ABSTRACT. This paper provides a method to construct simultaneous confidence bands for quantile functions and quantile effects in nonlinear network and panel models with unobserved two-way effects, strictly exogenous covariates, and possibly discrete outcome variables. The method is based upon projection of simultaneous confidence bands for distribution functions constructed from fixed effects distribution regression estimators. These fixed effects estimators are bias corrected to deal with the incidental parameter problem. Under asymptotic sequences where both dimensions of the data set grow at the same rate, the confidence bands for the quantile functions and effects have correct joint coverage in large samples. An empirical application to gravity models of trade illustrates the applicability of the methods to network data.

(preliminary and incomplete version, please do not circulate)

1. INTRODUCTION

Standard regression analyzes average effects of covariates on outcome variables. In many applications it is equally important to consider distributional effects. For example, a policy maker might be interested in the effect of an education reform not only on the mean but also the entire distribution of test scores or wages. Availability of panel data is very useful to identify \textit{ceteris paribus} average and distributional effects because allows the researcher to control for multiple sources of unobserved heterogeneity that might cause endogeneity or omitted variable problems. The idea is to use variation of the covariates over time for each individual or over individuals for each time period to account for unobserved individual and time effects. In this paper we develop inference methods for distributional effects in nonlinear models with two-way unobserved effects. They apply not only to traditional panel data models where the unobserved effects correspond to individual and time fixed effects, but also to models for other types of data where the unobserved effects reflect some grouping structure such as unobserved sender and receiver effects in network data models.
We develop inference methods for quantile functions and effects. The quantile function corresponds to the marginal distribution of the outcome in a counterfactual scenario where the treatment covariate of interest is set exogenously at a desired level and the rest of the covariates and unobserved effects are held fixed. The quantile effect is the difference of quantile functions at two different treatment levels. They apply to continuous and discrete treatments by appropriate choice of the treatment levels, and have causal interpretation under standard unconfoundedness assumptions; see for example Chernozhukov Fernandez-Val and Melly (2013b). The inference is based upon the generic method of Chernozhukov, Fernandez-Val, Melly, and Wuthrich (2016) that projects joint confidence bands for counterfactual distributions into joint confidence bands for quantile functions and effects. This method has the appealing feature that applies without modification to any type of outcome, let it be continuous, discrete or mixed.

The key input for the inference method is a joint confidence band for the counterfactual distributions at the treatment levels of interest. We construct this band from fixed effects distribution regression (FE-DR) estimators of the conditional distribution of the outcome given the observed covariates and unobserved effects. In doing so, we extend the distribution regression approach to model conditional distributions with unobserved effects. This version of the DR model is semiparametric because not only the DR coefficients can vary with the level of the outcome as in the cross section case, but also the distribution of the unobserved effects is left unrestricted. We show that the FE-DR estimator can be obtained as a sequence of binary response fixed effects estimators where the binary response is an indicator of the outcome passing some threshold. To deal with the incidental parameter problem associated with the estimation of the unobserved effects (Neyman and Scott (1948)), we extend the analytical bias corrections of Fernandez-Val and Weidner (2016) for single binary response estimators to multiple (possibly a continuum) of binary response estimators. In particular, we establish functional central limit theorems for the fixed effects estimators of the DR coefficients and associated counterfactual distributions, and show the validity of the bias corrections under asymptotic sequences where the two dimensions of the data set pass to infinity at the same rate. As in the single binary response model, the bias corrections remove the asymptotic bias of the fixed effects estimators without increasing their asymptotic variances.

We implement the inference method using multiplier bootstrap. This version of bootstrap constructs draws of an estimator as weighted averages of its influence function, where the weights are independent from the data. Compared to empirical bootstrap, multiplier bootstrap has the computational advantage that it does not involve any parameter reestimation. This advantage is particularly convenient in our setting because the parameter estimation require multiple nonlinear optimizations that can be highly dimensional due to the fixed effects. Multiplier bootstrap is also convenient to account for data dependencies. In network data, for example, it might be important to account for reciprocity or pairwise clustering. Reciprocity arises because observational units corresponding to the same pair of agents but reversing their roles as sender and receiver might be dependent even
after conditioning on the unobserved effects. By setting the weights of these observational units equal, we account for this dependence in the multiplier bootstrap. In addition to the previous practical reasons, there are some theoretical reasons for choosing multiplier bootstrap. Thus, Belloni et al. (2017) established bootstrap functional central limit theorems for multiplier bootstrap in high dimensional settings that can be extended to the network and panel models that we consider.

