H}_o : f(x) = g(x)$ for all x, against the alternative $\mathrm{H}_1 : f(x) \neq g(x)$ for some x.[9] The test, which works with either independent or dependent data, is often used, for example, when testing whether income distributions across two regions, groups, or times are identical. The test statistic, predicated on the integrated square error metric on a space of density functions, $I_1(f,g) = \int (f-g)^2 \, dx$, is

$$J_1 = \frac{N\sqrt{\widehat{\lambda}^c}\widehat{I}_1}{\widehat{\sigma}_1} \sim N(0,1), \tag{22}$$

where

$$\begin{aligned} \widehat{I}_1 &= \frac{1}{N^2\widehat{\lambda}^c}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left[l^c\left(\frac{x_i-x_j}{\widehat{\lambda}^c}\right)+l^c\left(\frac{y_i-y_j}{\widehat{\lambda}^c}\right)-l^c\left(\frac{y_i-x_j}{\widehat{\lambda}^c}\right)-l^c\left(\frac{x_i-y_j}{\widehat{\lambda}^c}\right)\right] \\ \widehat{\sigma}_1^2 &= \frac{1}{N^2\widehat{\lambda}^c\pi^{\frac{1}{2}}}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[l^c\left(\frac{x_i-x_j}{\widehat{\lambda}^c}\right)+l^c\left(\frac{y_i-y_j}{\widehat{\lambda}^c}\right)+2l^c\left(\frac{x_i-y_j}{\widehat{\lambda}^c}\right)\right] \end{aligned}$$

l^c is the standard normal kernel, and $\widehat{\lambda}^c$ is the optimally chosen bandwidth.[10]

3.3.2 Forecast Accuracy

Our final approach to assessing the competing models is to compare not only their in-sample predictive power, but also their forecast accuracy. This is a useful exercise as it may be the case that the greater flexibility afforded by the nonparametric estimation, while offering a better fit in-sample and suggesting additional insights into the process of bilateral trade, does little to improve forecasts beyond the estimation sample. To proceed, we estimate the various models on a randomly chosen sub-sample of the full data, indexed by $i = 1, ..., N_1$.[11] We then obtain the predicted values for both the estimation sample and the hold-out sample. In both cases, we plot the densities of the predicted values and use Li (1996) tests to test for equality between the distributions. Additionally, we compute four measures

[8] Note, one should not in general expect the predicted values to be identical to the observed values even if the model is correctly specified since the predicted values omit the residual. Nonetheless, the tests are useful as a measure of goodness of fit, as well as for examining differences between predictions from competing models.

[9] The explanation that follows assumes that $\{x_i\}$ and $\{y_i\}$ are two equally sized samples of size N, taken from f and g respectively. The extension to unequal sample sizes is trivial.

[10] For further details see Fan and Ullah (1999), Li (1996), and Pagan and Ullah (1999).

[11] Specifically, we draw $u_i \sim \mathrm{U}[0,1]$ for all i and retain those observations with $u \leq 0.5$.

of forecast accuracy:

$$\{\text{Corr}[x_j, \widehat{x_j}]\}^2 = \frac{\{\text{Cov}[x_j, \widehat{x_j}]\}^2}{\text{Var}[x_j]\,\text{Var}[\widehat{x_j}]} \tag{23}$$

$$PMSE = \frac{1}{N_0}\sum_j [x_j - \widehat{x_j}]^2 \tag{24}$$

$$PMAE = \frac{1}{N_0}\sum_j |x_j - \widehat{x_j}| \tag{25}$$

$$PMPE = \frac{1}{N_0}\sum_j \left|\frac{x_j - \widehat{x_j}}{x_j}\right| \tag{26}$$

where j, $j = 1, ..., N_0$, indexes the out-of-sample observations.[12] Equation (23) is the squared correlation coefficient between the actual value and predicted value, $\widehat{x_j}$. Equation (24) is the predicted mean squared error. Equations (25) and (26) reflect the predicted mean absolute error and predicted mean absolute percentage error, respectively. We calculate the same measures to evaluate the in-sample fit, where now j indexes observations in the estimation sample and N_0 is replaced with N_1.

4 Data

We utilize two data sources on bilateral trade. The first is panel data on shipments across U.S. states. Analyses of subnational 'trade' flows have been proliferating of late due to the fact that the data are more consistently measured and trade barriers that are difficult to measure at the country-level do not vary (or at least vary substantially less) at the subnational level (e.g., cultural differences, language barriers, institutions, infrastructure). These properties of subnational data are viewed as particularly desirable in analyses of border effects (Wolf 2000; Millimet and Osang 2005), as subnational borders are less likely to be correlated with unobservable trade barriers.

The inter- and intrastate trade flow data are from the 1993 and 1997 U.S. Commodity Flow Survey (CFS), collected by the Bureau of Transportation Statistics within the U.S. Department of Transportation. The data come from Millimet and Osang (2005); thus, we provide only limited details. The CFS tracks shipments between establishments by mode of transportation: rail, truck, air, water, and pipeline. The survey covers 25 two-digit SIC industries (codes 10 (except 108), 12 (except 124), 14 (except) 148, 20-26, 27 (except 279), 28-39, 41, and 50) and two three-digit SIC industries (codes 596 and 782). The 1993 (1997) survey randomly sampled 200,000 (100,000) establishments. Total shipments from one state to another (or within state) are reported. We measure shipments in real dollars.

Before proceeding there are two important limitations of the CFS data worth noting. First, the CFS tracks all shipments, not just shipments to the final user. For example, shipments of a single good from a factory to a warehouse and then to a retail store would each be included in the data. The fact that wholesale trade is included in the CFS data will likely affect inferences pertaining to the existence and size of the state border effect (Hillberry 2001). Second, Hillberry (2001) and Hillberry and Hummels (2000) argue that utilization of the aggregate CFS data

[12] In the log-linear models, x_j is replaced with $\ln(x_j)$ in equations (23) - (26).

may affect the interpretation of the border effect. However, the emphasis of this paper is not the size of the border effect (or other determinants of trade costs) per se, but rather the sensitivity of inferences regarding determinants of bilateral trade to functional form. Thus, these limitations are not viewed as problematic.

In terms of the remaining variables, y is measured by Gross State Product (GSP), obtained from the U.S. Bureau of Economic Analysis (BEA). The binary variables in the trade cost function, b, include a home dummy (equal to one if $i = j$, zero otherwise) and an adjacency dummy (equal to one if i and j share a common border). The continuous variables in the trade cost function, c, include distance and remoteness measures. The distance measure is borrowed from Wolf (2000) and is defined as the minimum driving distance in miles between the largest city in states i and j. Intrastate distance is computed as one-half the distance between a state and its closest neighboring state, where distance to neighboring states is measured as indicated above. The remoteness variable measures how remote states are vis-a-vis all other trading partners (i.e., excluding the current state in question). Specifically, as in Wolf (2000), $Remote_{ij} = \sum_{k=1,k\neq j}^{48} (d_{ik}/GSP_k)$, where d_{ik} is the distance from i to k; $Remote_{ji}$ is analogously defined. In general, a state located in the middle of a country will be less 'remote' than coastal or international border states (on average, Iowa is the least remote, while Oregon is the most remote). Finally, when estimating the generalized gravity model, three price variables are included: (i) regional consumer price indices, available for four regions – West, East, South, Midwest – from the U.S. Bureau of Labor Statistics; (ii) average wage per job for each state, available from the BEA; and, (iii) state-specific GSP deflators, calculated as $(NGSP_{it}/RGSP_{it})/(NGSP_{89}/RGSP_{89})$, $t = 1993, 1997$, where $NGSP$ denotes nominal GSP and $RGSP$ denotes real GSP (all measures taken from the BEA). Summary statistics for all variables are provided in Table 1.

The second data set is a cross-section of bilateral trade flows across 132 non-industrial countries. The data are from Glick and Rose (2002); again we provide limited details.[13] While the data are available for the entire period 1948-1997, as well as for industrial countries, we utilize this sub-sample for two main reasons. First, the computation time for the nonparametric models increases exponentially with sample size. However, as shown in Glick and Rose (2002), the results are fairly stable over time. In light of this, we focus on 1995 simply because it lies in the middle period spanned by the subnational data. Second, trade barriers are higher for non-industrial countries, indicating that functional form issues may be more relevant for the sub-sample of non-industrial countries (Anderson and van Wincoop 2004).

For the cross-country data, the dependent variable is the (log) real dollar value of bilateral trade. The binary variables in the trade cost function, b, include a currency union dummy (equal to one if i and j share a currency, zero otherwise), an adjacency dummy (equal to one if i and j share a common border), a common language dummy (equal to one if i and j share a language, zero otherwise), and a regional trade agreement (RTA) dummy (equal to one if i and j are in a RTA, zero otherwise). Distance is the only continuous variable in the trade cost function. In addition, in the terminology relevant for the nonparametric model, the trade cost function also includes two ordered, discrete variables: the number of landlocked countries (0, 1, or 2) and the number of islands (0, 1, or 2). Summary

[13]The authors are grateful to Andrew Rose for making the data available (http://faculty.haas.berkeley.edu/arose/).

statistics are displayed in Table 1.[14]

5 Results

5.1 U.S. Subnational Trade

5.1.1 Functional Form

Prior to discussing the results, Table 2 presents the bandwidths for the nonparametric models. Panel I corresponds to the level specification; Panel II corresponds to the log specification. The bandwidths in the level specification reveal three salient points. First, the bandwidths on the three continuous variables each exceed 1.60E+06. Since the regressors behave linearly as the bandwidths approach infinity, this suggests that a linear approximation may not be far off the mark. Second, the bandwidths on the home dummy, adjacency dummy, and importer by time dummies are much smaller than their respective upper bounds, implying that these variables are relevant in the model. Moreover, since the bandwidths are extremely close to zero, this implies that observations with different values of these covariates are given (essentially) no weight in the estimation; the kernel reduces to an indicator function. In other words, the estimation is equivalent to dividing the sample into cells and performing separate estimations within each cell. Finally, the bandwidth on the exporter by time dummies is very close to its upper bound; thus, this variable is (likely) irrelevant in explaining trade flows across U.S. states. Since only one set of state effects are relevant in the model, this suggests that symmetric trade costs may also be a reasonable approximation.

