Estimation of some nonlinear panel data models with both time-varying and time-invariant explanatory variables*

Bo E. Honoré[†] Michaela Kesina[‡]

July 2015

Abstract

The so-called "fixed effects" approach to the estimation of panel data models suffers from the limitation that it is not possible to estimate the coefficients on explanatory variables that are time-invariant. This is in contrast to a "random effects" approach, which achieves this by making much stronger assumptions on the relationship between the explanatory variables and the individual-specific effect. In a linear model, it is possible to obtain the best of both worlds by making random effects-type assumptions on the time-invariant explanatory variables while maintaining the flexibility of a fixed effects approach when it comes to the time-varying covariates. This paper attempts to do the same for some popular nonlinear models.

*This research was supported by the National Science Foundation and the Gregory C. Chow Econometric Research Program at Princeton University. We thank the referee, the associate editor and the editor for helpful suggestions.

†Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: honore@Princeton.edu.

‡ETH Zürich, Leonhardstrasse 21, 8092 Zürich.

1 Introduction

When analyzing panel data one is often interested in the coefficients on both time-varying and time-invariant explanatory variables. This generates a tension among competing assumptions, since the estimation of the coefficients on time-varying explanatory variables typically allows for so-called fixed effects, while estimating the coefficients on the time-invariant explanatory variables requires stronger assumptions. This, for example, is one of the issues[1] that motivated the work of Hausman and Taylor (1981). They studied linear panel data models and their approach is (essentially) to first estimate the coefficients on the time-varying explanatory variables under a set of fixed effects assumptions on the relationship between the individual-specific effect and the time-varying explanatory variables. They then combine this with a set of random-effects assumptions on the relationship between the time-invariant explanatory variables and the individual–specific effect.

Specifically, suppose that

$$y_{it} = x_{it}'\beta + z_i'\gamma + \alpha_i + \varepsilon_{it} \quad \text{for} \quad t = 1, \cdots, T_i \quad \text{and} \quad i = 1, \cdots, n.$$

By defining $\widetilde{\alpha}_i = z_i'\gamma + \alpha_i$, this becomes $y_{it} = x_{it}'\beta + \widetilde{\alpha}_i + \varepsilon_{it}$. This is a standard linear panel data model and β can be estimated using well-established methods. Since $y_{it} - x_{it}'\beta = z_i'\gamma + \alpha_i + \varepsilon_{it}$, the coefficient on the time-invariant explanatory variable, γ, can then be estimated by linear regression techniques such as least squares or instrumental variables methods, depending on the assumption on the relationship between $\alpha_i + \varepsilon_{it}$ and z_i.

The purpose of this paper is to generalize this idea to some nonlinear models of the type

$$y_{it} = g\left(x_{it}'\beta + z_i'\gamma + \alpha_i + \varepsilon_{it}\right) \tag{1}$$

where the aim is to provide estimates of both β and γ. Examples of nonlinear models where this

[1] The paper by Hausman and Taylor (1981) was also concerned with endogenous explanatory variables. The focus here is only on the introduction of time-invariant explanatory variables.

could be relevant include the discrete choice model, the censored regression model and the Poisson regression model, as well as more general multiplicative regression models. Clearly, in these models it is impossible to make progress without assumptions about the relationship between z_i and α_i. On the other hand, if one treats $z_i'\gamma + \alpha_i$ as a fixed effect, then β can be estimated using the methods proposed in, for example, Rasch (1960), Manski (1987), Honoré (1992) and Hausman, Hall, and Griliches (1984). The contribution here is to demonstrate how combining these estimates with stronger assumptions on the relationship between z_i and α_i, can be used to estimate γ.

Section 2 outlines the estimation procedures we investigate. They all take existing estimators for nonlinear fixed effects models as a starting point, and then impose additional assumptions to estimate the coefficients on the time-invariant covariates. In section 3, we investigate the practical performance of the proposed estimators. We do this by performing Monte Carlo studies and by applying our approach to empirical questions considered in the existing literature.

Multiplicative models have been used extensively to study trade flows. The explanatory variables often include both time-varying and time-invariant variables. Some of these are specific to the importer, some are specific to the exporter, and some are specific to the pair. In section 4, we discuss how to generalize our approach to allow for importer-specific, exporter-specific and pair-specific fixed effects. This follows the approach of Charbonneau (2013). Section 5 concludes.

2 Outline of Estimators

In this section, we outline the estimators investigated in this paper. We first consider the Poisson regression model studied in Hausman, Hall, and Griliches (1984). Since the conditional expectation is multiplicative, it is fairly easy to mimic the strategy for linear models to construct estimates for the time-invariant explanatory variables. The resulting approach applies to general multiplicative models and not just to Poisson regression models.

It is also possible to construct estimators for nonlinear models that are not multiplicative. In section 2.3 we propose a matching approach that works for many nonlinear models but requires a fairly strong statistical assumption on the errors. We also propose a method that is specific to the censored regression model but requires weaker assumptions.

2.1 Poisson Regression

Suppose that y_{it} is Poisson distributed with conditional mean $\exp(x_{it}'\beta + z_i'\gamma + \alpha_i)$. Hausman, Hall, and Griliches (1984) showed that the conditional distribution of $\{y_{it}\}_{t=1}^{T_i}$ given $(\{x_{it}\}_{t=1}^{T_i}, z_i, \alpha_i, \sum_{t=1}^{T_i} y_{it})$ depends on β but not on (z_i, α_i). This suggests estimating β by a conditional likelihood approach. Lancaster (2002) proved that the resulting estimator for β is identical to the estimator that treats all the $(z_i'\gamma + \alpha_i)$'s as parameters to be estimated.

Once β has been estimated, we can proceed by noting that $E\left[e^{-x_{it}'\beta} y_{it} \middle| (z_i, \alpha_i)\right] = \exp(z_i'\gamma + \alpha_i)$. So if $E\left[e^{\alpha_i} \middle| z_i\right]$ is constant, then γ can be estimated by nonlinear least squares where the dependent variable is $e^{-x_{it}'\widehat{\beta}} y_{it}$ and the mean function is specified as $\exp(z_i'\gamma + \gamma_0)$. Of course, γ can also be estimated by (pseudo) Poisson maximum likelihood. See Gourieroux, Monfort, and Trognon (1984).

Alternatively, we note that $E\left[\sum_{t=1}^{T_i} y_{it} \middle| \left(\{x_{it}\}_{t=1}^{T_i}, z_i, \alpha_i, T_i\right)\right] = \sum_{t=1}^{T_i} \exp(x_{it}'\beta + z_i'\gamma + \alpha_i)$, so

$$E\left[\left(\sum_{t=1}^{T_i} y_{it}\right)\left(\sum_{t=1}^{T_i} \exp\left(x_{it}'\beta\right)\right)^{-1} \middle| \left(\{x_{it}\}_{t=1}^{T_i}, z_i, \alpha_i, T_i\right)\right] = \exp\left(z_i'\gamma + \alpha_i\right),$$

which suggests using $\left(\sum_{t=1}^{T_i} y_{it}\right)\left(\sum_{t=1}^{T_i} \exp\left(x_{it}'\widehat{\beta}\right)\right)^{-1}$ as the dependent variable in a nonlinear least squares estimation where the mean function is specified as $\exp(z_i'\gamma + \gamma_0)$. This is intuitively appealing because $\left(\sum_{t=1}^{T_i} y_{it}\right)\left(\sum_{t=1}^{T_i} \exp\left(x_{it}'\widehat{\beta}\right)\right)^{-1}$ is the maximum likelihood estimator of $\exp(z_i'\gamma + \alpha_i)$ if all the $\exp\left(z_j'\gamma + \alpha_j\right)$'s and β were estimated jointly by maximum likelihood.

When γ is estimated by nonlinear least squares in a second step, the estimate can be interpreted without the assumption that $E\left[e^{\alpha_i} \middle| z_i\right]$ is constant. Specifically, consider the standard multiplicative model with fixed effects, $E\left[y_{it}|x_{it}, \phi_i\right] = \exp\left(x_{it}'\beta\right)\phi_i$. This model includes the fixed effects Poisson model as a special case. Without additional assumptions it is clearly impossible to include time-invariant explanatory variables in x_{it}. On the other hand, it may still be interesting to understand the relationship between ϕ_i and a set of time-invariant explanatory variables, z_i. For example, inspired by OLS one might consider the best linear approximation to $E\left[\phi_i|z_i\right]$, where "best" is defined in a mean squared error sense. However, since ϕ_i is nonnegative, it seems more natural to approximate $E\left[\phi_i|z_i\right]$ by functions of the form $\exp\left(c + z_i'\gamma\right)$. Nonlinear least squares estimation of γ as described in the previous paragraph estimates the parameters in the best approximation of this type (in a mean squared error sense).[2]

2.2 Connection to Multiplicative Models

The Poisson regression model is popular in part because maximum likelihood estimation of this model provides a consistent estimator in a more general multiplicative model, in which the mean of a dependent variable is modelled as the exponential of a linear index. See Gourieroux, Monfort, and Trognon (1984). This robustness carries over to applications of the maximum conditional Poisson likelihood estimator to panel data multiplicative models. Specifically, assume that $E\left[y_{it}|\left(\left\{x_{it}\right\}_{t=1}^{T_i}, z_i, \alpha_i, T_i\right)\right] = \exp\left(x_{it}'\beta\right)\exp\left(z_i'\gamma\right)e^{\alpha_i}$. Then the maximum conditional Poisson likelihood estimator that treats $z_i'\gamma + \alpha_i$ as a fixed effect will be consistent and asymptotically normal. See Cameron and Trivedi (2005). Afterwards, γ can be estimated as above, treating $e^{-x_{it}'\widehat{\beta}}y_{it}$

[2]Of course, the phrase "mean squared" assumes a distribution with which to calculate the mean. If one wants to refer to the cross-sectional distribution of (ϕ_i, z_i) then only one observation for each cross-sectional unit should be used (as in the previous paragraph). On the other hand, if "mean" refers to the distribution in the data set, then each individual should have an observation corresponding to each time period.

as a dependent variable with mean $\exp(z_i'\gamma + \gamma_0)$.