The methods developed in this paper apply to models that include unobserved effects to capture grouping or clustering structures in the data such as models for panel and network data. These effects allow us to control for unobserved group heterogeneity that might be related to the covariates causing endogeneity or omitted variable bias. They also serve to parsimoniously account for clustering dependences in the data. We illustrate the wide applicability with an empirical example to gravity models of trade. In this case the outcome is the volume of trade between two countries and each observational unit corresponds to a country pair indexed by exporter country (sender) and importer country (receiver). We estimate the distributional effects of gravity variables such as the geographical distance controlling for exporter and importer country effects that pick up unobserved heterogeneity that might be correlated with the gravity variables. We uncover significant heterogeneity in the effects of distance and other gravity variables across the distribution, which is missed by traditional mean methods. We also find that the Poisson model, which is commonly used in the trade literature to deal with zero trade in many country pairs, does not provide a good approximation to the distribution of the volume of trade due to severe overdispersion.

**Literature review.** Unlike mean effects, there are different ways to define distributional and quantile effects. For example, we can distinguish conditional effects versus unconditional or marginalized effects, or quantile effects versus quantiles of the effects. Here we give a brief review of the recent literature on distributional and quantile effects in panel data models emphasizing the following aspects: (1) type of effect considered; (2) type of unobserved effects in the model; and (3) asymptotic approximation. For the unobserved effects, we distinguish models with one-way effects versus two-way effects. For the asymptotic approximation we distinguish short panels with large \( N \) and fixed \( T \) versus long panels with large \( N \) and large \( T \), where \( N \) and \( T \) denote the dimensions of the panel. We focus mainly on fixed effects approaches where the unobserved effects are treated as parameters to be estimated, but also mention some correlated random effects approaches that impose restrictions on the distribution of the unobserved effects. This paper deals with inference on marginalized quantile effects in large panels with two-way effects.

estimators without shrinkage and developed bias corrections. All these papers require that
$T$ pass to infinity faster than $N$, making difficult to extend the theory to models with two-
way individual and time effects. Graham, Hahn and Powell (2009) found a special case
where the fixed effects quantile regression estimator does not suffer of incidental parameter
problem.

In short panels, Rosen (2012) showed that a linear quantile restriction is not sufficient
to point identify conditional effects in a panel linear quantile regression model with un-
observed individual effects. Chernozhukov, Fernandez-Val, Hahn and Newey (2013a) and
Chernozhukov et al. (2015) discussed identification and estimation of marginalized quantile
effects in nonseparable panel models with unobserved individual effects and location and
scale time effects under a time homogeneity assumption. They showed that the effects
are point identified only for some subpopulations and characterized these subpopulations.
Graham, Hahn, Poirier and Powell (2015) considered quantiles of effects in linear quantile
regression models with two-way effects. Finally, Abrevaya and Dahl (2008) and Arellano
and Bonhomme (2016) developed estimators for conditional quantile effects in linear quant-
ile regression model with unobserved individual effects using a correlated random effects
approach.

Plan of the paper. Section 2 introduces the distribution regression model with unob-
served effects for network and panel data, and describes the quantities of interest including
model parameters, distributions, quantiles and quantile effects. Section 3 discusses fixed
effects estimation, bias corrections to deal with the incidental parameter problem, and
uniform inference methods. Section 4 provides asymptotic theory for the fixed effects esti-
mators, bias corrections, and multiplier bootstrap. Section 5 and 6 report results of
the empirical application to the gravity models of trade and a Monte Carlo simulation
calibrated to the application, respectively. All the proofs are given in the Appendix.

Notation. For any two real numbers $a$ and $b$, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For
a real number $a$, $\lfloor a \rfloor$ denotes the integer part of $a$. For a set $A$, $|A|$ denotes the cardinality
or number of elements of $A$.

2. Model and Parameters of Interest

2.1. Distribution Regression Model with Unobserved Effects. We observe the data
set $\{(y_{ij}, x_{ij}) : (i, j) \in D\}$, where $y_{ij}$ is a scalar outcome variable with region of interest
$\mathcal{Y}$, and $x_{ij}$ is a vector of covariates with support $\mathcal{X} \subseteq \mathbb{R}^{d_x}$.\footnote{If $y_{ij}$ has unbounded support, then the region $\mathcal{Y}$ is usually a subset of the support to avoid tail estimation.} The variable $y_{ij}$ can be
discrete, mixed or continuous. The subscripts $i$ and $j$ index individuals and time periods in
traditional panels, but they might index other dimensions in more general data structures.
In our empirical application, for example, we use a panel where $y_{ij}$ is the volume of trade
between country \(i\) and country \(j\), and \(x_{ij}\) includes gravity variables such as the distance between country \(i\) and country \(j\). Both \(i\) and \(j\) index countries as exporters and importers respectively. The set \(D\) contains the indexes of the units that are observed. It is a subset of the set of all possible pairs \(D_0 = \{(i, j) : i = 1, \ldots, I; j = 1, \ldots, J\}\), where \(I\) and \(J\) are the dimensions of the panel. We introduce \(D\) to allows for missing data that are common in panel and network applications. For example, in the trade application \(I = J\) and \(D = D_0 \setminus \{(i, i) : i = 1, \ldots, I\}\) because we do not observe trade of a country with itself. We denote the total number of observed units by \(n\), i.e. \(n = |D|\).