The bandwidths for the log specification (Panel II) differ only modestly in terms of their implications. First, the continuous variables continue to enter (approximately) linearly. Second, the home dummy, adjacency dummy, and importer by time dummies are (likely) relevant in the model, while the exporter by time dummies are (likely) irrelevant (again, consonant with symmetric trade costs). However, as the bandwidths for the home and adjacency dummies are not as close to zero as in Panel I, there is more smoothing "across cells" in the log specification.

In sum, examination of the bandwidths suggest that linearity coupled with symmetric trade costs may offer a fairly good representation of the theoretically consistent gravity model. However, linearity is not synonymous with homogeneous effects of the covariates. For instance, since the implication is that estimation is "by cell" (at least defined by importer and time), this suggests linear, but heterogeneous, gravity models across cells. We now turn to the actual results, as well as more formal statistical tests of functional form.

5.1.2 Border Effect

Table 3 presents the first set of results, focusing on one particular determinant of subnational trade: political borders. We provide a detailed analysis of the results for the home dummy in light of the fact that the economically significant impact of borders on trade flows has been extensively documented. McCallum's (1995) well known result that, even

[14] Both data sets exclude observations with zero trade. As noted in Santos Silva and Tenreyro (2005), an additional advantage of PPML estimation is the ability to naturally handle zeros. However, since the goal of the paper is to assess functional form issues within and across the level and log specifications, we utilize a consistent sample for comparability.

after controlling for the usual determinants of bilateral trade flows, trade between Canadian provinces is significantly larger (by a factor of 22) than cross-border trade with U.S. states has led to a number of subsequent analyses. Using post-NAFTA data for the period 1994–1996, Helliwell (1998) finds that the border effect declined by almost 50% compared to McCallum's pre-NAFTA estimate. However, using Helliwell's data and controlling for unobserved time invariant attributes, Wall (2000) estimates a U.S.-Canada border effect that is 40% larger than that reported by McCallum. Anderson and Van Wincoop (2003), relying on a parametric theoretically consistent specification of the gravity equation (as discussed in Section 2), report a border effect much smaller than McCallum; the border reduces trade between Canada and the U.S. by about 44%. Obstfeld and Rogoff (2000) label this border effect one of the six major puzzles in international macroeconomics.

Because national borders may be correlated with unobservable determinants of trade, such as linguistic or cultural differences, many researchers have focused on analyzing border effects at the subnational level. However, despite the implicit controls for such unobservables at the subnational level, Wolf (2000) still finds a statistically and economically significant border effect using data from the 1993 CFS.[15] Millimet and Osang (2005) utilize the 1993 and 1997 CFS, verifying the robustness of the border effect along many dimensions. Thus, one conclusion is that other factors must account for the border effect. Prior to reaching such a conclusion, it is important to know whether these findings are an artifact of a mis-specified model.

Turning to the results, Table 3 displays the mean nonparametric estimate of the border effect, as well as the coefficient at each decile of the distribution, for each specification. For comparison, the corresponding parametric coefficients are also reported. Finally, we display the home bias, equal to the anti-log of the coefficient, as well as the tariff equivalent of the border effect assuming an elasticity of substitution of eight. The results from the level (log) specification are presented in Panel I (II).

In Panel I, the parametric estimate, obtained from (8), is 1.72 (s.e. = 0.04) and highly statistically and economically significant, implying that *intrastate* trade is nearly six times greater than *interstate* trade, ignoring the effect on multilateral resistance, and the tariff equivalent is approximately 28%.[16] The parametric estimate is well above the corresponding mean and 90^{th} percentile of the nonparametric estimates; the median (90^{th} percentile) nonparametric counterpart is 0.79 (0.95), yielding a tariff equivalent of roughly 12% (15%). Thus, on the one hand, the nonparametric level specification indicates that the parametric level specification vastly overstates the border effect. On the other hand, the border effect is still large and economically significant, constituting at least a 10% tariff equivalent for roughly 70% of the sample.

In Panel II, the parametric estimate, obtained from the log specification in (9), of 1.82 (s.e. = 0.06) is marginally greater than the PPML estimate. However, while the parametric estimate does not change drastically from the level specification, the mean (median) nonparametric estimate increases to 1.42 (1.48). Moreover, the distribution of nonparametric estimates is much more disperse, as the tariff equivalent is roughly 3.5 times as large at the 90^{th}

[15] Wolf's (2000) border coefficient implies a tariff equivalent of the border cost of approximately 24% (assuming an elasticity of substitution of eight). However, one should not overlook the caveats discussed in Section 4.

[16] Fixed effects estimates of the theoretically consistent model yield unbiased estimates of the *average* border effect (assuming no mis-specification), despite ignoring the multilateral resistance terms (Feenstra 2004; Baier and Bergstrand 2005).

percentile (36%) as it is at the 10^{th} percentile (10%). Despite these changes, the parametric estimate still falls above the 80^{th} percentile of the nonparametric estimates.

In sum, relaxation of the linear functional form reveals two findings. First, the nonparametric level and log specifications both suggest a smaller, but still statistically and economically significant, impact of borders on subnational trade. Second, while switching from the level to the log specification results in a modest increase in the estimated size of the border effect in the parametric model, the increase is much more pronounced in the nonparametric model (increasing both the magnitude and dispersion of the home bias).

5.1.3 Continuous Covariates

Table 4 presents the results for the continuous covariates. For each specification, we present the parametric results, as well as the nonparametric mean estimate, and the nonparametric estimates corresponding to the 25^{th}, 50^{th}, and 75^{th} percentiles of the distribution (labelled $Q1$, $Q2$, and $Q3$).

In terms of the level specification (Panel I), three findings stand out. First, the median nonparametric estimates of the distance and remoteness elasticities of trade are much smaller in absolute value than the corresponding parametric estimates.[17] For instance, the parametric model suggests a much stronger impact of distance on subnational trade, with an elasticity of -0.29 (s.e. = 0.03) as compared with -0.10 (s.e. = 0.07) at the median of the nonparametric estimates. Second, while the parametric model finds strong effects of both importer and exporter remoteness (elasticities above 20, standard errors under three), the nonparametric model indicates that only the remoteness of the importing state matters (median elasticity of 6.73 with a standard error of 3.40); remoteness of the exporting state is statistically insignificant at conventional levels for the majority of the sample. Finally, the assumption of homogeneous effects of the continuous variables appears reasonable for distance and exporter remoteness. Specifically, the interquartile ranges for these two variables are sizeable in economic terms, but are not overly large relative to the standard errors. This contrasts with the statistically and economically significant differences in the border effect across states in Panel I of Table 3, as well as importer remoteness.

In terms of log specification (Panel II), again three points are noteworthy. First, the mean and median nonparametric estimates, as well as the parametric estimate, of the distance elasticity of trade are not statistically distinguishable (roughly -0.50, with larger standard errors in the nonparametric model), and indicate a larger impact of distance relative to the level results in Panel I. In combination with the results in Table 3, then, the level specification suggests a *smaller* effect of geography on subnational trade; borders and distance both matter less. Second, as with the nonparametric level results in Panel I, the nonparametric log specification indicates that only importer remoteness impacts trade patterns (median elasticity of 38.85 with a standard error of 7.07). Finally, the assumption of homogeneous effects of the continuous variables seems even more reasonable in the log, as opposed to the level, specification. While the interquartile ranges for the three variables continue to be large in absolute terms,

[17] Note, the mean nonparametric estimates are very different from the median estimates, indicating an asymmetrc distribution (and, most likely, the presence of outliers). In the nonparametric level specification, the covariates in (12) enter in levels and elasticities are evaluated at the sample average using the observation-specific slope coefficients. In other words, the elasticity for observation i for variable z_j is calculated as $\widehat{\beta}(z_{ij}) * (\overline{z}_j/\overline{x})$, where overbars indicate sample averages.

the confidence intervals for the point estimates overlap in each case.

Taking a step back, one fact becomes clear from these results: all four models yield different conclusions regarding the determinants of subnational trade flows. Specifically, the *parametric level specification* indicates the importance of remoteness and borders, whereas the *parametric log specification* also points to a much larger role for distance. On the other hand, the *nonparametric level specification* finds a large impact of importer remoteness and a smaller role for borders; the *nonparametric log specification* finds a large role for importer remoteness, borders, and distance. In light of these differences, we now seek to more formally assess the validity of the various models.

5.1.4 Model Comparisons

To assess the reasonableness of the linearity and homogeneity assumptions in a statistical sense, we first conduct the Hsiao et al. (2003) test for functional form. Strikingly, we *fail* to reject the parametric functional form in the level specification (p = 0.63) and log specification (p = 0.54), consonant with the interpretation of the bandwidths discussed previously.[18] This result is extremely powerful, implying that despite the differences in the nonparametric and parametric estimates, the parametric specifications are preferable (on efficiency grounds) since the restrictions of the parametric models cannot be rejected.

Despite this finding, we also assess the models along other criteria. To examine the predictive power of the parametric and nonparametric models, we obtain kernel density estimates of the distributions of the parametric and nonparametric predicted values from each model, as well as the observed data. The densities are plotted in Figures 1 (levels) and 2 (logs). Table 8 presents the results of the Li (1996) pairwise tests for equality of the distributions. In terms of the level specification (Figure 1 and Panel I), the parametric model fits the observed data remarkably well. The nonparametric model does not perform as well, under-predicting the level of trade.[19] The Li tests confirm these observations: equality between the nonparametric predicted distribution and observed distribution is strongly rejected (p = 0.00); we also reject equality between the parametric predicted distribution and observed distribution (p = 0.01) as well as the two predicted distributions (p = 0.01). The lack of equality between the two predicted distributions, combined with the fact that the value of the test statistic for the parametric model (2.35) is much lower than for the nonparametric model (10.44), indicates that the parametric model is significantly closer (in a statistical sense) to the observed data. In combination with the results of the Hsiao et al. (2003) functional form test, the evidence indicates the parametric model is quite reasonable in terms of fitting the data.