More generally, suppose that $E\left[y_{it} \middle| \left(\{x_{i\tau}\}_{\tau \leq t}, z_i, \alpha_i\right)\right] = \exp(x_{it}'\beta)\exp(z_i'\gamma)\, e^{\alpha_i}$. Following Wooldridge (1997), β then satisfies the conditional moment condition

$$E\left[e^{x_{is}'\beta} y_{it} - e^{x_{it}'\beta} y_{is} \middle| \left(\{x_{i\tau}\}_{\tau \leq s}, z_i, \alpha_i\right)\right] = 0 \tag{2}$$

for $s < t$. Equation (2) implies the unconditional moment conditions

$$E\left[\left(e^{x_{is}'\beta} y_{it} - e^{x_{it}'\beta} y_{is}\right) g\left(\{x_{i\tau}\}_{\tau \leq s}, z_i\right)\right] = 0 \tag{3}$$

for arbitrary functions g (provided that the moments exists). If one is willing to assume that (3) actually identifies β, then β can be estimated by generalized method of moments. Afterwards, γ can be estimated as above.

2.3 Other Limited Dependent Variable Models

The approach for estimating the Poisson and other multiplicative models depends crucially on the multiplicative structure of the conditional mean. We now discuss a more general strategy for estimating (β, γ) in (1). Again, the idea is to proceed in the same way as for the linear model and think of $z_i'\gamma + \alpha_i$ as an individual-specific fixed effect. We can then estimate β using some existing estimator. Then the question is how to estimate γ in a model of the type $y_{it} = g(w_{it} + z_i'\gamma + \alpha_i + \varepsilon_{it})$, where $w_{it} = x_{it}'\beta$. The complication is that a potential correlation between w_{it} and α_i will cause problems even if z_i is independent of $\alpha_i + \varepsilon_{it}$.

One potential solution is to identify pairs of observations indexed by it and js such that $w_{it} = w_{js}$. If z_i is independent of $\alpha_i + \varepsilon_{it}$ conditional on w_{it} , then such pairs will allow one to essentially treat w_{it} as a fixed effect for that pair. This strategy benefits from being general, but it is not easy to digest the assumption that z_i is independent of $\alpha_i + \varepsilon_{it}$ conditional on w_{it}. One case where it is satisfied is if z_i is a treatment that is assigned at random independently of all other characteristics of the individual. Such cases are quite special and we therefore also consider alternative approaches that are more specific to the models mentioned in the introduction.

2.3.1 Discrete Choice

Consider first the logit model $y_{it} = 1\left\{x_{it}'\beta + z_i'\gamma + \alpha_i + \varepsilon_{it} \geq 0\right\}$, where ε_{it} is *i.i.d.* and logistically distributed and independent of $(\{x_{it}\}, z_i, \alpha_i)$. Rasch (1960) proposed estimating β by maximizing the likelihood conditional on $\sum_{t=1}^{T_i} y_{it}$. The resulting estimator is consistent and asymptotically normal. Having estimated β, the question then becomes how to estimate γ in a model of the type $y_{it} = 1\left\{w_{it} + z_i'\gamma + \alpha_i + \varepsilon_{it} \geq 0\right\}$ where $w_{it} = x_{it}'\beta$. If α_i and z_i are independent conditional on w_{it} then

$$P\left(y_{it} = 1 | \, w_{it} = k, z_i\right) \lesseqgtr P\left(y_{js} = 1 | \, w_{js} = k, z_j\right) \qquad \text{iff} \qquad z_i'\gamma \lesseqgtr z_j'\gamma.$$

If β were known and $x_{it}'\beta$ discretely distributed, then the approach of Manski (1987) and Han (1987) suggests estimating γ by

$$\widetilde{\gamma} = \arg\max_g \binom{n}{2}^{-1} \sum_{i<j} w_{ij} \sum_{t=1}^{T_i} \sum_{s=1}^{T_j} 1\left\{x_{it}'\beta = x_{js}'\beta\right\} sign\left(y_{it} - y_{js}\right) sign\left((z_i - z_j)'g\right).$$

With β unknown and $x_{it}'\beta$ continuously distributed, one could then use

$$\widehat{\gamma} = \arg\max_g \binom{n}{2}^{-1} \frac{1}{h_n} \sum_{i<j} w_{ij} \sum_{t=1}^{T_i} \sum_{s=1}^{T_j} K\left(\frac{(x_{it} - x_{js})'\widehat{\beta}}{h_n}\right) sign\left(y_{it} - y_{js}\right) sign\left((z_i - z_j)'g\right), \quad (4)$$

where K is a kernel and h_n a bandwidth chosen so that in the limit, only pairs with $x_{it}'\beta = x_{js}'\beta$ will contribute to $\widehat{\gamma}$. If it is reasonable to assume that z_i and α_i are independent conditional on $(x_{i1}, \cdots, x_{iT})$, but not conditional on $x_{it}'\beta$, then it would be appropriate to match on $x_i = x_j$ rather than $x_{it}'\beta = x_{js}'\beta$. Although this should work in theory, we suspect that the high dimensional matching will make the estimator perform poorly in small samples.

Rasch's (1960) approach to estimating β does not work if ε_{it} is not *i.i.d.* logistic. One could argue that the logistic assumption is strong, and that it would be better to apply Manski's (1987) maximum score estimator in the first step. We do not pursue this here because Chamberlain (2010) has shown that if the support of the explanatory variables is bounded, then point identification is possible only in the logit case. Moreover, he showed that even if the support is unbounded, β can be estimated at the usual $\sqrt{n}$- rate only in the logit case.

We finally note that if z_i is independent of $v_{it} = \alpha_i + \varepsilon_{it}$, then it is possible to estimate γ by the method in Lewbel (2000). Specifically, suppose we know that the coefficient of the first element of z_i is positive. We can then write $y_{it} = 1\left\{x_{it}'\beta/\gamma_1 + \widetilde{z}_i'\widetilde{\gamma}/\gamma_1 + z_{i1} + v_{it}/\gamma_1 \geq 0\right\}$ with $z_i = (z_{i1}, \widetilde{z}_i')'$. Since z_{i1} is independent of v_{it}, we can then think of z_{i1} as a "special regressor" in the sense of Lewbel (2000) and construct estimators of β/γ_1 and $\widetilde{\gamma}/\gamma_1$ using his method. If the errors are logistically distributed, we can combine this with an estimate of β to recover γ_1 (and hence γ) as well. On the other hand, if the distribution of ε_{it} is left unspecified, then one can at most hope to identify β and γ up to scale. Lewbel's (2000) approach to estimation requires strong assumptions on the support and tail-behavior of the special regressor. In practice, it is probably useful to impose additional assumptions such as symmetry, and then follow the approach of Chen, Khan, and Tang (2013).

2.3.2 Censored Regression

Consider the fixed effects censored regression model given by $y_{it} = \max\{0, x_{it}'\beta + z_i'\gamma + \alpha_i + \varepsilon_{it}\}$. Honoré (1992) showed that with two observations for each individual, and with the error terms being i.i.d. for a given individual,[3]

$$\beta = \arg\min_b E[q\left(y_{i1}, y_{i2}, (x_{i1} - x_{i2})' b\right)], \tag{5}$$

[3]The assumption on the error terms made in Honoré (1992) allowed for very general serial correlation. However, for the discussion in this paper we will restrict ourselves to the i.i.d. assumption.

where

$$q(y_1, y_2, \delta) = \begin{cases} \Xi(y_1) - (y_2 + \delta)\xi(y_1) & \text{if } \delta \le -y_2; \\ \Xi(y_1 - y_2 - \delta) & \text{if } -y_2 < \delta < y_1; \\ \Xi(-y_2) - (\delta - y_1)\xi(-y_2) & \text{if } y_1 \le \delta; \end{cases}$$

and $\Xi(d)$ is given by[4] either $\Xi(d) = |d|$ or $\Xi(d) = d^2$. See also Arellano and Honoré (2001). We can therefore construct estimators for β by minimizing the sample analog of (5),

$$\mathcal{S}_n(b) \equiv \frac{1}{n} \sum_i w_i \sum_{s<t} q\left(y_{is}, y_{it}, (x_{is} - x_{it})' b\right),$$

where w_i is a weight that is assumed to be independent of $\{\varepsilon_{it}\}_{t=1}^{T_i}$.

Once β has been estimated, we can proceed as above and assume that z_i and α_i are independent conditional on $x_{it}'\beta$. Under random sampling this implies that $\alpha_i + \varepsilon_{it}$ will be distributed like $\alpha_j + \varepsilon_{js}$ conditional on $x_{it}'\beta = x_{js}'\beta$. We can apply a pairwise comparison estimator in the spirit of Honoré and Powell (1994) and Honoré and Powell (2005) to estimate γ

$$\widehat{\gamma} = \arg\min_g \binom{n}{2}^{-1} \frac{1}{h_n} \sum_{i<j} w_{ij} \sum_{t=1}^{T_i} \sum_{s=1}^{T_j} K\left(\frac{(x_{it} - x_{js})' \widehat{\beta}}{h_n}\right) q(y_{it}, y_{js}, (z_i - z_j)' g). \tag{6}$$

Again, if z_i and α_i are independent conditional on $(x_{i1}, \cdots, x_{iT})$, but not conditional on $x_{it}'\beta$, then one would match on $x_i = x_j$ rather than on $x_{it}'\beta = x_{js}'\beta$.

The censored regression model is an example for which it is possible to construct an estimator under weaker assumptions than those needed for $\widehat{\gamma}$ above. Suppose, for example, that x_{it} is composed of a set of binary variables. Then $x_{it}'\beta$ is bounded from below by $\sum_\ell \min\{\beta_\ell, 0\}$ and

$$\max\left\{y_{it} - x_{it}'\beta, -\sum_\ell \min\{\beta_\ell, 0\}\right\} = \max\left\{\max\left(z_i'\gamma + \alpha_i + \varepsilon_{it}, -x_{it}'\beta\right), -\sum_\ell \min\{\beta_\ell, 0\}\right\}$$
$$= \max\left\{z_i'\gamma + \alpha_i + \varepsilon_{it}, -x_{it}'\beta, -\sum_\ell \min\{\beta_\ell, 0\}\right\} = \max\left\{z_i'\gamma + \alpha_i + \varepsilon_{it}, -\sum_\ell \min\{\beta_\ell, 0\}\right\}$$

or

$$\max\left\{y_{it} - x_{it}'\beta + \sum_\ell \min\{\beta_\ell, 0\}, 0\right\} = \max\left\{z_i'\gamma + \widetilde{\alpha}_i + \varepsilon_{it}, 0\right\},$$

[4]Other convex loss functions, $\Xi(\cdot)$, could be used as well.

where $\widetilde{\alpha}_i = \alpha_i + \sum_\ell \min\{\beta_\ell, 0\}$. Depending on the assumptions made on the distribution of $\alpha_i + \varepsilon_{it}$ given z_i, this can be used to construct a number of estimators of γ. For example, with $\widetilde{y}_{it} = \max\left\{y_{it} - x_{it}'\widehat{\beta} + \sum_\ell \min\left\{\widehat{\beta}_\ell, 0\right\}, 0\right\}$, Powell's (1984) CLAD estimator would minimize

$$\sum_{i,t} \left|\widetilde{y}_{it} - \max\left\{0, z_i'g + g_0\right\}\right|$$

with respect to[5] (g, g_0). Honoré and Powell's (1994) pairwise comparison estimator would minimize

$$\sum_{i<j} \sum_{t=1}^{T_i} \sum_{s=1}^{T_j} q(\widetilde{y}_{it}, \widetilde{y}_{js}, (z_i - z_j)'g). \tag{7}$$

The estimation approaches discussed here imply a two-step procedure. Deriving the joint asymptotic distribution of such estimators is well understood when both steps are based on (generalized) method of moments. See, for example, Newey (1984). On the other hand, this does not cover the matched pairwise comparison estimator in equations (4) and (6). It is straightforward to derive the asymptotic variance for this estimator, mimicking the line of argument in, for example, Aradillas-Lopez, Honoré, and Powell (2007).