Let \(v_i\) and \(w_j\) denote vectors of unspecified dimension that contain unobserved random variables or effects that might be related to the covariates \(x_{ij}\). In traditional panels, \(v_i\) are individual effects that capture unobserved individual heterogeneity and \(w_j\) are time effects that account for aggregate shocks. More generally, these variables serve to capture some forms of endogeneity and group dependencies in a parsimonious fashion. We specify the conditional distribution of \(y_{ij}\) given \((x_{ij}, v_i, w_j)\) using the distribution regression (DR) model with unobserved effects

\[
F_{y_{ij}}(y \mid x_{ij}, v_i, w_j) = F_y(P(x_{ij})'\beta(y) + \alpha(v_i, y) + \gamma(w_j, y)), \quad y \in \mathcal{Y}, \quad (i, j) \in D, \tag{2.1}
\]

where \(F_y\) is a known link function such as the normal or logistic distribution, which may vary with \(y\), \(x \mapsto P(x)\) is a dictionary of transformations of \(x\) such as polynomials, b-splines and tensor products with fixed dimension that has a one as the first component, \(\beta(y)\) is an unknown parameter vector, which can vary with \(y\), and \((v, y) \mapsto \alpha(v, y)\) and \((w, y) \mapsto \gamma(w, y)\) are unspecified measurable functions. This DR model is a semiparametric model for the conditional distribution because \(y \mapsto \theta(y) := (\beta(y), \alpha_1(y), \ldots, \alpha_f(y), \gamma_1(y), \ldots, \gamma_f(y))\) is a function-valued parameter and the dimension of \(\theta(y)\) varies with \(I\) and \(J\), although we do not make this dependence explicit.

When \(y_{ij}\) is continuous, the model (2.1) has the following representation as an implicit nonseparable model by the probability integral transform

\[
F_{y_{ij}}(P(x_{ij})'\beta(y_{ij}) + \alpha(v_i, y_{ij}) + \gamma(w_j, y_{ij})) = u_{ij}, \quad u_{ij} \mid x_{ij}, v_i, w_j \sim U(0, 1),
\]

where the error \(u_{ij}\) represents the unobserved ranking of the observation \(y_{ij}\) in the conditional distribution. We shall use this representation as the data generating process in Monte Carlo simulations. The parameters of the model are related to derivatives of the conditional quantiles. Let \(Q_{y_{ij}}(u \mid x_{ij}, v_i, w_j)\) be the \(u\)-quantile of \(y_{ij}\) conditional on \((x_{ij}, v_i, w_j)\) defined as the left-inverse of \(y \mapsto F_{y_{ij}}(y \mid x_{ij}, v_i, w_j)\) at \(u\), namely

\[
Q_{y_{ij}}(u \mid x_{ij}, v_i, w_j) = \inf\{y \in \mathcal{Y} : F_{y_{ij}}(y \mid x_{ij}, v_i, w_j) \geq u\},
\]

and \(x_{ij} = (x_{ij1}, \ldots, x_{ijd_x})\). Then, it can be shown that if \(x_{ij} \mapsto Q_{y_{ij}}(u \mid x_{ij}, v_i, w_j)\) is differentiable,

\[
\beta_k(Q_{y_{ij}}(u \mid x_{ij}, v_i, w_j)) \propto -\partial_{x_{ij}} Q_{y_{ij}}(u \mid x_{ij}, v_i, w_j), \quad k = 1, \ldots, d_x, \quad \partial_{x_{ij}} := \partial/\partial x_{ij}.
\]
and
\[
\frac{\beta_k(y)}{\beta_l(y)} \bigg|_{y = Q_{yij}(u \mid x_{ij}, v_i, w_j)} = \frac{\partial_{x_{ij}} Q_{yij}^\ell(u \mid x_{ij}, v_i, w_j)}{\partial_{x_{ij}} Q_{yij}^\ell(u \mid x_{ij}, v_i, w_j)}, \quad \ell, k = 1, \ldots, d_x,
\]
provided that \( \partial_{x_{ij}} Q_{yij}^\ell(u \mid x_{ij}, v_i, w_j) \neq 0 \). The DR coefficients therefore are proportional to (minus) derivatives of the conditional quantile function, and ratios of DR coefficients correspond to ratios of derivatives.