In terms of the log specification (Figure 2 and Panel II), the results are qualitatively unaltered. Specifically, Figure 2 and the Li tests continue to reject equality in each pairwise test, but the parametric distribution is closer to the actual data visually as well as measured by the magnitude of the test statistic (3.69 versus 20.48). In particular, the nonparametric model performs worse in terms of fitting the lower tail of the distribution. In combination with the Hsiao et al. test, this provides further confirmation that a parametric theoretically consistent model based on a

[18] The p-values are obtained using 399 bootstrap repetitions.

[19] Note, the nonparametric level specification does not restrict the predicted values to be non-negative. The kernel densities plotted in the corresponding figures use Silverman's (1986) reflection method to correct for this.

log-linear trade cost function and homogeneity performs quite well.

As a final means of assessing the models, we split the sample and re-estimate each model using a randomly chosen 2,155 observations; 2,073 observations are relegated to the hold-out sample. The four measures of forecast accuracy – equations (23) - (26) – are presented in Table 9. Table 10 presents the corresponding results of the Li tests of equal distributions; Figure A1 in the Appendix contains the corresponding densities. Two striking findings emerge. First, the *parametric level specification* outperforms the corresponding nonparametric model in-sample and out-of-sample according to the four measures reported in Table 9. Moreover, the Li test *fails* to reject equality between the parametric predicted and actual densities in-sample ($p = 0.32$) *and* at marginal levels out-of-sample ($p = 0.07$). However, the Li test easily rejects equality between the nonparametric distributions and the actual data (in-sample: $p = 0.00$; out-of-sample: $p = 0.00$). Moreover, as evidenced in Figure A1, the nonparametric model continues to under-predict the volume of trade both in- and out-of-sample.

Second, while the *nonparametric log specification* outperforms the corresponding parametric model in-sample, the *parametric log specification* fares better out-of-sample according to the four measures reported in Table 9. In addition, the Li test indicates a better fit by the *parametric log specification* both in-sample and out-of-sample. The Li test statistic is 1.41 (1.86) in-sample for the parametric (nonparametric) model; 1.67 (6.31) out-of-sample. Moreover, the Li test fails to reject equality between the parametric and actual in-sample densities at the $p < 0.05$ level. As seen in Figure A1, the improvement in the parametric model comes from its ability to better fit the tails of the distributions.

In sum, formal statistical tests fail to reject the parametric assumptions in the usual log gravity specification, as well as the level specification of Santos Silva and Tenreyro (2005). Moreover, tests based on predictive power indicate the level specification fits the data even better than log specification. Thus, at least with U.S. subnational data, the linearity and homogeneity assumptions do not appear problematic, and estimation in levels seems preferable. To see if this conclusion holds using cross-country data, we turn to the international results.

5.2 Cross-Country Trade

5.2.1 Functional Form

As with the U.S. data, we begin by discussing the bandwidths. Table 5 presents the results for the nonparametric level (Panel I) and log (Panel II) specifications. The bandwidths yield two findings. First, in both specifications, the bandwidth for one set of country dummies is approximately equal to its upper bound, indicating that this variable is (likely) irrelevant in explaining trade flows as the data are completely smoothed across observations with different values of this variable. As such, the resulting estimates are not sensitive to the value assumed by this variable. As with the U.S. subnational data, this suggests that symmetric trade costs may be a reasonable approximation. Second, the results for both specifications suggest that distance most likely enters linearly (the bandwidths exceed 1.0E+10), currency unions and sharing a border are irrelevant in explaining trade flows, and that common language and the number of islands are relevant. However, the level and log specifications yield different conclusions with respect to the relevance of regional trade agreements (level: relevant; log: irrelevant) and the number of landlocked countries (level: irrelevant; log: relevant), as well as the amount of smoothing across observations with different number of islands (level: no smoothing; log: some smoothing).

5.2.2 Currency Union Effect

Table 6 presents the first set of results, focusing on one specific determinant of international trade: currency unions. As with the subnational data, we provide a detailed analysis of the results for one of the binary variables given its importance in the literature; the impact of currency unions on trade flows has been widely analyzed (see Alesina et al. 2003). Rose (2000, 2001) and Rose and van Wincoop (2001) find that currency unions increase trade by over 300% (50%) ignoring (accounting for) the effect on the multilateral resistance variables; Rose and van Wincoop report an ad valorem tax equivalent of 14% (again, with an elasticity of substitution of eight). Persson (2001), however, finds a much smaller effect, albeit still large, utilizing a semiparametric propensity score matching to estimate the impact of currency unions on trade. Rose (2004) conducts a meta-analysis, concluding that currency unions increase trade by about 100%.[20]

Table 6 displays the mean nonparametric estimate of the currency union effect, as well as the coefficient at each decile of the distribution. For comparison, the corresponding parametric coefficients are also reported. Finally, we display the currency union effect, equal to the anti-log of the coefficient, as well as the tariff equivalent of the currency union effect assuming an elasticity of substitution of eight. The parametric result, obtained from the level specification, is 0.60 (s.e. = 0.29) and statistically significant, implying that *intra-union* trade is nearly twice is large as *inter-union* trade (ignoring the effect on multilateral resistance), consonant with Rose (2004), and the tariff equivalent is approximately 9%.[21] The parametric estimate is larger than all the nonparametric estimates; the nonparametric estimates are essentially zero, consistent with the discussion above concerning the bandwidth interpretation. In the parametric log specification, the currency union effect is 1.09 (s.e. = 0.32), with a tariff equivalent of 17%. Again, the nonparametric estimates are essentially zero. Thus, there are obviously substantial differences in the estimated currency union effect across the various models. However, prior to concluding that the currency union effect is an artifact of mis-specification in cross-country gravity models, we must assess the statistical properties of the parametric models.

[20] Many researchers question the exogeneity of currency unions. Given the infancy of nonparametric instrumental variable estimation techniques, along with the questions of endogeneity that could be asked of many determinants of trade costs, we ignore this issue at present (see footnote 3). However, Alesina et al. (2003) find *larger* currency union effects when treating such unions as endogenous. Thus, our estimates may be viewed as lower bounds (although the parametric IV model in Alesina et al may be mis-specified).

[21] As noted previously with respect to the border effect, fixed effects estimates of the theoretically consistent model yield unbiased estimates of the *average* currency union effect (assuming no mis-specification), despite ignoring the multilateral resistance terms (Feenstra 2004; Baier and Bergstrand 2005).

5.2.3 Continuous Covariates

Before turning to the statistical tests, Table 7 presents the results for the effect of the sole continuous covariate, distance. As before, we present the parametric results, as well as the nonparametric mean estimate, and the nonparametric estimates corresponding to the 25^{th}, 50^{th}, and 75^{th} percentiles of the distribution (labelled $Q1$, $Q2$, and $Q3$). Three results emerge. First, the parametric level estimate of the distance elasticity is smaller in magnitude than the parametric log estimate (-1.05 (s.e. = 0.09) versus -1.52 (s.e. = 0.07)); similarly, the nonparametric level estimates are smaller than their corresponding log estimates. Second, the nonparametric level estimates are statistically insignificant at the three quartiles; the elasticities are statistically significant at the first and second quartile in the nonparametric log specification. Finally, whereas the parametric level estimate lies in extreme lower tail of the distribution of nonparametric level estimates, the parametric log estimate lies above the 25^{th} percentile of the distribution of the nonparametric log estimates. As with the currency union results, the nonparametric estimates of the distance elasticities are interesting in light of the "puzzle" noted by some that the usual gravity model fails to provide evidence of a diminishing effect of distance on trade over time as globalization becomes more pronounced (see, e.g., Brun et al. 2005). Nonetheless, despite the apparent resolution of this "puzzle" using nonparametric techniques, we must assess the statistical validity of the parametric models.

5.2.4 Model Comparisons

To assess the parametric models, we begin with the Hsiao et al. (2003) test for functional form. As with the U.S. subnational data, we once again *fail* to reject the parametric functional form in the level specification (p = 0.74) and log specification (p = 0.16), consonant with the interpretation of the bandwidths discussed previously. Consequently, despite the significant differences in coefficient estimates across the nonparametric and parametric models, the parametric results are again the preferable estimates (on efficiency grounds).

Turning to the predictive power of the parametric and nonparametric models, the kernel density estimates of the distributions of the parametric and nonparametric predicted values from each model, as well as the observed data, are plotted in Figures 3 (levels) and 4 (logs). Table 8 presents the results of the Li (1996) pairwise tests for equality of the distributions (Panels III and IV). In terms of the level specification (Figure 3 and Panel III), the parametric model fits the observed data remarkably better than the nonparametric model. As in the U.S. subnational data, the nonparametric model vastly under-predicts the level of trade.[22] The Li tests confirm these observations: equality between the nonparametric predicted distribution and observed distribution is strongly rejected (p = 0.00), whereas we *fail* to reject equality between the parametric predicted distribution and observed distribution (p = 0.24). Moreover, we reject equality of the two predicted distributions: parametric versus nonparametric (p = 0.00). Thus, as with the U.S. data, the graphical results suggest the parametric level specification performs quite well. In fact, this result is perhaps more surprising using cross-country data given the additional heterogeneity in the sample.