3 Practical Performance

This section illustrates the properties of some of the estimation strategies discussed above. We begin by investigating the performance of the Poisson estimation strategy in a setting where the data-generating process has the more general multiplicative structure. We present the results from a Monte Carlo study designed to compare the performance of the proposed two-step fixed effects estimator to the random effects Poisson estimator implemented in Stata and to a correlated random effects estimator (CRE) that included the average of the time-varying explanatory variables as a covariate. After discussing the Monte Carlo results, we illustrate the method in an empirical setting.

[5] The inclusion of g_0 is required to allow for a non-zero median of $\widetilde{\alpha}_i + \varepsilon_{it}$.

We also compare the matched pairwise comparison estimator (MPCE) defined in equation (6) for the censored regression model to the random effects and correlated random effects Tobit estimators implemented in Stata. Again, we do this in a Monte Carlo study as well as in an empirical example.

All the experiments use 400 replications and all bootstrap procedures are based on 200 draws.

3.1 Monte Carlo Evidence for Poisson Estimation of the Multiplicative Model

In this subsection we present the results from a Monte Carlo experiment that illustrates the Poisson regression approach discussed in section 2.1. Since the Poisson model is quite restrictive, we consider a more general data-generating process with the same structure on the conditional expectation as the Poisson regression model. In this design, which we label Design 1a, the dependent variable is

$$y_{it} = \exp\left(c + x_{it}'\beta + z_i'\gamma + \alpha_i + \varepsilon_{it}\right).$$

We have four time-varying covariates $x_{1it} \sim U[0,1] + \alpha_i + \frac{1}{2}(z_{i1} + z_{i2})$, $x_{2it} \sim U[0,1] \cdot \alpha_i$, $x_{kit} \sim U[0,1]$ for $k = 3,4$ with the true parameter vector $\beta = (0.5, 0, 0, 0)'$. We consider two time-invariant variables $z_{1i} \sim U[0,1]$, $z_{2i} \sim N(0,1)$ with the true parameter vector $\gamma = (0, 0.2)$. The constant, c, is set to -3. Finally, $\alpha_i \sim U[0.5, 1.5]$ and $\varepsilon_{it} \sim U[0,1]$.[6]

We consider two sample sizes, $n = 125$ and $n = 250$, and set $T = 4$. By microeconometric standards, these are modest sample sizes. One would expect the normal approximations to the distributions of the estimators and test statistics to be better in larger samples. On the other hand, there is no reason to think that biases in inconsistent estimators will depend on sample size.

We analyze the performance of three estimators. The first is the random effects Poisson maximum likelihood estimator (RE Poisson) implemented in Stata as `xtpoisson`. It imposes the restriction that e^{α_i} is gamma distributed. The second estimator is the correlated random effects

[6] With this design, the variance of $x_{it}'\beta + z_i'\gamma$ is approximately 50% higher than the variance of the combined error, $\alpha_i + \varepsilon_{it}$.

estimator (CRE Poisson), which includes individual-specific means of the time-varying variables in the regression equation (Mundlak (1978)) and estimates the regression equation with `xtpoisson`. The third estimator is the two-step fixed effects estimator (TSFE) proposed in section 2.1. First, we estimate β by maximum conditional pseudo-likelihood. This estimator is used to calculate $y_{it}e^{-x'_{it}\widehat{\beta}}$, which is then the dependent variable in a Poisson regression model with the mean $\exp(z_i\gamma + \gamma_0)$. As mentioned in section 2.1, we could also have estimated γ by nonlinear least squares. This would have given the estimates a nice semiparametric interpretation as the coefficients in a well-defined approximation. The reason why we do not do this is that Poisson pseudo maximum likelihood implicitly down-weights large observations. This makes it less likely that the estimates are driven by a few large outliers.

Table 1 presents the performance of the three estimators in terms of median bias and median absolute error. The RE Poisson is clearly biased for both β and γ. CRE Poisson, which is intended to alleviate the problem of the unobserved heterogeneity and to allow for some correlation between α and the covariates, performs well for the coefficients on the time-varying variables. However, it is more biased for the estimates on the time-invariant variables than RE Poisson, which ignores this dependence entirely. The TSFE performs well for both types of variables and outperforms the other two estimators. Increasing the sample size improves the TSFE, while the measures for the RE and CRE Poisson remain largely unchanged. This is to be expected as these estimators are inconsistent and the bias in the estimators tends to dominate their variability for the sample sizes considered.

Design 1a is somewhat special because the whole data-generating process is log-linear. To illustrate that the performance of the two-step fixed effects estimator is not driven by this, we also consider a slightly modified version of the previous design. Specifically, for the results in Table 2, we first generate a dependent variable as above. We then change it to zero with probability 0.1 and multiply it by 0.9^{-1} with probability 0.9. This preserves the conditional expectation of Design 1a, but the zeros prevent one from estimating the parameters of interest after taking logs. The results of this experiment with $n = 125$ are presented in Table 2. In terms of biases, the results are similar to those in Table 1, and – as one would expect – the additional noise increases the variance of the estimators compared to the same estimators in Table 1.

Of course, if the random effects Poisson model is correctly specified, then one would expect the random effects maximum likelihood estimator to be efficient. We therefore modify the experiment above to a model in which the random effects specification is correct. We keep β, γ, and the distribution of (x_{it}, z_i), and draw the random effect from a $\Gamma(2, 2)$ distribution independently of (x_{it}, z_i). We label this Design 1b. Table 3 presents the results for the three estimators with $n = 125$. As expected, RE Poisson displays a lower median absolute error than TSFE and CRE Poisson.

Table 4 shows the mean and standard deviation for the p–values for testing the true hypothesis that $(\gamma_1, \gamma_2) = (0, 0.2)$ for the three data-generating processes and for each of the three estimators.[7] As expected, the TSFE approach proposed here performs well for all three designs. The random effects approaches only perform well in Design 1b.

3.2 Empirical Illustration for Poisson Estimation of the Multiplicative Model: Estimation of Doctor Consultations

As an illustration of the Poisson regression model, we estimate a model of doctor consultations using data from the German Socio-Economic Panel. We use the data in Winkelmann (2004).[8] The data set is an unbalanced panel over the years 1995–1999. Detailed information about the data

[7]If the tests are valid, then the p-values should be $U(0, 1)$-distributed with mean 0.5 and standard deviation approximately 0.29.

[8]This data set was downloaded from http://qed.econ.queensu.ca/jae/2004-v19.4/winkelmann/.

is given in that paper. Here we focus on individuals with completed education. Thus, we use individuals who are between 30 and 60 years old. For each individual, we calculate the maximum level of education and delete years in which the education level is lower than that maximum. In our estimation procedure, only individuals who are observed for at least two years contribute. Keeping these units, this leads to a data set of $5,214$ individuals and a total number of observations of $20,969$. Our dependent variable of interest is the number of doctor consultations. We estimate a Poisson model with mean $\exp\left(x_{it}'\beta + z_i'\gamma + \alpha_i\right)$.

The time-invariant variables in the model are the individual's age in 1995 (Age), years of schooling (Education), and a gender dummy (Male). Age2 is the square of a person's age in each year; it is time-varying. Married indicates whether a person is married. Household size is the number of persons living in the household. Sport is one if a person engages in sports at least once per week. Good health and Bad health are binary variables indicating whether the self-reported health status is good or bad. The binary variable Social assistance is one if the person receives welfare payments and Income is the logarithm of the income per month. The individual's working status is measured by three dummy variables: Full-time employed, Part-time employed, and Unemployed. The dummy variables Winter, Spring, and Fall are one depending on the quarter of the year in which the interview took place. We additionally include year dummies.

Table 4 shows the results of using the random effects Poisson estimator (RE Poisson), the correlated random effects Poisson estimator (CRE Poisson) and the two-step fixed effects estimator (TSFE) proposed here. Men go to the doctor less frequently than women. Better education leads to fewer doctor consultations. Good health decreases the number of doctor visits; bad health has the opposite effect. Financial factors such as income and welfare payments exhibit no significant impact on the number of doctor visits. Also, employment status does not tend to affect the number of visits to the doctor. The sign of the significant coefficients is the same but the size of the parameter estimators differs slightly between the three estimation approaches.

For all estimators we report the bootstrapped standard errors. We also ran the RE and CRE Poisson with nonrobust analytical standard errors. The bootstrap standard errors were twice as big as the analytical standard errors, suggesting misspecification of the random effects model. To further investigate this, we estimated the variance of the difference in two estimators by the bootstrap test. We used this to test the null hypothesis that two estimators estimate the same parameter vector for three pairs. The χ^2 test statistics for testing the equality of RE and CRE Poisson, RE Poisson and TSFE, and CRE Poisson and TSFE are 230.73, 275.33, and 179.75, respectively. These values clearly reject the corresponding null hypothesis. We interpret this as evidence against the random effects assumptions.