**Remark 1** (Parametric models). There are many parametric models that are special cases of the DR model. For example, the standard linear model with two-way effects
\[
y_{ij} = x_{ij}' \beta + \alpha(v_i) + \gamma(w_j) + \sigma \varepsilon_{ij}, \quad \varepsilon_{ij} \mid x_{ij}, v_i, w_j \sim G,
\]
is a special case of the model (2.1) with link function \( F_y = G \) for all \( y \),
\[
\beta(y) = (e_1 y - \beta)/\sigma, \quad \alpha(v_i, y) = -\alpha(v_i)/\sigma, \quad \gamma(w_j, y) = -\gamma(w_j)/\sigma,
\]
where \( e_1 \) is a unit vector of dimension \( d_x \) with a one in the first component. This location-shift model imposes that all the parameters, other than the intercept, are fixed with respect to the level \( y \). This greatly restricts the heterogeneity in the conditional quantile effects,
\[
\partial_{x_{ij}} Q_{yij}^\ell(u \mid x_{ij}, v_i, w_j) = \beta_{\ell}, \quad \forall (u, x_{ij}, v_i, w_j),
\]
which are constant not only with respect to the quantile index but also with respect to the observed and unobserved covariates.

The Cox proportional hazard panel model for duration data
\[
t(y_{ij}) = P(x_{ij})^\beta + \alpha(v_i) + \gamma(w_j) + \varepsilon_{ij},
\]
where \( \varepsilon_{ij} \) has an extreme value distribution independent of \((x_{ij}, v_i, w_j)\) and \( t(y) \) is an unknown monotone transformation, is also a special case of the model (2.1) with link function \( F_y(\ell) = 1 - \exp(-\exp(-\ell)) \), the complementary log-log distribution, and parameters
\[
\beta(y) = \beta - e_1 t(y), \quad \alpha(v_i, y) = \alpha(v_i), \quad \gamma(w_j, y) = \gamma(w_j),
\]
where again only the intercept varies with the level \( y \).

Another example is the Poisson model for count data with two-way effects where, for \( \lambda_{ij} = \exp(P(x_{ij})^\beta + \alpha(v_i) + \gamma(w_j)) \),
\[
F_{y_{ij}}(y \mid x_{ij}, v_i, w_j) = \exp(-\lambda_{ij}) \sum_{k=0}^{\lfloor y \rfloor} \lambda_{ij}^k/k!, \quad y \geq 0,
\]
where \( \lfloor y \rfloor \) denotes the integer part of \( y \). This conditional distribution is encompassed by the model (2.1) with link function \( F_y(\ell) = \exp(-\exp(\ell)) \sum_{k=0}^{\lfloor y \rfloor} \exp(\epsilon)^k/k! \), the incomplete gamma function, and parameters \( \beta(y) = \beta, \alpha(v_i, y) = \alpha(v_i)\), and \( \gamma(w_j, y) = \gamma(w_j) \) that do not vary with \( y \). In this case the link function changes with \( y \).
2.2. Estimands. We are interested in measuring the effect on the outcome of changing one of the covariates holding the rest of the covariates and the unobserved effects fixed. Let \( x = (t, z) \), where \( t \) is the covariate of interest or treatment and \( z \) are the rest of the covariates that usually play the role as controls. One effect of interest is the quantile (left-inverse) function \( Q_k(\tau) = F_k^-(\tau) := \inf\{y \in \mathcal{Y} : F_k(y) \geq \tau\}, \quad \tau \in (0, 1) \),

where

\[
F_k(y) = n^{-1} \sum_{(i,j) \in D} F_y(P(t_{ij}^k, z_{ij}^k)' \beta(y) + \alpha(v_i, y) + \gamma(w_j, y)),
\]

\( t_{ij}^k \) is a level of the treatment that may depend on \( t_{ij} \), and \( k \in \{0, 1\} \). We provide examples below. Note that in the construction of the counterfactual distribution \( F_k \), we marginalize \( (x_{ij}, v_i, w_j) \) using the empirical distribution. The resulting effects are finite population effects. We shall focus on these effects because conditioning on the covariates and unobserved effects is natural in the trade application. We construct the quantile effect function (QEF) by taking differences of the QF at two treatment levels

\[
\Delta(\tau) = Q_1(\tau) - Q_0(\tau), \quad \tau \in (0, 1).
\]