In terms of the log specification (Figure 4 and Panel IV), the parametric model continues to fare well relative

[22] Again, the nonparametric kernel densities utilize Silverman's (1986) reflection method to correct for negative predicted values.

to the nonparametric model. However, both models fit the observed data worse than their counterpart in the level specification. Specifically, the Li tests now reject equality in each pairwise test, but the parametric distribution continues to lie closer to the actual data both visually, as well as measured by the magnitude of the test statistic (39.89 versus 199.98). Moreover, as evidenced in Figure 4, neither model fits the tails of the distribution particularly well, although the parametric model does better. As a result, consonant with the U.S. subnational data, a theoretically consistent gravity model based on a linear, level specification fits the data best.

As a final means of assessing the models, we split the sample and re-estimate each model using a randomly chosen 2,171 observations; 2,137 observations are relegated to the hold-out sample. The results are presented in Table 9 (Panels III and IV). Table 10 presents the corresponding results of the Li tests of equal distributions (Panels III and IV); Figure A2 in the Appendix contains the corresponding densities. As with the U.S. subnational data, two salient findings emerge. First, the *parametric level specification* outperforms the corresponding nonparametric model in-sample according to three of four forecast measures, and out-of-sample according to all four measures. Moreover, the Li test *fails* to reject equality between the parametric predicted and actual densities in-sample ($p = 0.45$), but not out-of-sample ($p = 0.00$). However, the Li test easily rejects equality between the nonparametric distributions and the actual data (in-sample: $p = 0.01$; out-of-sample: $p = 0.00$), and the value of test statistic is larger (in absolute value) out-of-sample for the nonparametric model, indicating a worse fit (parametric: 15.45; nonparametric: 51.53).[23]

Second, the *parametric log specification* fares better both in-sample *and* out-of-sample according to all four measures reported in Table 9. In addition, the Li test indicates a better fit by the parametric model both in-sample and out-of-sample. The Li test statistic is 18.35 (109.16) in-sample for the parametric (nonparametric) model; 14.43 (141.49) out-of-sample. However, unlike in the parametric level specification, the Li test rejects equality between the parametric and actual in-sample densities ($p = 0.00$).

In sum, formal statistical tests fail to reject the parametric assumptions in the level specification, as well as the usual log specification, of the gravity equation using cross-country data. Furthermore, tests based on predictive power indicate the level specification of Santos Silva and Tenreyro (2005) outperforms the nonparametric level specification, as well as the parametric log specification, confirming the conclusions from the U.S. subnational data.

6 Conclusion

The gravity model of bilateral trade has assumed a prominent place in empirical analyses of bilateral trade flows, as well as being recently extended to bilateral capital and equity flows. A cynic might conjecture that the model's

[23] Viewing the out-of-sample plot in Figure A2, it appears that the nonparametric model fits the data better than the parametric model despite the larger Li test statistic. There exist two possible explanations for this apparent discrepancy. First, the nonparametric distribution may be estimated less precisely. Second, the plots in the figures use Silverman's (1986) reflection method to account for the fact that some nonparametric predicted values are negative. However, the Li test is based on the actual predicted values, inclusive of negative values. Performing the Li tests on the data used to generate the plots does not affect the qualitative results in the paper concerning rejection of the null of equality between any pair of distributions.

popularity is as much attributable to its ease of estimation as its theoretical underpinning and apparent empirical success (e.g., Baier and Bergstrand 2005). Nonetheless, the gravity model has been used to assess the impact of some very policy relevant variables such as WTO membership, currency unions, political borders, and more. The sensitivity of the resulting conclusions to assumptions about which the underlying theory is absent, namely linearity and homogeneity of the trade cost function, is predominantly absent in the literature. Here, we assess the impact of these assumptions along a host of dimensions using panel data on subnational trade flows from the U.S. and cross-sectional data at the country-level.

The findings indicate that relaxation of the usual restrictions embedded in log-linear parametric gravity models has a substantial impact on the marginal effects of specific trade determinants. For example, the effects of political borders and currency unions appear highly overstated. However, formal statistical tests *fail to reject* these restrictions, implying that the parametric estimates are more efficient. Furthermore, in terms of explaining the distribution of bilateral trade – both in-sample and out-of-sample – parametric models, particularly the level specification of Santos Silva and Tenreyro (2005), fare better than their nonparametric counterparts that impose little structure. Thus, the bottom line is that the gravity model along with the assumption of linear, homogeneous trade costs and estimation in levels is a fair representation of data on bilateral trade flows.

References

[1] Aitchison, J. and C.G.G. Aitken (1976), "Multivariate Binary Discrimination by Kernel Method," *Biometrika*, 63, 413-420.

[2] Alesina, A., R.J. Barro, and S. Tenreyro (2003), "Optimal Currency Areas,"in M. Gertler and K. Rogoff (eds.) *NBER Macroeconomics Annual 2002*, Cambridge, MA: MIT Press.

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

[4] Anderson, J.E. and E. van Wincoop (2004), "Trade Costs," *Journal of Economic Literature*, 42, 691-751.

[5] Baier, S.L. and J.H. Bergstrand (2001), "The Growth of World Trade: Tariffs, Transport Costs and Income Similarity," *Journal of International Economics*, 53, 1-27.

[6] Baier, S.L. and J.H. Bergstrand (2005), "*Bonus Vetus* OLS: A Simple OLS Approach for Addressing the 'Border Puzzle' and Other Gravity-Equation Issues," unpublished manuscript, University of Notre Dame.

[7] Brun, J.-F., Brun, C. Carrére, P. Guillaumont, and J. Melo (2005), "Has Distance Died? Evidence from a Panel Gravity Model," *World Bank Economic Review*, 19, 99-120.

[8] Combes, P.-P., M. Lafourcade, and T. Mayer (2005), "The Trade-Creating Effects of Business and Social Networks: Evidence from France," *Journal of International Economics*, 66, 1-29.

[9] Deardorff, A. (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in J.A. Frankel (ed.) *The Regionalization of the World Economy*, Chicago: University of Chicago Press.

[10] Eaton, J. and S. Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, 70, 1741-1799.

[11] Estevadeordal, A., B. Frantz, and A.M. Taylor (2003), "The Rise and Fall of World Trade,"*Quarterly Journal of Economics*, 68, 359-408.

[12] Fan, Y. and A. Ullah (1999), "On Goodness-of-Fit Tests for Weekly Dependent Processes Using Kernel Method," *Journal of Nonparametric Statistics*, 11, 337-360.

[13] Fratianni, M. and H. Kang (2006), "Heterogeneous Distance–Elasticities in Trade Gravity Models," *Economics Letters*, 90, 68-71.

[14] Glick, R. and A.K. Rose (2002), "Does a Currency Union Affect Trade? The Time-Series Evidence," *European Economic Review*, 46, 1125-1151.

[15] Hall, P., J. Racine and Q. Li (2004), "Cross-Validation and the Estimation of Conditional Probability Densities," *Journal of the American Statistical Association*, 99, 1015-1026.

[16] Helliwell, J.F. (1998), *How Much Do National Borders Matter?*, Washington, D.C.: Brookings Institution Press.

[17] Hillberry, R. (2001), "Aggregation Bias, Compositional Change, and the Border Effect," Office of Economics Working Paper No. 2001-04-B, U.S. International Trade Commission.

[18] Hillberry, R. and D. Hummels (2000), "Explaining Home Bias in Consumption: Production Location, Commodity Composition, and Magnification," unpublished manuscript, Krannert School of Management, Purdue University.

[19] Hillberry, R. and D. Hummels (2003), "Intra-national Home Bias: Some Explanations," *Review of Economics and Statistics*, 85, 1089-1092.

[20] Hsiao, C., Q. Li and J. Racine (2003), "A Consistent Model Specification Test with Mixed Categorical and Continuous Data," *Journal of Econometrics*, forthcoming.

[21] Hummels, D. (2001), "Toward a Geography of Trade Costs," unpublished manuscript, Krannert School of Management, Purdue University.

[22] Hurvich, C.M., J.S. Simonoff and C.-L. Tsai (1998), "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society, Series B*, 60, 271-293.

[23] Imbens, G.W. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics & Statistics*, 86, 4-29.

[24] Leamer, E.E. and J. Levinsohn (1995), "International Trade Theory: The Evidence," in G.M. Grossman and K. Rogoff (eds.) *Handbook of International Economics*, vol. 3, Amsterdam: Elsevier.

[25] Li, Q. (1996), "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15, 261-274.

[26] Li, Q. and J. Racine (2004), "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14, 485-512.

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

[28] Millimet, D.L. and T. Osang (2005), "Do State Borders Matter for U.S. Intranational Trade? The Role of History and Internal Migration," *Canadian Journal of Economics*, forthcoming.

[29] N ©, Nonparametric software by Jeff Racine (http://www.economics.mcmaster.ca/racine/).

[30] Obstfeld, M. and K. Rogoff (2000), "The Six Major Puzzles in International Economics: Is There a Common Cause?" in B.S. Bernanke and J. Rotemberg (eds.) *NBER Macroeconomics Annual 2000*, Cambridge, MA: MIT Press.

[31] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge, Cambridge University Press.

[32] Persson, T. (2001), "Currency Unions and Trade: What Can the Data Say?," *Economic Policy*, 434-448.

[33] Racine, J. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics,* 119, 99-130.

[34] Ranjan, P. and J. Tobias (2005), "Bayesian and Gravity," *Journal of Applied Econometrics*, forthcoming.

[35] Redding, S. and A.J. Venables (2004), "Economic Geography and International Inequality," *Journal of International Economics*, 62, 53–82.

[36] Rose, A.K. (2000), "One Money, One Market: The Effect of Common Currencies on Trade," *Economic Policy*, 9-45.

[37] Rose, A.K. (2001), "Currency Unions and Trade: The Effect is Large," *Economic Policy*, 449-461.

[38] Rose, A.K. (2004), "The Effect of Common Currencies on International Trade: A Meta-Analysis," in V. Alexander, J. Melitz, and G.M. von Furstenberg (eds.) *Monetary Unions and Hard Pegs: Effects on Trade, Financial Development and Stability*, Oxford University Press.