3.3 Monte Carlo Evidence for Censored Regression

The matched pairwise comparison estimator (MPCE) requires that z_i and α_i be independent conditional on $x_{it}'\beta$. This is a strong assumption. One of the simplest designs in which it is satisfied is when α_i is independent of (x_{it}, z_i). We therefore first illustrate the performance of the MPCE estimator in that case. Specifically, in Design 2a $x_{kit} \sim N(0,1)$ for $k = 1, \cdots, 4$ and $x_{5it} = x_{1it} \cdot x_{2it}$. There are two independent elements of z_i with $z_{i1} \sim N(0,1) + \frac{1}{2}\sum_{t=1}^{4} x_{1,it}$ and $z_{2i} \sim \chi^2(1)$. The true parameters are $\beta = (2, -0.5, 0, 0, 0)'$ and $\gamma = (2, -2)'$. Finally $\alpha_i = \left(\chi^2(1) - 1\right)/\sqrt{2}$ and $\varepsilon_{it} = 3\left((1+\alpha_i)\varpi_{it} + \alpha_i\right)$, where ϖ_{it} is a mixture of an $N\left(-\frac{3}{10}, \frac{1}{4}\right)$ and an $N\left(\frac{6}{5}, \frac{11}{5}\right)$; the mixing probabilities are 0.8 and 0.2, respectively.[9] The constant, c, is set to 4. With this specification, the censoring probability is approximately 38% and the variance of $x_{it}'\beta + z_i'\gamma$ is slightly smaller than that of $\alpha_i + \varepsilon_{it}$. Specifically, the R^2 in an infeasible regression using the y's before censoring would

[9] This implies that the mean of ϖ_{it} is 0 and the variance 1. This particular mixed normal distribution was chosen because it was also used in the paper by Honoré and Powell (1994).

be approximately 0.47. This is a random effects design in which the individual-specific effect and the transitory errors are both quite nonnormal, and the main question addressed by the Monte Carlo experiment is the sensitivity of the MPCE and MLE estimators to nonnormality. The choice of the distribution of the explanatory variables is arbitrary except that we avoid joint normality, since some estimators perform unexpectedly with joint normality. As before, the sample sizes are $n = 125$ and $n = 250$ with $T = 4$. For the MPCE we use $\Xi(d) = d^2$.

Implementing the MPCE requires a bandwidth and a kernel. Since semi- and nonparametric estimators typically are much more sensitive to the choice of bandwidth than the kernel, we consider only estimators with the standard normal density as the kernel and ignore the possibility that the small sample performance of the estimator would improve if a bias-reducing kernel was used.

We apply Silverman's rule of thumb (equation 3.31 in Silverman (1986)), $h_n = 0.9\widehat{\sigma}n^{-1/5}$, as the point of departure for the bandwidth. Since asymptotic normality of the estimator (centered at the truth) will require undersmoothing, this bandwidth is likely to be too large. We therefore refer to this as "large bandwidth." We also consider bandwidths of $0.9\widehat{\sigma}n^{-1/5}/2$ ("medium") and $0.9\widehat{\sigma}n^{-1/5}/4$ ("small") as well as bandwidths that are ten times smaller ("very small") and ten times larger than the medium bandwidth ("very large").

Table 6 presents the median bias and median absolute deviation (from the median) for the estimators of γ_1 and γ_2 for the five bandwidths when $n = 125$. Overall, the estimators are surprisingly insensitive to the choice of bandwidth, although the results are consistent with the prediction that larger bandwidths give larger bias but lower variation. To evaluate the usefulness of the asymptotic normal distribution, we also calculate the p-value for the Wald tests of the true hypothesis that $(\gamma_1, \gamma_2) = (2, -2)$. The second part of Table 6 shows the resulting means and standard deviations, when the variance of the estimator is estimated using the analytical expression and using a simple bootstrap. Most of the p-values have mean and standard deviation close to 0.5 and 0.29, which suggests that the asymptotics provide a good approximation to the behavior of the estimator. The exception is that the analytic estimator of the variance seems to perform poorly with the very small bandwidth. Because of the low sensitivity of the estimator to the bandwidth, we focus on only the medium bandwidth, the normal kernel and bootstrap standard errors when discussing the MPCE in the remaining part of this subsection.

We compare the performance of the MPCE with two other estimators: the panel data random effects Tobit (RE Tobit) estimator implemented in Stata's `xttobit` command and the correlated random effects estimator, which augments the regression equation by the individual-specific means of the time-varying variables. The augmented equation is then estimated with the `xttobit` command. The resulting estimator is labelled the CRE Tobit. Table 7 presents the performance of the three estimators in terms of median bias and median absolute error for sample sizes $n = 125$ and $n = 250$. All standard errors are calculated using a bootstrap. Assuming normality – as in the RE Tobit – results in biased estimators for the time-invariant variables. The CRE Tobit performs even worse for the time-invariant variables and shows only little improvements for the time-varying parameter estimates. In contrast, the MPCE performs well for both types of variables.

The first line in Table 8 presents summary statistics for the p-values for testing $(\gamma_1, \gamma_2) = (2, -2)$. While the mean and standard deviation for the p-values of the MPCE are close to those of a $U[0, 1]$-distribution, the bias in the point estimates results in severely downwardly biased p-values of the RE Tobit and CRE Tobit.

The nonnormality of the error components leads to the biases of the RE Tobit. On the other hand, this estimator will be consistent and asymptotically efficient under normality. To understand the trade-off between the two approaches, we carry out another experiment, which we label Design 2b, to investigate their relative performance under normality. We keep the construction of x_{kit} and z_i and the parameter values β and γ, but we now assume that $\alpha_i \sim N(0, 25)$ and $\varepsilon_{it} \sim N(0, 25)$.

As before, the censoring probability is approximately 38% and the R^2 calculated on the basis of the uncensored observations is 0.49. We do this experiment for $n = 125$ and $T = 4$. Table 9 reports the numerical measures for the MPCE, RE and CRE Tobit estimators, and the second line of Table 8 presents summary statistics for the p-values for the Wald test for testing $(\gamma_1, \gamma_2) = (2, -2)$. As expected, the biases of the RE and CRE Tobit estimators disappear and the moments of the p-values are closely approximated by those of the uniform distribution. According to the measures in Table 9, the RE Tobit outperforms the MPCE but the differences are small. Based on these experiments, we conclude that all three estimators work well under normality, with the RE Tobit being slightly superior, as one would expect. On the other hand, deviations from normality result in biased RE and CRE Tobit estimates, while the MPCE continues to perform well.

We next turn to a fixed effects setting. The MPCE assumes independence of z_i and α_i conditional on $x_{it}'\beta$. One setting where this assumption is reasonable is a randomized experiment. We therefore consider a second design where we treat the time-invariant variable z_i as a binary treatment indicator. In Design 3a the time-varying variables are $x_{kit} \sim N(0, 1) + \alpha_i$ for $k = 1, 2$, $x_{kit} \sim N(0, 1)$ for $k = 3, 4$ and $x_{5it} = x_{1it} \cdot x_{2it}$ where the time-invariant individual-specific effect is generated as $\alpha_i \sim N(0, 1)$. Note that, unlike the previous design, some of the time-varying variables are correlated with the unobserved heterogeneity. This leads to a data-generating process in the spirit of fixed effects models. The time-invariant variable z_i is a binary ("treatment") and independent of everything else with $p = \frac{1}{2}$. The true parameters are $\beta = (1, 0, 0, 0, 0)'$ and $\gamma = 2$. The constant, c, is -1. Finally, $\varepsilon_{it} = \sqrt{2} \cdot \exp(\alpha_i - 1) \cdot \nu_{it}$ where $\nu_{it} \sim N(0, 1)$. The population R^2 in an infeasible regression using the uncensored y's is around 0.63 and the censoring probability is 53%. The experiments are conducted for $n = 125$ and $T = 4$.

The first line of Table 10 shows the mean and standard deviation of the p-values for testing the hypothesis that $\gamma = 2$. The moments of the p-values for the MPCE match the theoretical $U[0, 1]$-

distribution, whereas, as expected, the p-values based on RE and CRE Tobit are downwardly biased. The bias is also evident in Table 11. The RE Tobit is clearly biased for all variables that are correlated with the unobserved heterogeneity. The CRE Tobit performs better than the RE Tobit for the time-varying variables. However, the bias and median absolute errors of the CRE Tobit are worse for the time-invariant variable compared to the RE Tobit. In contrast, the MPCE performs well for both types of variables and clearly outperforms the RE and CRE Tobit.

The correlation between (x_{1it}, x_{2it}) and α_i, and the fact that $(\alpha_i, \varepsilon_{it})$ is not normally distributed both lead to inconsistency of `xttobit`. We therefore also perform the experiment in a random effects setting that satisfies the parametric assumptions for `xttobit`. Specifically, we keep the design from above but eliminate the dependence between (x_{1it}, x_{2it}) and α_i, and set $\varepsilon_{it} \sim N(0, 1)$. The population R^2 in an infeasible regression using the uncensored y's is around 0.60 and the censoring probability is approximately 50%. The second line of Table 10 shows the mean and standard deviation of the p-values for testing the hypothesis that $\gamma = 2$. The moments of the p-values match those of the theoretical distribution for all the estimators. Table 12 shows the numerical measures for the estimators. They are fairly similar for all three approaches. As expected, the RE Tobit maximum likelihood estimator performs best, but the MPCE is not far behind.

3.4 Empirical Illustration for Censored Regression: Earnings and Race

As an illustration of the censored regression model, we estimate a simple model of earnings using longitudinal data from the Social Security Administration before and after the Civil Rights Act of 1964. The application builds heavily upon (especially) Chay (1995), Chay and Honoré (1998), Hu (2002) and Chay and Powell (2001), and we refer the reader to those papers for a detailed description of the data. We have an unbalanced panel on 9,225 individuals over the years 1963, 1964, 1970, and 1971, yielding a total number of observations of 33,720.

The basic model is

$$\log(w_i) = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{Education}_i + \gamma_3 \text{Black}_i + \beta \text{Age2}_{it} + \sum_t \beta_t \text{Year}_t + \sum_t \beta_{Black\times t} (\text{Black}_i \times \text{Year}_t) + \alpha_i + \varepsilon_{it}, \quad (8)$$

where w denotes earnings, which is censored from above due to the Social Security cap. Age is time-invariant because of the inclusion of year dummies. Thus, the age variable is the individual's age in 1963. Age-squared (Age2), on the other hand, is time-varying. Education measures the years of schooling and is time-invariant along with the binary variable black. We include year dummies and set β_{1963} to zero. To assess the evolution of the black-white earnings gap, we include interactions between the race dummy and the year dummies. γ_0 denotes a constant. In this data set, wages are censored from above with a censoring probability between 0.46 and 0.49. Twelve percent of the individuals are black. On average, the education level and wages of blacks are lower than those of whites. The average gap in wages between blacks and whites decreased over time, from -0.47 to -0.34.

Table 13 displays the results of estimating (8) using the random effects Tobit estimator (RE Tobit), the correlated random effects (CRE) Tobit estimator and the MPCE proposed here. The standard errors are based on the bootstrap. For the MPCE, we use $\Xi(d) = d^2$ and employ the small, medium and large bandwidths. The estimates using the different bandwidths are fairly similar, a result that is in line with the results in the previous subsection. The estimators of the RE Tobit, the CRE Tobit[10] and the MPCE differ in size but have the same sign. Our results confirm the findings in the literature. Older people have higher wages. More education has a positive impact

[10] In this example, there is little difference between the random effects and the correlated random effects estimator. In general, the difference is that the correlated random effects estimator includes the average of the time-varying explanatory variables as a covariate. In this case only the average of age-squared is added since the average of the year-race interactions are perfectly correlated with race.

on wages. Blacks have lower wages than whites. The biggest difference in the parameter estimates is that our estimates are smaller in magnitude than those obtained by the random effects approach.