The choice of the levels \( t_{ij}^0 \) and \( t_{ij}^1 \) is usually based on the scale of the treatment:

- If the treatment is binary, \( \Delta(\tau) \) is the \( \tau \)-quantile treatment effect with \( t_{ij}^0 = 0 \) and \( t_{ij}^1 = 1 \).
- If the treatment is continuous, \( \Delta(\tau) \) is the \( \tau \)-quantile effect of a unitary or one standard deviation increase in the treatment with \( t_{ij}^0 = t_{ij} \) and \( t_{ij}^1 = t_{ij} + d \), where \( d \) is 1 or the standard deviation of \( t_{ij} \).
- If the treatment is the logarithm of a continuous treatment, \( \Delta(\tau) \) is the \( \tau \)-quantile effect of doubling the treatment (100\% increase) with \( t_{ij}^0 = t_{ij} \) and \( t_{ij}^1 = t_{ij} + \log 2 \).

For example, in the trade application we use the levels \( t_{ij}^0 = 0 \) and \( t_{ij}^1 = 1 \) for binary covariates such as the indicators for common legal system and free trade area, and \( t_{ij}^0 = t_{ij} \) and \( t_{ij}^1 = t_{ij} + \log 2 \) for the logarithm of distance.

3. Fixed Effects Estimation and Uniform Inference

To simplify the notation in this section we write \( P(x_{ij}) = x_{ij} \) without loss of generality, and define \( \alpha_i(y) := \alpha(v_i, y) \) and \( \gamma_j(y) := \gamma(w_j, y) \). We denote by \( Y \) a finite subset of \( \mathcal{Y} \).
where the estimation and inference are performed. If \( \mathcal{Y} \) is finite, then we can set \( \bar{\mathcal{Y}} = \mathcal{Y} \). Otherwise, \( \bar{\mathcal{Y}} \) is a finite grid covering \( \mathcal{Y} \) with a size that might increase with \( n \).

### 3.1. Fixed Effects Distribution Regression Estimator.

The parameters of the PDR model can be estimated from multiple binary regressions with two-way effects. To see this, note that the conditional distribution in (2.1) can be expressed as

\[
F_y(x_{ij}'\beta(y) + \alpha_i(y) + \gamma_j(y)) = \mathbb{E}[1\{y_{ij} \leq y\} | x_{ij}, v_i, w_j].
\]

Accordingly, we can construct a collection of binary variables,

\[
1\{y_{ij} \leq y\}, \quad (i, j) \in \mathcal{D}, \quad y \in \bar{\mathcal{Y}},
\]

and estimate the parameters for each \( y \) by conditional maximum likelihood with fixed effects. Thus,

\[
\hat{\theta}(y) := (\hat{\beta}(y), \hat{\alpha}_1(y), \ldots, \hat{\alpha}_I(y), \hat{\gamma}_1(y), \ldots, \hat{\gamma}_J(y)),
\]

the fixed effects distribution regression estimator of \( \theta(y) := (\beta(y), \alpha_1(y), \ldots, \alpha_I(y), \gamma_1(y), \ldots, \gamma_J(y)) \), is obtained as

\[
\hat{\theta}(y) \in \arg \max_{\theta \in \Theta_{I,J}} \sum_{i,t} (1\{y_{ij} \leq y\} \log F_y((x_{ij}'\beta + \alpha_i + \gamma_j)) + 1\{y_{ij} > y\} \log[1 - F_y(x_{ij}'\beta + \alpha_i + \gamma_j))],
\]

(3.1)

where \( \Theta_{I,J} \subseteq \mathbb{R}^{d_x + I + J} \) is the parameter space for all the \( \theta(y), y \in \bar{\mathcal{Y}} \). When the link function is the normal or logistic distribution, the previous program is concave and smooth in parameters and therefore has good computational properties. See Fernandez-Val and Weidner (2016) and Cruz-Gonzalez et al. (2016) for a discussion of computation of logit and probit regressions with two-way effects.