[39] Rose, A.K. and E. van Wincoop (2001), "National Money as a Barrier to Trade: The Real Case for Currency Union," *American Economic Review*, 91, 386-390.

[40] Santos Silva, J.M.C. and S. Tenreyro (2005), "The Log of Gravity," *Review of Economics and Statistics*, forthcoming.

[41] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.

[42] Wall, H.J. (2000), "Gravity Model Specification and the Effects of the Canada-U.S. Border," Working Paper 2000-024A, Federal Reserve Bank of St. Louis.

[43] Wang, M.C. and J. Van Ryzin (1981), "A Class of Smooth Estimators for Discrete Estimation," *Biometrika*, 68, 301-309.

[44] Wolf, H.C. (2000), "Intranational Home Bias in Trade," *Review of Economics and Statistics,* 82, 555-63.

Figure A1: Forecasting – US Data

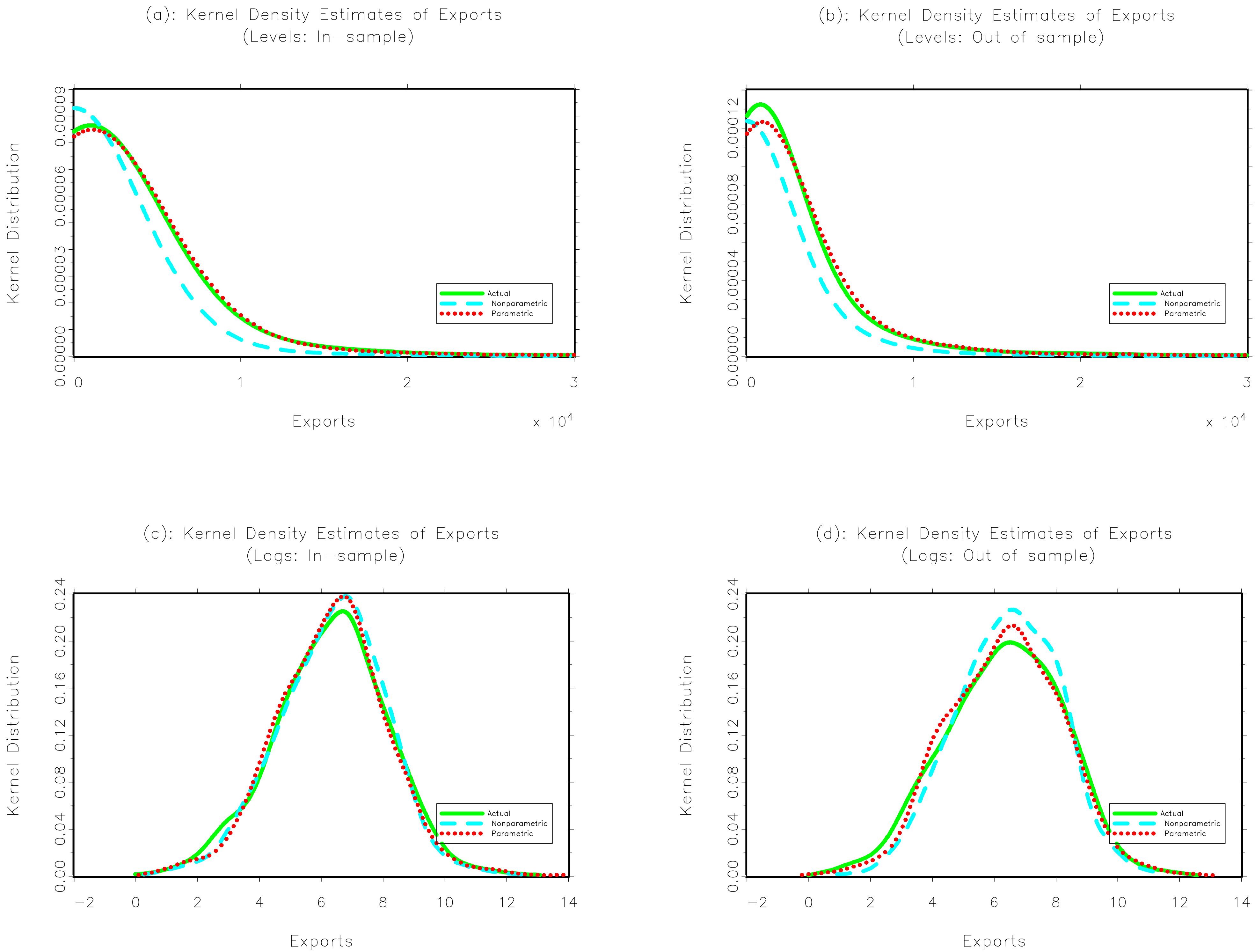


Figure A2: Forecasting – International Data

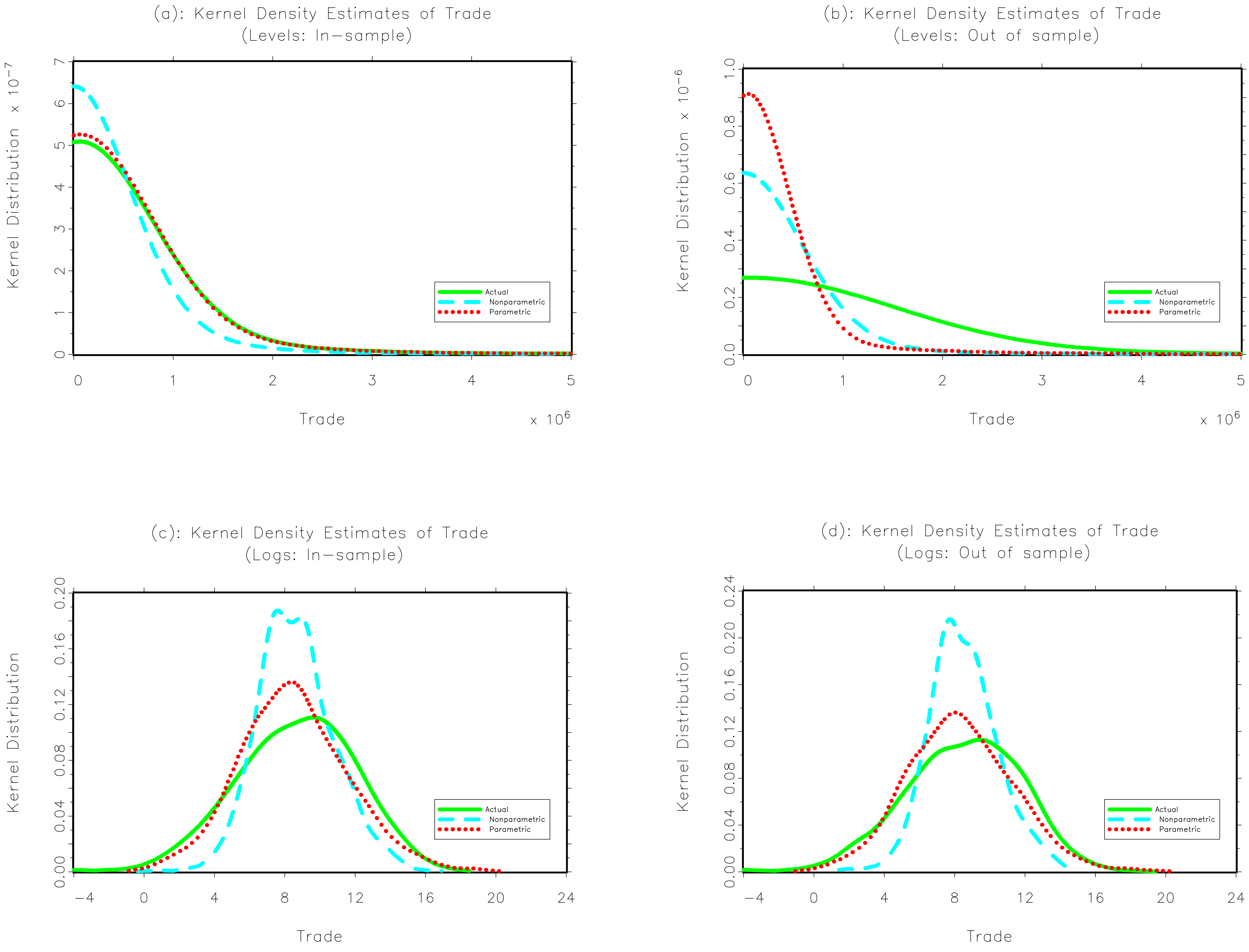


Table 1. Summary Statistics

Variable	Mean (Standard Deviation)		
	1993	1997	Full Sample
I. U.S. Subnational Data			
Shipments (mills 1996 US$)	2836.09	3187.23	3009.75
	(13720.72)	(15743.57)	(14755.13)
Distance (miles)	1156.76	1156.05	1156.41
	(705.51)	(713.01)	(709.14)
Gross State Product (mills 1996 US$)	145351.70	173753.70	159398.20
	(164287.40)	(191997.70)	(179073.00)
Remoteness	1.19	0.95	1.07
	(0.23)	(0.19)	(0.24)
Adjacent (1 = bordering states)	0.12	0.13	0.12
	(0.33)	(0.33)	(0.33)
Home (1 = intrastate)	0.02	0.02	0.02
	(0.15)	(0.15)	(0.15)
Observations	2137	2091	4228
II. Cross-Country Data			
Bilateral Trade			370826.80
			(4987185.00)
Distance (miles)			4403.70
			(2837.31)
Currency Union (1 = yes)			0.02
			(0.13)
Common Language (1 = yes)			0.22
			(0.42)
Regional Trade Agreement			0.02
			(0.13)
Adjacent (1 = bordering countries)			0.04
			(0.19)
Number Landlocked			0.34
			(0.53)
Number of Islands			0.33
			(0.54)
Observations			4315

Notes: Cross-country data are from Glick and Rose (2002) for 132 non-industrial countries and territories.