The last column of Table 13 presents the results from estimating the coefficients on the time-invariant explanatory variables by minimizing (7). The point estimates resulting from this are in line with those obtained by the other estimators, but the standard errors are substantially larger. This is not surprising. This estimator is based on weaker assumptions than the others and the construction of $\widetilde{y}_{it}$ adds to the already very heavy censoring.

4 Empirical Application and Extension

As a final empirical application, we use an extension of the multiplicative model proposed above to estimate a model of bilateral trade flows between countries. Empirical gravity models of the type considered here have been of great interest in the trade literature for more than a decade. See, for example, Head and Mayer (2013) for a survey.

We employ the same data as Helpman, Melitz, and Rubinstein (2008).[11] It is a balanced panel covering the years 1980 to 1989 and it reports bilateral trade flows between 158 countries. The data set also contains information on country and country-pair specific variables such as distance, sharing a border, having the same legal system, belonging to the WTO, etc.

For this application, the cross-sectional unit denotes the country pair ij, where the first index i denotes the exporting country and the second index j denotes the importing country. The dependent variable of interest, y_{ijt}, is the trade flow from country i to j in year t, and we estimate a multiplicative model with a linear index that depends on time-varying covariates, x_{ijt}, time-

[11] This data set was downloaded from http://scholar.harvard.edu/melitz/publications.

invariant covariates, z_{ij}, and a country-pair specific fixed effect, α_{ij},

$$E\left[y_{ijt}|\left\{x_{ijt}\right\}_{t=1}^{T}, z_{ij}, \alpha_{ij}\right] = \exp\left(x_{ijt}'\beta + z_{ij}'\gamma + \alpha_{ij}\right).$$

Although x_{ijt} and z_{ij} are indexed by both i and j, our setup allows for the possibility that they do not vary across all of the dimensions. For example, x_{ijt} will contain a set of time dummies.

In the application, the time-varying variables are as follows: CU is a binary variable indicating whether two countries have the same currency or have a 1:1 exchange rate. FTA is binary; it is one if both countries are part of a free trade agreement. WTOnone and WTOboth are one if neither or both countries are members of the WTO, respectively. We also include year dummies for 1981–1989. The time-invariant covariates are variables that are specific to the country pair ij. Distance is the logarithm of the geographical distance between the capitals. Border is binary; it is one if both countries share a common border. Island is one if both countries in the pair are islands, and zero otherwise. Similarly, Landlock is one if both countries in the pair are landlocked, i.e., have no access to the coast. Colony is a binary variable that is one if one of the countries in the pair colonialized the other one. Legal is a binary variable and equals one if both countries in the pair have the same legal system. Similarly, Language is binary and equals one if both countries in the pair speak the same language. Religion is a variable measuring the similarity in the shares of Protestants, Catholics and Muslims in the countries in the pair. A higher number indicates a bigger similarity.

In the data, more than half of all observations have zero trade flows. Trading countries are closer to each other in distance. They are more likely to share a common border and less likely to be completely landlocked. Former colonial relations are slightly more frequent among countries that are trading with each other. Among trading countries, the probability of both countries being in the WTO is higher than in the sample in general.

We first estimate the parameters of the model by the two-step method of moments estimator outlined in Section 2.1. The results are presented in column three of Table 14. The standard errors are based on 200 bootstrap replications, where a bootstrap "unit" is an exporter-importer pair. For comparison, we also estimate the model with the Poisson random effects estimator implemented in Stata and with the correlated random effects estimator as proposed by Mundlak (1978), which includes individual-specific means of the time-varying variables as additional covariates in the random effects specification. These results are reported in columns one and two of Table 14. The difference in the results from the three estimation strategies is much more pronounced for the coefficients on the time-invariant coefficients than for the time-varying coefficients, and the most striking finding is the insensitivity of the latter to the estimation method.

We also used Stata's implementation of the fixed effects maximum conditional Poisson likelihood estimator to estimate the parameters for the time-varying explanatory variables. As expected, this resulted in the same point estimates for those parameters as those presented in column three of Table 14. As a result, they are not included in the table. It is worth noting, however, that different methods for calculating the standard errors resulted in substantially different results. It is not surprising that calculating robust standard errors makes a difference, since there is no reason to believe that trade flows are literally Poisson distributed. On the other hand, the analytic robust standard errors and those calculated by the bootstrap should be comparable even if the model is completely misspecified (provided that all theoretical moments are finite). We attribute the difference to the fact that CU is a dummy variable with mean less than 0.01. One can therefore think of all observations of one as outliers. In other words, all of the variation in CU is driven by outliers.[12] It is well known that robust standard errors can be sensitive to outliers. We therefore

[12]Some of the other dummy variables also have very low means. The reason why we focus on Currency Union is that the estimated standard errors for its coefficient were much more sensitive to whether it was calculated analytically

also estimated the model without CU. The results are presented in columns four to six of Table 14.

It is natural to modify the gravity model estimated above to allow for exporter and importer country-specific effects among the time-invariant components. This suggests a model of the type

$$E\left[y_{ijt}|\left(\{x_{ijt}\}_{t=1}^{T},z_{ij},w_{1i},w_{2j},\eta_i^1,\eta_j^2,\alpha_{ij}\right)\right]=\exp\left(x'_{ijt}\beta\right)\exp\left(z'_{ij}\gamma\right)\exp\left(w'_{1i}\delta_1+w'_{2j}\delta_2\right)e^{\eta_i^1}e^{\eta_j^2}e^{\alpha_{ij}},$$

where w_{1i} and w_{2j} are exporter- and importer-specific time-invariant explanatory variables such as whether the country is landlocked. η_i^1, η_j^2 and α_{ij} are unobserved country- and pair-specific effects to be discussed below.

The parameters, β, γ and δ can then be estimated in three steps.

4.1 Estimation of β

The effect of x_{ijt}, β, can be estimated as before. This requires no assumptions on η_i^1, η_j^2 and α_{ij}. Specifically, we treat the term $\exp\left(z'_{ij}\gamma\right)\exp\left(w'_{1i}\delta_1+w'_{1i}\delta_2\right)e^{\eta_i^1}e^{\eta_j^2}e^{\alpha_{ij}}$ as a combined unobserved pair-specific effect, and we then have the moment condition

$$E\left[y_{ijt}\exp\left(x'_{ijs}\beta\right)-y_{ijs}\exp\left(x'_{ijt}\beta\right)\middle|\,\{x_{ijt}\}_{t=1}^{T},z_{ij},w_{1i},w_{2j}\right]=0.$$

This allows us to estimate β. In practice we do this by calculating the maximum conditional likelihood estimator proposed by Hausman, Hall, and Griliches (1984).

4.2 Estimation of γ

In the estimation of β, $\exp\left(z'_{ij}\gamma\right)\exp\left(w'_{1i}\delta_1+w'_{1i}\delta_2\right)e^{\eta_i^1}e^{\eta_j^2}e^{\alpha_{ij}}$ was treated as a combined unobserved pair-specific effect. Distinguishing the effect of the pair-specific time-invariant variables, z_{ij}, from the other country-specific time-invariant variables, w_{1i} and w_{2j}, requires further assumptions. We therefore make a random effects-type assumption on the relationship between $e^{\alpha_{ij}}$ and

or by bootstrap.

$\left(z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2\right)$. Specifically, we assume that $E\left[e^{\alpha_{ij}} \middle| z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2\right]$ is constant. Since we have not specified the scale of $e^{\eta_i^1}$ and $e^{\eta_j^2}$, there is no loss of generality in assuming that this constant is one, $E\left[e^{\alpha_{ij}} \middle| z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2\right] = 1$. This yields

$$E\left[y_{ijt} \exp\left(-x'_{ijt}\beta\right) \middle| z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2, \alpha_{ij}\right] = \exp\left(z'_{ij}\gamma\right) \exp\left(w'_{1i}\delta_1 + w'_{2j}\delta_2\right) e^{\eta_i^1} e^{\eta_j^2} e^{\alpha_{ij}}.$$

This relationship also holds unconditional on α_{ij}, which gives the moment condition

$$E\left[y_{ijt} \exp\left(-x'_{ijt}\beta\right) \middle| z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2\right] = \exp\left(z'_{ij}\gamma\right) \exp\left(w'_{1i}\delta_1 + w'_{2j}\delta_2\right) e^{\eta_i^1} e^{\eta_j^2}.$$

While this does not depend on α_{ij}, we still have to deal with the country-specific effects $e^{\eta_i^1}$ and $e^{\eta_j^2}$. To eliminate those, we follow the approach first proposed in Charbonneau (2013). Consider a second exporter-importer pair, kl, where i, j, k and ℓ all differ from each other. Multiplying y_{ijt} with $y_{k\ell t}$ and rearranging yields

$$\begin{aligned} & E[\left(y_{ijt} \exp\left(-x'_{ijt}\beta\right) \exp\left(-z'_{ij}\gamma\right)\right) \\ & \qquad \left(y_{k\ell t} \exp\left(-x'_{k\ell t}\beta\right) \exp\left(-z'_{k\ell}\gamma\right)\right) \mid \left(z_{ij}, w_{1i}, w_{2j}, \eta_i^1, \eta_j^2\right), \left(z_{k\ell}, w_{1k}, w_{2\ell}, \eta_k^1, \eta_\ell^2\right)] \\ = & \left(\exp\left(w'_{1i}\delta_1 + w'_{2j}\delta_2\right) e^{\eta_i^1} e^{\eta_j^2}\right) \left(\exp\left(w'_{1k}\delta_1 + w'_{2\ell}\delta_2\right) e^{\eta_k^1} e^{\eta_\ell^2}\right). \end{aligned}$$