The quantile functions and effects are estimated via plug-in rule, i.e.,

\[
\hat{Q}_k(\tau) = \hat{F}_k^{-1}(\tau) \wedge \sup\{y \in \bar{\mathcal{Y}}\}, \quad \tau \in (0, 1), \quad k \in \{0, 1\},
\]

where

\[
\hat{F}_k(y) = n^{-1} \sum_{(i,j) \in \mathcal{D}} F_y(t^k_{ij}, z^k_{ij}') \hat{\beta}(y) + \hat{\alpha}_i(y) + \hat{\gamma}_j(y), \quad y \in \bar{\mathcal{Y}},
\]

and

\[
\hat{\Delta}(\tau) = \hat{Q}_1(\tau) - \hat{Q}_0(\tau) \quad \tau \in (0, 1).
\]

### 3.2. Incidental Parameter Problem and Bias Corrections.

Fixed effects estimators can be severely biased in nonlinear models because of the incidental parameter problem (Neyman and Scott, 1948). These models include the binary regressions that we estimate to obtain the DR coefficients and parameters of interest. We deal with the incidental parameter problem using the analytical bias corrections of Fernandez-Val and Weidner (2016) for parameters and average partial effects (APE) in binary regressions with two-way effects. We note here that the distributions \( F_k(y), k \in \mathcal{K}, \) can be seen as APE, i.e. they are averages of functions of the data, unobserved effects and parameters.
The bias corrections are based on expansions of the bias of the fixed effects estimators as $I, J \to \infty$. For example,

$$
\mathbb{E}[\tilde{F}_k(y) - F_k(y)] = \frac{I}{n} B_k^{(F)}(y) + \frac{J}{n} D_k^{(F)}(y) + R_k^{(F)}(y), \quad (3.2)
$$

where $nR_k^{(F)}(y) = o(I \vee J)$.

In Section 4 we establish that this expansion holds uniformly in $y \in \mathcal{Y}$ and $k \in \{0, 1\}$, i.e.

$$
\sup_{k \in \{0, 1\}, y \in \mathcal{Y}} \|nR_k^{(F)}(y)\| = o(I \vee J).
$$

This result generalizes the analysis of Fernandez-Val and Weidner (2016) from a single binary regression to multiple (possibly a continuum) of binary regressions. This generalization is required to implement the inference methods for quantile functions and effects.

The expansion (3.2) is the basis for the bias corrections. Let $\hat{B}_k^{(F)}(y)$ and $\hat{D}_k^{(F)}(y)$ be estimators of $B_k^{(F)}(y)$ and $D_k^{(F)}(y)$, which are uniformly consistent in $y \in \mathcal{Y}$. Bias corrected fixed effects estimators of $F_k$ and $Q_k$ are formed as

$$
\hat{Q}_k(\tau) = \tilde{F}_k^-(\tau) \wedge \sup\{y \in \bar{\mathcal{Y}}\},
$$

$$
\tilde{F}_k(y) = \hat{F}_k(y) - \frac{I}{n} \hat{B}_k(y) - \frac{J}{n} \hat{D}_k(y), \quad y \in \bar{\mathcal{Y}}.
$$

We also use the corrected estimators $\tilde{F}_k$ as the basis for inference.

**Remark 2** (Monotonization). If the bias corrected estimator $y \mapsto \tilde{F}_k(y)$ is non-monotone on $\bar{\mathcal{Y}}$, we can rearrange it into a monotone function by simply sorting the values of function in a nondecreasing order. Chernozhukov et al. (2009) showed that the rearrangement improves the finite sample properties of the estimator.

### 3.3. Uniform Inference.

Our inference goal is to construct confidence bands that cover the QF $\tau \mapsto Q_k(\tau)$ and the QEF $\tau \mapsto \Delta(\tau)$ simultaneously over a set of quantiles $\mathcal{T} \subseteq [\varepsilon, 1 - \varepsilon]$, for some $0 < \varepsilon < 1/2$. The set $\mathcal{T}$ is chosen such that $Q_k(\tau) \in [\inf\{y \in \mathcal{Y}\}, \sup\{y \in \mathcal{Y}\}]$, for all $\tau \in \mathcal{T}$ and $k \in \{0, 1\}$.

We use the generic method of Chernozhukov et al. (2016) to construct confidence bands for quantile functions and effects from confidence bands for the corresponding distributions. Let $\mathbb{D}$ denote the space of weakly increasing functions, mapping $\bar{\mathcal{Y}}$ to $[0, 1]$. Assume we have a confidence band $I_k = [L_k, U_k]$ for $F_k$, with lower and upper endpoint functions $L_k$ and $U_k$. Specifically, given two functions $y \mapsto U_k(y)$ and $y \mapsto L_k(y)$ in the set $\mathbb{D}$ such that $L_k \leq U_k$, pointwise, we define a band $I_k = [L_k, U_k]$ as the collection of intervals