Table 2. Bandwidths

Variables By Type	Bandwidth	Upper Bound	Interpretation
I. Levels Model			
Unordered Categorical Variables:			
Home Dummy	8.01E-04	0.5000	0/1 weight
Exporter State x Time Effects	9.82E-01	0.9896	most likely irrelevant
Importer State x Time Effects	1.41E-17	0.9896	0/1 weight
Adjacent Dummy	6.11E-06	0.5000	0/1 weight
Continuous Variables:			
ln(Distance)	6.15E+08	∞	most likely linear
ln(Exporter Remoteness)	9.64E+05	∞	most likely linear
ln(Importer Remoteness)	1.65E+06	∞	most likely linear
II. Log-Linear Model			
Unordered Categorical Variables:			
Home Dummy	8.70E-02	0.5000	relevant
Exporter State x Time Effects	9.86E-01	0.9896	irrelevant
Importer State x Time Effects	1.30E-17	0.9896	0/1 weight
Adjacent Dummy	1.71E-01	0.5000	relevant
Continuous Variables:			
ln(Distance)	4.84E+06	∞	most likely linear
ln(Exporter Remoteness)	2.30E+06	∞	most likely linear
ln(Importer Remoteness)	1.37E+06	∞	most likely linear

Table 3. Estimates of the Distribution of the Border Effect

	Estimate	Standard Error	Home Bias	Tariff Equivalent
I. Levels Model				
Mean	0.6972	0.0021	2.0081	0.1047
10%	0.2890	0.0003	1.3350	0.0421
20%	0.6253	0.0021	1.8687	0.0934
30%	0.6994	0.0001	2.0126	0.1051
40%	0.7528	0.0010	2.1229	0.1135
50%	0.7879	0.0003	2.1987	0.1191
60%	0.8227	0.0004	2.2767	0.1247
70%	0.8463	0.0001	2.3309	0.1285
80%	0.8679	0.0021	2.3818	0.1320
90%	0.9527	0.0075	2.5926	0.1458
Parametric (PPML)	1.7155	0.0374	5.5595	0.2777
II. Log-Linear Model				
Mean	1.4218	0.0082	4.1444	0.2252
10%	0.6496	0.0163	1.9149	0.0972
20%	0.8605	0.0042	2.3643	0.1308
30%	1.1453	0.0235	3.1434	0.1778
40%	1.3476	0.0012	3.8483	0.2123
50%	1.4772	0.0106	4.3808	0.2350
60%	1.5599	0.0010	4.7582	0.2496
70%	1.6645	0.0167	5.2829	0.2684
80%	1.7791	0.0022	5.9247	0.2894
90%	2.1490	0.0029	8.5764	0.3593
Parametric (FE)	1.8166	0.0581	6.1509	0.2963

NOTE: Parametric result obtained by pooling 1993 and 1997 data. Home Bias = exp(coefficient). Tariff Equivalent = [exp(coefficient/(s-1)-1], for σ=8.

Table 4. Quartile Estimates for the Continuous Regressors (Elasticities)

Variable	Mean	Q1	Q2	Q3	Parametric
I. Levels Model					
ln(Distance)	-1.6532	-0.2700	-0.1027	0.0160	-0.2941
	(1.3590)	(0.1208)	(0.0659)	(0.0495)	(0.0262)
ln(Exporter Remoteness)	2.4338	-0.0102	0.2081	0.5706	31.8736
	(1.9184)	(0.0956)	(0.1442)	(0.3046)	(2.1555)
ln(Importer Remoteness)	40.4854	2.2655	6.7273	15.5794	22.3704
	(15.9634)	(1.0740)	(3.3978)	(5.5375)	(2.4212)
II. Log-Linear Model					
ln(Distance)	-0.5683	-0.9033	-0.5382	-0.2498	-0.4906
	(0.2348)	(0.3963)	(0.3320)	(0.2268)	(0.0207)
ln(Exporter Remoteness)	0.2956	-0.0631	0.3973	0.9596	17.1565
	(0.6987)	(0.6143)	(0.7564)	(1.0261)	(0.7453)
ln(Importer Remoteness)	38.5748	35.2241	38.8501	43.1152	11.5525
	(7.0605)	(9.3982)	(7.0701)	(7.2165)	(0.7449)

NOTES: Standard errors in parentheses obtained via bootstrapping except for the parametric model.

Table 5. Bandwidths

Variables By Type	Bandwidth	Upper Bound	Interpretation
I. Levels Model			
Unordered Categorical Variables:			
Currency Union Dummy	0.5000	0.5000	irrelevant
Country *i* Effects	0.1281	0.9928	relevant
Country *j* Effects	0.9710	0.9928	most likely irrelevant
Common Language Dummy	0.4044	0.5000	relevant
Regional Trade Agreement	0.0304	0.5000	relevant
Adjacent Dummy	0.4711	0.5000	most likely irrelevant
Ordered Categorical Variables:			
Number Landlocked	1.0000	1.0000	irrelevant
Number Islands	2.59E-18	1.0000	0/1 weight
Continuous Variables:			
ln(Distance)	1.16E+10	∞	most likely linear
II. Log-Linear Model			
Unordered Categorical Variables:			
Currency Union Dummy	0.5000	0.5000	irrelevant
Country *i* Effects	0.9924	0.9928	most likely irrelevant
Country *j* Effects	0.1652	0.9928	relevant
Common Language Dummy	0.3866	0.5000	relevant
Regional Trade Agreement	0.5000	0.5000	irrelevant
Adjacent Dummy	0.5000	0.5000	irrelevant
Ordered Categorical Variables:			
Number Landlocked	0.0621	1.0000	relevant
Number Islands	0.2274	1.0000	relevant
Continuous Variables:			
ln(Distance)	5.98779E+13	∞	most likely linear

Table 6. Estimates of the Distribution of the Currency Union Effect

	Estimate	Standard Error	Union Effect	Tariff Equivalent
I. Levels Model				
Mean	-6.86E-08	3.82E-07	1.00E+00	-9.81E-09
10%	-9.84E-08	5.96E-07	1.00E+00	-1.41E-08
20%	-1.41E-08	8.85E-07	1.00E+00	-2.01E-09
30%	-3.51E-09	1.45E-08	1.00E+00	-5.01E-10
40%	-1.57E-09	8.63E-08	1.00E+00	-2.24E-10
50%	-9.41E-10	2.45E-09	1.00E+00	-1.34E-10
60%	-4.54E-10	3.24E-10	1.00E+00	-6.49E-11
70%	-1.63E-10	2.20E-09	1.00E+00	-2.33E-11
80%	-6.70E-11	2.85E-10	1.00E+00	-9.57E-12
90%	-4.52E-12	5.54E-10	1.00E+00	-6.46E-13
Parametric (PPML)	0.6016	0.2852	1.8250	0.0897
II. Log-Linear Model				
Mean	8.79E-12	1.01E-11	1.00E+00	1.26E-12
10%	-5.98E-12	8.99E-11	1.00E+00	-8.54E-13
20%	-9.68E-13	9.58E-12	1.00E+00	-1.38E-13
30%	1.45E-12	4.07E-11	1.00E+00	2.07E-13
40%	4.94E-12	1.39E-12	1.00E+00	7.05E-13
50%	7.08E-12	4.55E-11	1.00E+00	1.01E-12
60%	9.86E-12	1.64E-10	1.00E+00	1.41E-12
70%	1.43E-11	1.71E-10	1.00E+00	2.05E-12
80%	1.98E-11	3.43E-10	1.00E+00	2.83E-12
90%	2.60E-11	3.57E-10	1.00E+00	3.72E-12
Parametric (FE)	1.0857	0.3228	2.9615	0.1678

NOTE: Union Effect = exp(coefficient). Tariff Equivalent = [exp(coefficient/(σ-1)-1], for σ=8.

Table 7. Quartile Estimates for the Continuous Regressors (Elasticities)

Variable	Mean	Q1	Q2	Q3	Parametric
I. Levels Model					
ln(Distance)	-2.2757	-0.5276	-0.1328	0.0785	-1.0502
	(0.4502)	(0.2256)	(0.1378)	(0.1851)	(0.0935)
II. Log-Linear Model					
ln(Distance)	-1.0486	-1.7166	-0.9639	-0.4227	-1.5229
	(0.4433)	(0.5208)	(0.3787)	(1.0309)	(0.0745)

NOTES: Standard errors in parentheses obtained via bootstrapping except for the parametric model.

Table 8. Li Tests for Equality of Distributions

I. Levels Model (US Data)	
Actual vs Predicted (Nonparametric)	
Test statistic:	10.4410
P-value:	0.0000
Actual vs Predicted (Parametric)	
Test statistic:	2.3521
P-value:	0.0093
Predicted (Nonparametric) vs Predicted (Parametric)	
Test statistic:	2.3917
P-value:	0.0083
II. Log-Linear Model (US Data)	
Actual vs Predicted (Nonparametric)	
Test statistic:	20.4762
P-value:	0.0000
Actual vs Predicted (Parametric)	
Test statistic:	3.6894
P-value:	0.0000
Predicted (Nonparametric) vs Predicted (Parametric)	
Test statistic:	13.0707
P-value:	0.0000
III. Levels Model (International Data)	
Actual vs Predicted (Nonparametric)	
Test statistic:	-63.0802
P-value:	0.0000
Actual vs Predicted (Parametric)	
Test statistic:	0.7025
P-value:	0.2399
Predicted (Nonparametric) vs Predicted (Parametric)	
Test statistic:	-21.7905
P-value:	0.0000
IV. Log-Linear Model (International Data)	
Actual vs Predicted (Nonparametric)	
Test statistic:	199.9783
P-value:	0.0000
Actual vs Predicted (Parametric)	
Test statistic:	39.8915
P-value:	0.0000
Predicted (Nonparametric) vs Predicted (Parametric)	
Test statistic:	48.2320
P-value:	0.0000

Note: P-values based on asymptotic normal approximation.