We can next eliminate the η's by subtracting the same expression, but now obtained from the pairs y_{kjt} and $y_{i\ell t}$. This gives the moment condition

$$\begin{aligned} 0 = & E[\left(y_{ijt} \exp\left(-x'_{ijt}\beta\right) \exp\left(-z'_{ij}\gamma\right)\right) \left(y_{k\ell t} \exp\left(-x'_{k\ell t}\beta\right) \exp\left(-z'_{k\ell}\gamma\right)\right) \\ & - \left(y_{kjt} \exp\left(-x'_{kjt}\beta\right) \exp\left(-z'_{kj}\gamma\right)\right) \left(y_{i\ell t} \exp\left(-x'_{i\ell t}\beta\right) \exp\left(-z'_{i\ell}\gamma\right)\right) \\ & \qquad |z_{ij}, z_{k\ell}, z_{kj}, z_{i\ell}, w_{1i}, w_{2j}, w_{1k}, w_{2\ell}, \eta_i^1, \eta_j^2, \eta_k^1, \eta_\ell^2] \end{aligned}$$

and hence

$$\begin{aligned} 0 \quad = \quad & E[\left(y_{ijt}\exp\left(-x_{ijt}'\beta\right)\exp\left(-z_{ij}'\gamma\right)\right)\left(y_{k\ell t}\exp\left(-x_{k\ell t}'\beta\right)\exp\left(-z_{k\ell}'\gamma\right)\right) \\ & -\left(y_{kjt}\exp\left(-x_{kjt}'\beta\right)\exp\left(-z_{kj}'\gamma\right)\right)\left(y_{i\ell t}\exp\left(-x_{i\ell t}'\beta\right)\exp\left(-z_{i\ell}'\gamma\right)\right) \\ & \qquad\qquad |z_{ij}, z_{k\ell}, z_{kj}, z_{i\ell}, w_{1i}, w_{2j}, w_{1k}, w_{2\ell}]. \end{aligned} \tag{9}$$

This allows us to construct many moment conditions for γ once β has been estimated. In order to use moment conditions that uniquely identify γ and that can be solved relatively efficiently, we base our estimation on a set of moment conditions that are also in the spirit of Hausman, Hall, and Griliches (1984). With $y_{ijk\ell t} = y_{ijt}\exp(-x_{ijt}'\beta)\ y_{k\ell t}\exp(-x_{k\ell t}'\beta)$ and $z_{ijk\ell} = z_{ij} + z_{k\ell}$, (9) becomes

$$0 = E\left[y_{ijk\ell t}\exp\left(-z_{ijk\ell}'\gamma\right) - y_{kji\ell t}\exp\left(-z_{kji\ell}'\gamma\right)\middle|\, z_{ij}, z_{k\ell}, z_{kj}, z_{i\ell}, w_{1i}, w_{2j}, w_{1k}, w_{2\ell}\right]. \tag{10}$$

Now consider maximizing the objective function

$$E\left[y_{ijk\ell t}\log\Lambda\left(\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right) + y_{kji\ell t}\log\Lambda\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)\right],$$

where $\Lambda(\cdot)$ is the logistic CDF and $\Lambda'(\cdot) = \Lambda(\cdot)(1-\Lambda(\cdot))$. The first-order condition is

$$\begin{aligned} 0 \quad &= \quad E\left[y_{ijk\ell t}\frac{\Lambda'\left(\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)}{\Lambda\left(\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)}\left(z_{ijk\ell} - z_{kji\ell}\right) + y_{kji\ell t}\frac{\Lambda'\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)}{\Lambda\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)}\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)\right)\right] \\ &= \quad E\left[\left(y_{ijk\ell t}\ \Lambda\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right) - y_{kji\ell t}\Lambda\left(\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right)\right)\left(z_{ijk\ell} - z_{kji\ell}\right)\right] \\ &= \quad E\left[\left(y_{ijk\ell t}\ \exp\left(-z_{ijk\ell}' g\right) - y_{kji\ell t}\exp\left(-z_{kji\ell}' g\right)\right)\frac{1}{\exp\left(-z_{ijk\ell}' g\right) + \exp\left(-z_{kji\ell}' g\right)}\left(z_{ijk\ell} - z_{kji\ell}\right)\right]. \end{aligned}$$

(10) implies that this is 0 at $g = \gamma$. Moreover, the objective function has the same mathematical structure as the textbook logit log-likelihood function. Hence, it is concave. As a result $g = \gamma$ is the unique maximizer. We therefore estimate γ by minimizing

$$-\sum_{t,i\neq j\neq k\neq \ell} y_{ijk\ell t}\log\Lambda\left(\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right) + y_{kji\ell t}\log\Lambda\left(-\left(z_{ijk\ell} - z_{kji\ell}\right)' g\right). \tag{11}$$

Note that the summation over t affects just the y's. Also note that there is no reason to both compare the exporter-importer pair ij to the exporter-importer pair kl and the exporter-importer pair kl to the exporter-importer pair ij. These insights can lead to a significant reduction in the computational demands for estimating γ.

4.3 Estimation of δ

Finally, if $E\left[e^{\eta_i^1}\middle| w_{1i}, w_{2j}\right]$ and $E\left[e^{\eta_j^2}\middle| w_{1i}, w_{2j}\right]$ are constant, then

$$E\left[y_{ijt}\exp\left(-x_{ijt}'\beta\right)\exp\left(-z_{ij}'\gamma\right)\middle| w_{1i}, w_{2j}\right] = \exp\left(\delta_0 + w_{1i}'\delta_1 + w_{2j}'\delta_2\right). \tag{12}$$

This allows us to estimate δ_1 and δ_2 (if β and γ have already been estimated).

We implement the estimation of δ_1 and δ_2 by first replacing β and γ with $\widehat{\beta}$ and $\widehat{\gamma}$ and then applying the Poisson maximum likelihood estimator to (12). The first-order condition for this is

$$\sum\left(y_{ijt}\exp\left(-x_{ijt}'\widehat{\beta}\right)\exp\left(-z_{ij}'\widehat{\gamma}\right) - \exp\left(w_{1i}'\delta_1 + w_{2j}'\delta_2\right)\right)\begin{pmatrix} w_{1i} \\ w_{2j} \end{pmatrix} = 0 \tag{13}$$

which corresponds to (12).

4.4 Empirical Findings

Table 15 presents the results of estimating β, γ and δ by the approach discussed in sections (4.1)-(4.3). The first two columns present the results using Stata's canned random effects routine. We list the time-invariant, pair-specific explanatory variables first because the results for those are the most striking. The parameter estimates for our approach are typically much smaller in magnitude. Intuitively, this suggests that the fixed effects capture some of the relationships that would otherwise have been attributed to the covariates. We also note that the random effects estimates suggest that there is a negative relationship between trade-flows and the exporter and importer countries having the same legal system. Our results suggest the more plausible positive relationship.

Deriving the estimates of the standard errors of the three-step estimator of (β, γ, δ) is complicated, because the second step involves sums of the form $\sum_{t,i\neq j\neq k\neq \ell}$ which cannot easily be rewritten as U–statistics. We also cannot perform a simple bootstrap, because a bootstrap sample will pair some countries as importers with themselves as exporters. We therefore estimate the standard errors by subsampling. Specifically, for each subsample, we randomly choose $n_0 = 20$ countries and estimate the model based on the trade flow between those countries using all time periods. The standard errors based on 400 subsamples are then scaled down by $\sqrt{\frac{n_0(n_0-1)}{n(n-1)}}$. This will work for the estimation of the standard errors of the estimators for β and γ, but not for δ. The reason is that when δ is estimated, we treat the η_i^1s (and η_j^2s) as random effects. Each observation in the subsample will contribute $(n_0 - 1)$ observations that all share the same η_i^1. In the original sample, each i will contribute $(n-1)$ observations that all share the same η_i^1. So at this point, the correlation structure is different in the subsample and in the original sample.[13] As a result, we do not present standard errors associated with the estimator of δ. One way that one could have tackled them would have been to observe that the β and the γ are estimated on the basis of $n\,(n-1)$ importer-exporter combinations. On the other hand, δ is based on a sample in which there are only n draws of the η_i^1s (and η_j^2s). So asymptotically, one could justify ignoring the estimation error in β and γ when estimating the standard errors of coefficients on w_{1i} and w_{2j}. We do not pursue this, because it takes the error structure in the model more seriously than we want to.

To further illustrate the potential of the approach, we perform a simple Monte Carlo simulation. To mimic the empirical application, we generate data with $n = 158$ and $T = 10$ from the model

$$y_{ijt} = \exp\left(x_{ijt}'\beta\right)\exp\left(z_{ij}'\gamma\right)\exp\left(w_{1i}'\delta_1 + w_{2j}'\delta_2\right)e^{\eta_i^1}e^{\eta_j^2}e^{\alpha_{ij}}e^{\varepsilon_{ijt}}.$$

In order to avoid large outliers in the distribution of y_{ijt}, we focus on bounded explanatory variables. The fixed effects, α_{ij}, η_i^1 and η_j^2, are all uniformly distributed on the interval $(0, 1)$. There are two pair-specific time-varying explanatory variables. x_{ijt1} follows a standard uniform distribution and

[13] If we used only a subset of the time periods, then the α_{ij}'s would have caused the same problem for the estimation of the standard errors of γ.

x_{ijt2} is generated as a standard uniform plus α_{ij}. There are also two time-invariant pair-specific explanatory variables. z_{ij1} follows a standard uniform distribution and z_{ij2} is generated as a standard uniform plus $\eta_i^1 + \eta_j^2$. Finally, there are two i-specific and two j-specific time-invariant explanatory variables. These are all also drawn from uniform distributions. We draw the errors, ε_{ijt}, from a normally distributed stationary $AR(1)$ with $V[\varepsilon_{ijt}] = 1$ and $cov[\varepsilon_{ijt-1}, \varepsilon_{ijt}] = 0.3$. The results presented in Table 16 are consistent with our expectations. Our approach tends to produce estimators with lower bias, whereas the likelihood-based estimators have less variability.

In the simulation, we also mimicked the subsampling method for constructing standard errors for the estimation of β and γ. The standard deviations of the t-statistics for testing the true parameter values for those four parameters lies between 1.08 and 1.13. We see this as evidence in support of the subsampling method for calculating the standard errors of β and γ.

5 Concluding Remarks

When studying panel data, one is often interested in the effects of both time-invariant and time-varying explanatory variables. This results in a tension: the effects of time-varying covariates can be investigated under fixed effects assumptions, whereas it is essentially only possible to estimate the effect of the time-invariant covariates under stronger random effects assumptions. It is well understood how to combine these assumptions, and proceed to estimation in linear panel data models. This paper attempts to do the same for some popular nonlinear models.

The Poisson regression model is used to illustrate how a specific approach can be used to estimate models with weak assumptions on the unobservables in a specific model. This approach applies to a general class of multiplicative models. We also introduce a more generic matching method. This leads to a more general approach than in a random effects model, but it still makes fairly demanding assumptions on the relationship between the covariates and the individual-specific effects. For both cases, we present the results of Monte Carlo experiments that illustrate the usefulness of the new methods in situations where a random effects model would result in misleading parameter estimates and where a pure fixed effects approach would make inference on the time-invariant explanatory variables impossible.