$$
I_k(y) = [L_k(y), U_k(y)], \quad y \in \bar{\mathcal{Y}}.
$$

---

3Fernandez-Val and Weidner (2016) considered the case where $n = IJ$, i.e. there is no missing data, so that $I/n = 1/J$ and $J/n = 1/I$. 
We say that $I_k$ covers $F_k$ if $F_k \in I_k$ pointwise namely $F_k(y) \in I_k(y)$ for all $y \in \bar{Y}$. If $U_k$ and $L_k$ are some data-dependent bands, we say that $I_k$ is a confidence band for $F_k$ of level $p$, if $I_k$ covers $F_k$ with probability at least $p$. Similarly, we say that the set of bands \{ $I_k : k \in \mathcal{K}$ \} is a joint confidence band for the set of functions \{ $F_k : k \in \mathcal{K}$ \} of level $p$, if $I_k$ covers $F_k$ with probability at least $p$ simultaneously over $k \in \mathcal{K}$. The index set $\mathcal{K}$ can be a singleton to cover individual confidence bands or a finite set such as $\mathcal{K} = \{0, 1\}$ to cover joint confidence bands. In Section 4 we provide a multiplier bootstrap algorithm for computing joint confidence bands based on the joint asymptotic distribution of the bias corrected estimators \{ $\tilde{F}_k : k \in \mathcal{K}$ \}.

Note that if $[L_k', U_k']$ is a confidence band for $F_k$ that does not obey the constraint $L_k', U_k' \in \mathbb{D}$, we can transform $[U_k', L_k']$ into a new band $[L_k, U_k]$ such that $L_k, U_k \in \mathbb{D}$.

The following result provides a method to construct joint confidence bands for \{ $Q_k = F_k^- : k \in \mathcal{K}$ \}, from joint confidence bands for \{ $F_k : k \in \mathcal{K}$ \}.

**Lemma 1** (Chernozhukov, Fernandez-Val, Melly, and Wuthrich, 2016). Consider a set of distribution functions \{ $F_k : k \in \mathcal{K}$ \} and endpoint functions \{ $L_k : k \in \mathcal{K}$ \} and \{ $U_k : k \in \mathcal{K}$ \} with components in the class $\mathbb{D}$. Then, if \{ $F_k : k \in \mathcal{K}$ \} is jointly covered by \{ $I_k : k \in \mathcal{K}$ \} with probability at least $p$, then \{ $Q_k = F_k^- : k \in \mathcal{K}$ \} is jointly covered by \{ $I_k^- : k \in \mathcal{K}$ \} with probability at least $p$, where

$$I_k^- (\tau) := [U_k^- (\tau), L_k^- (\tau)], \quad \tau \in \mathcal{T}, \quad k \in \mathcal{K}.$$

This Lemma establishes that we can construct confidence bands for quantile functions by inverting the endpoint functions of confidence bands for distribution functions. The geometric intuition is that the inversion amounts to rotate and flip the bands, and these operations preserve coverage.

We next construct simultaneous confidence sets for the quantile effect function $\tau \mapsto \Delta(\tau)$ defined by

$$\Delta(\tau) = Q_1(\tau) - Q_0(\tau) = F_1^-(\tau) - F_0^-(\tau), \quad \tau \in \mathcal{T}.$$  

\footnote{Other monotonization operators, such as the projection on the set of weakly increasing functions, can also be used.}
The basic idea is to take the Minkowski differences of the bands for the quantile functions $Q_1$ and $Q_0$ as the confidence band for the quantile effect. Specifically, suppose we have the set of confidence bands $\{I_k^\pm : k = 0, 1\}$ for the set of functions $\{F_k^\pm : k = 0, 1\}$ of level $p$. For example, we can construct these bands using Lemma 1. Then we can convert these bands to confidence bands for $\Delta$ by taking the pointwise Minkowski difference $\ominus$ of each of the pairs of the two sets.

**Lemma 2** (Chernozhukov, Fernandez-Val, Melly, and Wuthrich, 2016). Consider the set of distribution functions $\{F_k : k = 0, 1\}$ and the band functions $\{L_k : k = 0, 1\}$ and $\{U_k : k = 0, 1\}$, with components in the class $\mathcal{D}$. If the set of distribution functions $\{F_k : k = 0, 1\}$ is jointly covered by the set of bands $\{I_k : k = 0, 1\}$ with probability at least $p$, then the quantile effect function $\Delta = F_1^\pm - F_0^\pm$ is covered by $I_\Delta^\pm$ with probability at least $p$, where $I_\Delta^\pm$ is defined by:

$$I_\Delta^\pm(\tau) := [U_1^\pm(\tau), L_1^\pm(\tau)] \ominus [U_0^\pm(\tau), L_0^\pm(\tau)] = [U_1^\pm(\tau) - L_0^\pm(\tau), L_1^\pm(\tau) - U_0^\pm(\tau)], \quad \tau \in \mathcal{T}.$$  