Table 9. Forecasting Accuracy

Model	Squared Correlation	Mean Squared Error	Mean Absolute Error	Mean Absolute Percentage Error
I. Levels Model (US Data)				
In-Sample				
Parametric	**0.9948**	**1.89E+06**	**535.5376**	**0.5179**
Nonparametric	0.9924	2.41E+06	792.8200	2.6771
Hold-out-Sample				
Parametric	**0.9042**	**1.26E+07**	**874.1655**	**0.5890**
Nonparametric	0.4165	8.26E+07	2507.8800	4.4318
II. Log-Linear Model (US Data)				
In-Sample				
Parametric	0.9367	0.2161	0.3484	0.1315
Nonparametric	**0.9484**	**0.1800**	**0.3214**	**0.0767**
Hold-out-Sample				
Parametric	**0.9246**	**0.2741**	**0.3941**	**0.0905**
Nonparametric	0.7924	0.7563	0.6213	0.2780
III. Levels Model (International Data)				
In-Sample				
Parametric	**0.9699**	2.90E+13	**126696**	**14589**
Nonparametric	0.9032	**1.20E+12**	308819	273200
Hold-out-Sample				
Parametric	**0.3904**	**3.14E+11**	**277999**	**35011**
Nonparametric	0.0173	4.23E+13	640534	1089949
IV. Log-Linear Model (International Data)				
In-Sample				
Parametric	**0.7021**	**4.0409**	**1.4509**	**0.3186**
Nonparametric	0.5387	6.5684	1.9086	0.4509
Hold-out-Sample				
Parametric	**0.5740**	**5.4284**	**1.6647**	**0.4010**
Nonparametric	0.2921	8.8187	2.2434	0.5651

NOTES: For definitions of the various models and accuracy measures, see text. US estimation based on N=2155; hold-out-sample on N=2073. International estimation based on N=2171; hold-out-sample on N=2137. Bold indicates most accurate forecast within the in-sample and hold-out-samples.

Table 10. Li Tests for Equality of Distributions: Forecasting

	In-Sample	Hold-Out-Sample
I. Levels Model (US Data)		
Actual vs Predicted (Nonparametric)		
Test statistic:	5.4290	9.1998
P-value:	0.0000	0.0000
Actual vs Predicted (Parametric)		
Test statistic:	0.4744	1.4457
P-value:	0.3175	0.0742
Predicted (Nonparametric) vs Predicted (Parametric)		
Test statistic:	3.2119	3.2313
P-value:	0.0000	0.0000
II. Log-Linear Model (US Data)		
Actual vs Predicted (Nonparametric)		
Test statistic:	1.8567	6.3075
P-value:	0.0318	0.0000
Actual vs Predicted (Parametric)		
Test statistic:	1.4144	1.6685
P-value:	0.0785	0.0480
Predicted (Nonparametric) vs Predicted (Parametric)		
Test statistic:	1.4957	4.3419
P-value:	0.0676	0.0000
III. Levels Model (International Data)		
Actual vs Predicted (Nonparametric)		
Test statistic:	-2.4704	-51.5293
P-value:	0.0067	0.0000
Actual vs Predicted (Parametric)		
Test statistic:	0.1211	-15.4482
P-value:	0.4503	0.0000
Predicted (Nonparametric) vs Predicted (Parametric)		
Test statistic:	-2.7559	89.5935
P-value:	0.0030	0.0000
IV. Log-Linear Model (International Data)		
Actual vs Predicted (Nonparametric)		
Test statistic:	109.1649	141.4883
P-value:	0.0000	0.0000
Actual vs Predicted (Parametric)		
Test statistic:	18.3535	14.4343
P-value:	0.0000	0.0000
Predicted (Nonparametric) vs Predicted (Parametric)		
Test statistic:	31.1365	49.1383
P-value:	0.0000	0.0000

Figure 1: Kernel Density Estimates of Exports
(US Data – Levels)

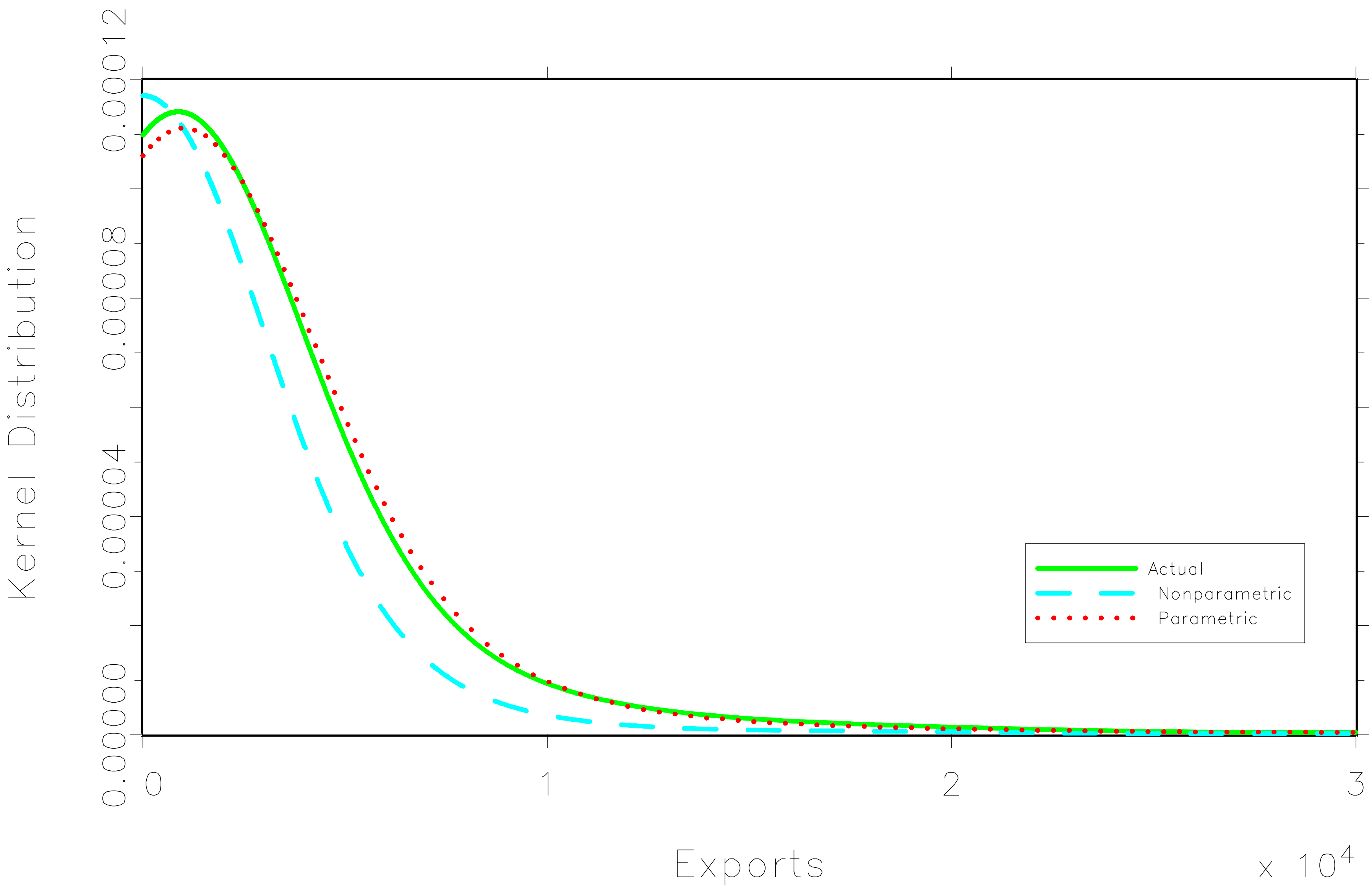
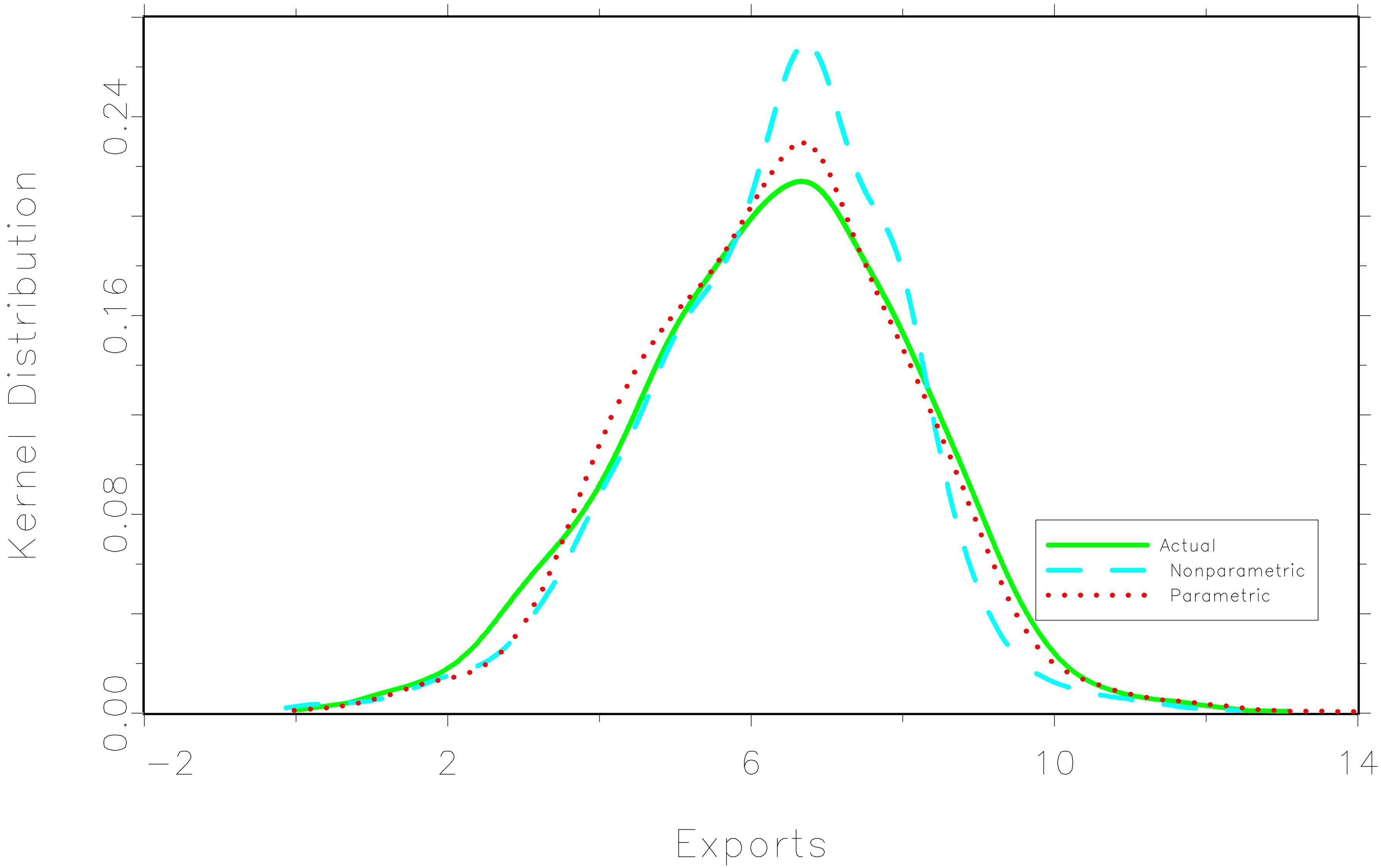