We use this to estimate Poisson regression panel data models for the number of doctor visits using data from the German Socio-Economic Panel. While point estimates are not extremely different from those obtained using a random effects model, a simple specification test soundly rejects the random effects model in favor of our more general assumptions. For the censored regression model we re-examine the effect of the Civil Rights Act of 1964 on earnings using top-coded Social Security earnings. In this case the main parameter of interest is the coefficient on the interaction of the time dummies and race. This can be estimated using an existing fixed effects approach, and the main added-value of our approach is to also allow for estimation of the coefficients on education, birth-cohort and race without making a full set of random effects assumptions.

We finally apply our approach to estimation of panel data models with both time-varying and time-invariant covariates to a fairly standard gravity model of trade. This part of the paper makes both a methodological and an empirical contribution by allowing for both exporter-importer pair-specific effects and pure exporter- or importer-specific effects. This section leaves open the question of how to derive an analytic expression for the asymptotic variance of the parameter vector. We leave this for future research.

References

Aradillas-Lopez, A., B. E. Honoré, and J. L. Powell (2007): "Pairwise Difference Estimation with Nonparametric Control Variables," *International Economic Review*, 48(4), 1119–1158.

Arellano, M., and B. E. Honoré (2001): "Panel Data Models. Some Recent Developments," *Handbook of Econometrics*, 5, Elsevier Science, Amsterdam.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chamberlain, G. (2010): "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78, 159–168.

Charbonneau, K. B. (2013): "Multiple Fixed Effects in Theoretical and Applied Econometrics," Ph.D. thesis, Princeton University.

Chay, K. Y. (1995): "Evaluating the Impact of the 1964 Civil Rights Act on the Economic Status of Black Men using Censored Longitudinal Earnings Data," Manuscript, Princeton University.

Chay, K. Y., and B. E. Honoré (1998): "Estimation of Semiparametric Censored Regression Models: An Application to Changes in Black-White Earnings Inequality during the 1960s," *Journal of Human Resources*, 33(1), 4–38.

Chay, K. Y., and J. L. Powell (2001): "Semiparametric Censored Regression Models," *Journal of Economic Perspectives*, 15(4), 29–42.

Chen, S., S. Khan, and X. Tang (2013): "Informational Content of Special Regressors in Heteroskedastic Binary Response Models," ERID Working Paper No. 144.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(3), 701–720.

Han, A. (1987): "Nonparametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909–38.

Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6), pp. 1377–1398.

Head, K., and T. Mayer (2013): "Gravity Equations: Workhorse, Toolkit, and Cookbook," CEPR Discussion Papers 9322, C.E.P.R. Discussion Papers.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123(2), 441–487.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–565.

Honoré, B. E., and J. L. Powell (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64(1-2), 241–78.

——— (2005): "Pairwise Difference Estimation of Nonlinear Models," in *Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg.*, ed. by D. W. K. Andrews, and J. H. Stock, chap. 22, pp. 520–553. Cambridge University Press.

Hu, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70(6), pp. 2499–2517.

Lancaster, T. (2002): "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69(3), 647–666.

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97(1), 145–177.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–62.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.

Newey, W. K. (1984): "A method of moments interpretation of sequential estimators," *Economics Letters*, 14(23), 201 – 206.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

Rasch, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Pædagogiske Institut, Copenhagen.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

Winkelmann, R. (2004): "Health care reform and the number of doctor visits: an econometric analysis," *Journal of Applied Econometrics*, 19(4), 455–472.

Wooldridge, J. M. (1997): "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption," *Econometric Theory*, 13, 667–678.

Table 1: Multiplicative model Design 1a − RE Poisson, CRE Poisson, and TSFE, Multiplicative data-generating process with fixed effects

	β_1	β_2	β_3	β_4	γ_1	γ_2
			$n = 125$			
			RE Poisson			
Median bias	0.433	0.166	−0.001	−0.003	−0.211	−0.216
Median absolute error	0.433	0.166	0.046	0.041	0.211	0.216
			CRE Poisson			
Median bias	−0.002	−0.003	−0.001	−0.002	−0.333	−0.338
Median absolute error	0.038	0.041	0.044	0.042	0.333	0.338
			TSFE			
Median bias	−0.004	−0.001	−0.004	−0.001	0.007	0.003
Median absolute error	0.036	0.040	0.045	0.042	0.068	0.028
			$n = 250$			
			RE Poisson			
Median bias	0.439	0.172	0.002	0.004	−0.218	−0.220
Median absolute error	0.439	0.172	0.031	0.030	0.218	0.220
			CRE Poisson			
Median bias	−0.001	0.002	−0.002	−0.001	−0.341	−0.341
Median absolute error	0.026	0.027	0.030	0.030	0.341	0.341
			TSFE			
Median bias	−0.004	0.002	−0.002	−0.002	0.005	0.001
Median absolute error	0.027	0.025	0.030	0.029	0.050	0.021

Table 2: Multiplicative model Design 1a with zeros − RE Poisson, CRE Poisson, and TSFE, Multiplicative data-generating process with fixed effects, $n = 125$

	β_1	β_2	β_3	β_4	γ_1	γ_2
			RE Poisson			
Median bias	0.437	0.178	0.007	0.007	−0.218	−0.222
Median absolute error	0.437	0.178	0.061	0.060	0.218	0.222
			CRE Poisson			
Median bias	0.001	0.000	0.007	0.009	−0.344	−0.341
Median absolute error	0.065	0.062	0.062	0.068	0.344	0.341
			TSFE			
Median bias	−0.002	−0.001	0.007	0.006	0.004	0.000
Median absolute error	0.063	0.061	0.064	0.067	0.089	0.041

Table 3: Multiplicative model Design 1b − RE Poisson, CRE Poisson, and TSFE, Poisson data-generating process with random effects, $n = 125$

	β_1	β_2	β_3	β_4	γ_1	γ_2
			RE Poisson			
Median bias	−0.000	−0.006	0.029	0.038	−0.013	−0.010
Median absolute error	0.125	0.151	0.182	0.186	0.186	0.069
			CRE Poisson			
Median bias	0.015	0.004	0.017	0.047	−0.006	0.004
Median absolute error	0.203	0.172	0.197	0.218	0.201	0.101
			TSFE			
Median bias	0.018	0.001	0.007	0.042	0.012	−0.023
Median absolute error	0.198	0.179	0.195	0.219	0.202	0.101

Table 4: P-values of the hypothesis test $H_0 : \gamma_1 = 0, \gamma_2 = 0.2$, $n = 125$
Bootstrap standard errors

	RE Poisson		CRE Poisson		TSFE	
	Mean	Sd	Mean	Sd	Mean	Sd
Design 1a	0.000	0.004	0.000	0.000	0.497	0.300
Design 1a with zeros	0.001	0.006	0.000	0.005	0.458	0.300
Design 1b	0.501	0.295	0.505	0.285	0.503	0.287

Table 5: Results for RE Poisson, CRE Poisson, and TSFE

	RE Poisson	CRE Poisson	TSFE
Dependent variable: Number of doctor visits			
Age $\times 10^{-1}$	-0.283	-0.276	-0.102
	(0.212)	(0.204)	(0.379)
Male	-0.194^{***}	-0.164^{***}	-0.241^{***}
	(0.036)	(0.039)	(0.051)
Education	-0.020^{***}	-0.018^{**}	-0.018^{**}
	(0.006)	(0.006)	(0.007)
Age2 $\times 10^{-3}$	0.420	0.514	0.231
	(0.233)	(0.319)	(0.404)
Married	0.007	0.003	0.004
	(0.044)	(0.070)	(0.071)
Household size	-0.025	-0.013	-0.016
	(0.014)	(0.025)	(0.026)
Sport	0.001	-0.027	-0.027
	(0.032)	(0.039)	(0.038)
Good health	-0.504^{***}	-0.444^{***}	-0.439^{***}
	(0.026)	(0.033)	(0.030)
Bad health	0.648^{***}	0.600^{***}	0.598^{***}
	(0.031)	(0.032)	(0.031)
Social assistance	0.055	0.035	0.036
	(0.067)	(0.073)	(0.075)
Income	0.050	0.012	0.011
	(0.038)	(0.049)	(0.046)
Full-time employed	-0.117^{*}	-0.047	-0.051
	(0.048)	(0.063)	(0.063)
Part-time employed	-0.071	0.003	0.001
	(0.050)	(0.058)	(0.057)
Unemployed	-0.051	-0.028	-0.030
	(0.047)	(0.050)	(0.050)
Winter	-0.001	-0.005	-0.004
	(0.035)	(0.040)	(0.042)
Spring	-0.003	-0.010	-0.010
	(0.030)	(0.034)	(0.037)
Fall	-0.034	-0.036	-0.038
	(0.074)	(0.077)	(0.089)
Number of observations	20,969	20,969	20,969

Bootstrap standard errors based on 200 replications in parentheses.
* $p < 0.05$, ** $p < 0.01$, $^{***}p < 0.001$. All regressions include year dummies.

Table 6: Design 2a − MPCE for different bandwidths, $n = 125$

	Very small		Small		Medium		Large		Very large	
	Panel A: Performance of estimators									
	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
MB	−0.015	−0.108	0.011	−0.125	0.026	−0.119	0.052	−0.105	0.322	−0.111
MAD	0.441	0.650	0.400	0.630	0.408	0.615	0.410	0.582	0.394	0.602
	Panel B: P-values of the hypothesis test $H_0 : \gamma_1 = 2, \gamma_2 = -2$									
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Analytical	0.916	0.124	0.522	0.282	0.462	0.284	0.454	0.283	0.449	0.280
Bootstrap	0.555	0.268	0.489	0.279	0.477	0.283	0.474	0.283	0.480	0.281

Notes: MB: Median bias, MAD: Median absolute deviation from median.