Figure 2: Kernel Density Estimates of Exports
(US Data – Logs)



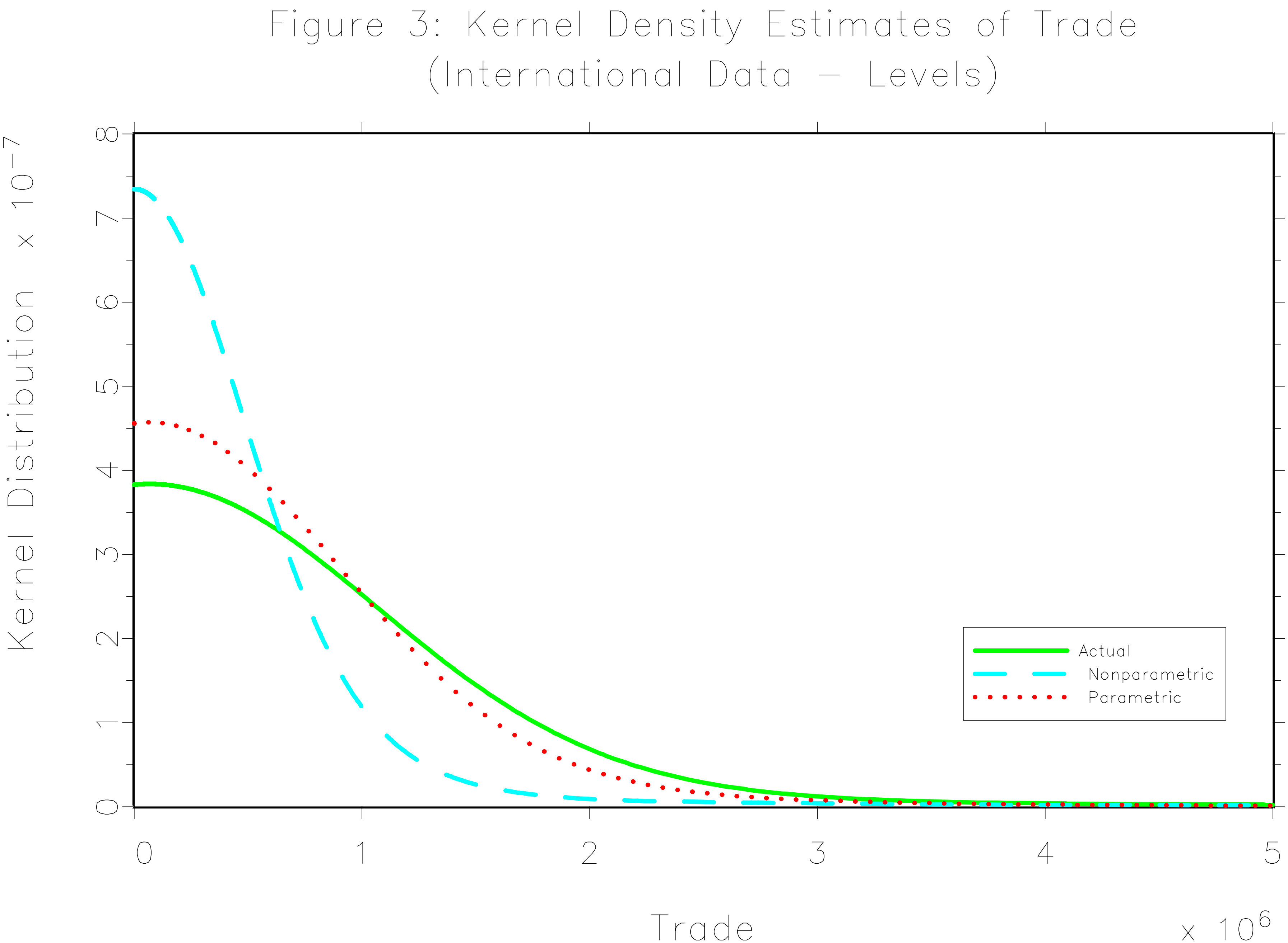


Figure 3: Kernel Density Estimates of Trade
(International Data – Levels)
Kernel Distribution x 10^-7
0
1
2
3
4
5
6
7
8
0
1
2
3
4
5
Trade
x 10^6
Actual
Nonparametric
Parametric

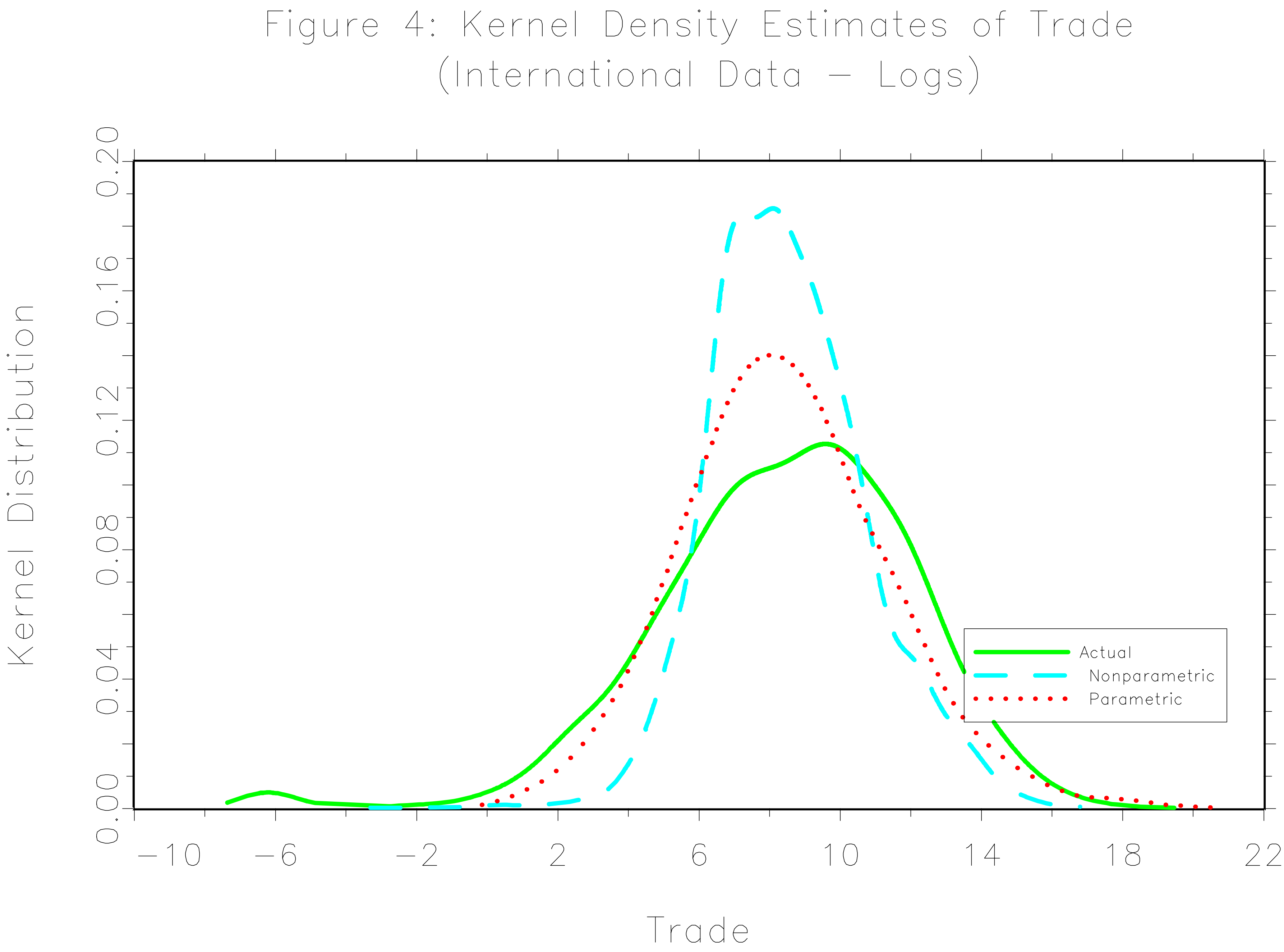


Figure 4: Kernel Density Estimates of Trade
(International Data – Logs)
Kernel Distribution
0.00
0.04
0.08
0.12
0.16
0.20
–10
–6
–2
2
6
10
14
18
22
Trade
Actual
Nonparametric
Parametric