Table 7: Censored regression model Design 2a − RE Tobit, CRE Tobit, and MPCE
Nonnormally distributed error components

	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2
				$n = 125$			
				RE Tobit			
Median bias	0.481	−0.123	−0.009	0.009	0.030	0.565	−0.552
Median absolute error	0.481	0.223	0.192	0.183	0.206	0.565	0.613
				CRE Tobit			
Median bias	0.485	−0.143	−0.006	0.008	0.021	0.596	−0.564
Median absolute error	0.487	0.233	0.197	0.178	0.207	0.598	0.634
				MPCE			
Median bias	0.037	−0.026	−0.019	−0.031	0.006	0.026	−0.119
Median absolute error	0.248	0.273	0.244	0.239	0.280	0.405	0.573
				$n = 250$			
				RE Tobit			
Median bias	0.481	−0.123	−0.009	0.009	0.030	0.565	−0.552
Median absolute error	0.481	0.223	0.192	0.183	0.206	0.565	0.613
				CRE Tobit			
Median bias	0.485	−0.143	−0.006	0.008	0.021	0.596	−0.564
Median absolute error	0.487	0.233	0.197	0.178	0.207	0.598	0.634
				MPCE			
Median bias	0.045	−0.019	−0.036	−0.008	0.063	−0.009	−0.118
Median absolute error	0.204	0.196	0.181	0.164	0.226	0.278	0.459

Table 8: P-values of the hypothesis test $H_0 : \gamma_1 = 2, \gamma_2 = -2$, $n = 125$
Bootstrap standard errors

	RE Tobit		CRE Tobit		MPCE	
	Mean	Sd	Mean	Sd	Mean	Sd
Design 2a	0.265	0.251	0.310	0.268	0.477	0.283
Design 2b	0.501	0.298	0.501	0.298	0.494	0.292

Table 9: Censored regression model Design 2b − RE Tobit, CRE Tobit, and MPCE,
Normally distributed error components, $n = 125$

	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2
				RE Tobit			
Median bias	0.019	0.004	0.001	0.018	0.017	−0.006	−0.003
Median absolute error	0.183	0.170	0.163	0.174	0.186	0.140	0.182
				CRE Tobit			
Median bias	0.027	−0.018	−0.011	-0.001	0.002	0.007	0.004
Median absolute error	0.197	0.210	0.183	0.188	0.215	0.180	0.182
				MPCE			
Median bias	0.004	0.015	−0.011	0.013	0.019	0.012	−0.009
Median absolute error	0.239	0.249	0.228	0.235	0.218	0.168	0.221

Table 10: P-values of the hypothesis test $H_0 : \gamma = 2$, $n = 125$
Bootstrap standard errors

	RE Tobit		CRE Tobit		MPCE	
	Mean	Sd	Mean	Sd	Mean	Sd
Design 3a	0.244	0.271	0.194	0.239	0.515	0.291
Design 3b	0.508	0.303	0.513	0.301	0.506	0.291

Table 11: Censored regression model Design 3a − RE Tobit, CRE Tobit, and MPCE, Normally distributed error components and fixed effects, $n = 125$

	β_1	β_2	β_3	β_4	β_5	γ
			RE Tobit			
Median bias	0.670	0.509	0.004	−0.011	−0.111	0.416
Median absolute error	0.670	0.509	0.066	0.068	0.120	0.416
			CRE Tobit			
Median bias	0.228	0.066	0.006	−0.011	−0.081	0.442
Median absolute error	0.228	0.081	0.074	0.073	0.091	0.442
			MPCE			
Median bias	0.023	−0.002	0.010	0.006	−0.006	0.047
Median absolute error	0.110	0.100	0.101	0.102	0.115	0.264

Table 12: Censored regression model Design 3b − RE Tobit, CRE Tobit, and MPCE, Normally distributed error components and random effects, $n = 125$

	β_1	β_2	β_3	β_4	β_5	γ
			RE Tobit			
Median bias	−0.010	−0.001	0.005	0.001	−0.001	0.009
Median absolute error	0.047	0.039	0.041	0.043	0.046	0.144
			CRE Tobit			
Median bias	−0.009	0.001	0.007	−0.000	−0.004	0.003
Median absolute error	0.050	0.038	0.042	0.045	0.046	0.149
			MPCE			
Median bias	0.011	−0.009	0.006	−0.008	−0.012	0.018
Median absolute error	0.065	0.051	0.052	0.050	0.052	0.167

Table 13: Results for RE Tobit, CRE Tobit, MPCE, and Pairwise Comparison

	RE Tobit	CRE Tobit	MPCE small bandwidth	MPCE medium bandwidth	MPCE large bandwidth	Pairwise Comparison
Dependent variable:	log wages					
Age	0.141***	0.133***	0.105***	0.092***	0.066***	0.086***
	(0.009)	(0.012)	(0.014)	(0.012)	(0.010)	(0.012)
Education	0.110***	0.110***	0.073***	0.072***	0.073***	0.021
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.021)
Black	−0.715***	−0.716***	−0.556***	−0.548***	−0.537***	−0.770***
	(0.047)	(0.045)	(0.057)	(0.052)	(0.049)	(0.217)
Age2	−0.002***	−0.002***	−0.001***	−0.001***	−0.001***	−0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Black × Year 1964	−0.040	−0.040	−0.017	−0.017	−0.017	−0.017
	(0.027)	(0.030)	(0.033)	(0.033)	(0.033)	(0.033)
Black × Year 1970	0.151***	0.151***	0.147***	0.147***	0.147***	0.147***
	(0.039)	(0.038)	(0.044)	(0.044)	(0.044)	(0.044)
Black × Year 1971	0.155***	0.154***	0.140**	0.140**	0.140**	0.140**
	(0.039)	(0.041)	(0.046)	(0.046)	(0.046)	(0.046)
Number of observations	33,720	33,720	33,720	33,720	33,720	33,720

Standard errors are in parentheses. Bootstrap standard errors for RE and CRE tobit and Pairwise Comparison are based on 200 replications. Analytical standard errors are reported for MPCE. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All regressions include year dummies.

Table 14: Panel data models

	RE Poisson	CRE Poisson	TSFE	RE Poisson	CRE Poisson	TSFE
Dependent variable: Trade/10000						
Distance	-0.253^{**}	-0.394^{***}	-0.255	-0.251^{**}	-0.394^{***}	-0.254
	(0.080)	(0.085)	(0.139)	(0.082)	(0.082)	(0.139)
Border	2.938^{***}	2.599^{***}	2.391^{***}	2.945^{***}	2.599^{***}	2.396^{***}
	(0.305)	(0.283)	(0.412)	(0.302)	(0.273)	(0.412)
Island	-0.484	-0.621^{*}	-0.274	-0.482	-0.621^{*}	-0.271
	(0.328)	(0.301)	(0.303)	(0.294)	(0.309)	(0.303)
Landlocked	-2.768^{***}	-2.477^{***}	-2.000^{***}	-2.765^{***}	-2.477^{***}	-1.998^{***}
	(0.266)	(0.272)	(0.323)	(0.268)	(0.268)	(0.323)
Colony	2.546^{***}	2.115^{***}	1.854^{***}	2.542^{***}	2.114^{***}	1.849^{***}
	(0.269)	(0.224)	(0.314)	(0.278)	(0.238)	(0.316)
Same legal system	-0.466^{***}	-0.202	-0.302^{*}	-0.466^{***}	-0.202	-0.303^{*}
	(0.128)	(0.140)	(0.120)	(0.127)	(0.151)	(0.120)
Same language	-0.414^{**}	-0.493^{***}	-0.391^{*}	-0.420^{**}	-0.492^{***}	-0.396^{*}
	(0.133)	(0.141)	(0.181)	(0.135)	(0.142)	(0.182)
Same religion	-0.620^{***}	-1.076^{***}	-0.234	-0.618^{***}	-1.076^{***}	-0.235
	(0.162)	(0.176)	(0.129)	(0.174)	(0.189)	(0.129)
CU	-0.226	-0.223	-0.222			
	(0.169)	(0.229)	(0.330)			
FTA	0.199	0.194	0.194	0.199	0.194	0.194
	(0.136)	(0.125)	(0.132)	(0.123)	(0.123)	(0.132)
WTO none	-0.565^{***}	-0.515^{**}	-0.513^{**}	-0.565^{***}	-0.515^{*}	-0.513^{**}
	(0.159)	(0.195)	(0.169)	(0.170)	(0.213)	(0.169)
WTO both	0.176^{**}	0.149^{**}	0.149^{**}	0.176^{**}	0.149^{**}	0.148^{**}
	(0.058)	(0.053)	(0.051)	(0.055)	(0.054)	(0.051)

Bootstrap standard errors based on 200 replications in parentheses.
* $p < 0.05$, ** $p < 0.01$, $^{***}p < 0.001$. All regressions include year dummies.

Table 15: Three-step estimation

	RE Poisson	CRE Poisson	3SE
Dependent variable:	Trade/10000		
Distance (in logs)	-0.364^{***}	-0.490^{***}	-0.700^{***}
	(0.066)	(0.071)	(0.048)
Border	3.142^{***}	3.180^{***}	0.674^{**}
	(0.276)	(0.452)	(0.218)
Island	-0.582	-0.442	-0.250
	(0.438)	(0.371)	(0.211)
Landlocked	0.714^{*}	0.738^{*}	0.510^{*}
	(0.309)	(0.309)	(0.222)
Colony	2.588^{***}	2.240^{***}	0.472^{***}
	(0.264)	(0.210)	(0.106)
Same legal system	-0.497^{***}	-0.194	0.347^{***}
	(0.121)	(0.127)	(0.082)
Same language	-0.591^{***}	-0.720^{***}	-0.121
	(0.120)	(0.125)	(0.106)
Same religion	-0.636^{***}	-1.098^{***}	-0.476^{*}
	(0.147)	(0.183)	(0.184)
FTA	0.198	0.194	0.194^{**}
	(0.135)	(0.124)	(0.056)
WTO none	-0.562^{***}	-0.513^{**}	-0.513^{***}
	(0.160)	(0.193)	(0.125)
WTO both	0.177^{**}	0.149^{**}	0.148^{*}
	(0.058)	(0.052)	(0.053)
Island i	0.008	-0.190	0.217
Landlocked i	-2.000	-1.929	-1.901
Island j	-0.288	-0.394	-0.194
Landlocked j	-2.052	-1.888	-1.995

Bootstrap standard errors based on 200 replications in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All regressions include year dummies.

Table 16: Three-step multiplicative model Design 4 − RE Poisson, CRE Poisson, and 3SE, $n = 158$, $T = 10$

	β_1	β_2	γ_1	γ_2	δ_1	δ_2	δ_3	δ_4
				RE Poisson				
Median bias	−0.000	−0.000	0.663	0.002	−0.000	−0.001	0.003	0.001
Median absolute error	0.011	0.010	0.663	0.010	0.019	0.019	0.019	0.020
				CRE Poisson				
Median bias	−0.000	−0.001	0.664	0.001	0.000	−0.000	0.003	0.001
Median absolute error	0.011	0.010	0.664	0.009	0.019	0.018	0.019	0.020
				3SE				
Median bias	−0.000	−0.001	−0.000	0.002	0.001	−0.001	0.009	0.002
Median absolute error	0.011	0.010	0.013	0.013	0.057	0.054	0.057	0.056