4. Asymptotic Theory

This section derives the asymptotic properties of the fixed effect estimators of $y \mapsto \beta(y)$ and $\{F_k : k \in \mathcal{K}\}$, as both dimensions $I$ and $J$ grow to infinity. We focus on the case where the link function is the logistic distribution at all levels, $F_y = \Lambda$, where $\Lambda(\xi) = (1 + \exp(-\xi))^{-1}$. We choose the logistic distribution for analytical convenience. In this case the Hessian of the log-likelihood function does not depend on $y$, what leads to several simplifications in the asymptotic expansions. For the case of single binary regressions, Fernandez-Val and Weidner (2016) showed that the properties of fixed effects estimators are similar for the logistic distribution and other smooth log-concave distributions such as the normal distribution. Accordingly, we expect that our results can be extended to other link functions using conceptually similar arguments but more tedious derivations.

We make the following assumptions:

**Assumption 1** (Sampling and Model Conditions).

(i) **Sampling:** The outcome variable $y_{ij}$ is independently distributed over $i$ and $j$ conditional on all the observed and unobserved covariates $C_B := \{(x_{ij}, v_i, w_j) : (i, j) \in D\}$.

(ii) **Model:** For all $y \in \mathcal{Y}$,

$$F_{y_{ij}}(y \mid C_B) = F_{y_{ij}}(y \mid x_{ij}, v_i, w_j) = \Lambda(x'_{ij} \beta(y) + \alpha(v_i, y) + \gamma(w_j, y)),$$

where $y \mapsto \beta(y)$, $y \mapsto \alpha(\cdot, y)$ and $y \mapsto \gamma(\cdot, y)$ are measurable functions.

(iii) **Compactness:** $\mathcal{X} \mathcal{Y} \mathcal{W}$, the support of $(x_{ij}, v_i, w_j)$, is a compact set, and $y \mapsto \alpha(v_i, y)$ and $y \mapsto \gamma(w_j, y)$ are a.s. uniformly bounded on $\mathcal{Y}$. 


(iv) Compactness and smoothness: Either $\mathcal{Y}$ is a discrete finite set, or $\mathcal{Y} \subset \mathbb{R}$ is a bounded interval. In the latter case, we assume that the conditional density function $f_{y_{ij}}(y \mid x_{ij}, v_i, w_j)$ exists, is uniformly bounded, and is uniformly continuous in $y$, uniformly on $\mathcal{Y} \times \mathcal{X} \mathcal{W}$.

(v) Missing data: There is only a fixed number of missing observations for every $i$ and $j$, that is, $\max_i (J - |\{(i', j') \in \mathcal{D} : i' = i\}|) \leq c_2$ and $\max_j (I - |\{(i', j') \in \mathcal{D} : j' = j\}|) \leq c_2$ for some constant $c_2 < \infty$ that is independent of the sample size.

(vi) Non-collinearity: The regressors $x_{ij}$ are non-collinear after projecting out the two-way fixed effects, that is, there exists a constant $c_3 > 0$, independent of the sample size, such that

$$\min_{\{\delta \in \mathbb{R}^{\dim \beta} : ||\delta||=1\}} \min_{(a,b)\in \mathbb{R}^{I+J}} \left[ \frac{1}{n} \sum_{(i,j)\in \mathcal{D}} (x'_{ij} \delta - a_i - b_j)^2 \right] \geq c_3.$$ 

(vii) Asymptotics: We consider asymptotic sequences where $I_n, J_n \to \infty$ with $I_n/J_n \to c$ for some positive and finite $c$, as the total sample size $n \to \infty$. We drop the indexing by $n$ from $I_n$ and $J_n$, i.e. we shall write $I$ and $J$.

**Remark 3** (Sufficient conditions for Assumption 1). Part (i) holds if $(y_{ij}, x_{ij}, v_i, w_j)$ is i.i.d. over $i$ and $j$, but it is more general as it does not restrict the distribution of $(x_{ij}, v_i, w_j)$ nor its dependence across $i$ and $j$. We show how to relax this assumption allowing for a form of weak conditional dependence in Section 4.4. Part (ii) holds if the observed covariates are strictly exogenous conditional on the unobserved effects and the conditional distribution is correctly specified for all $y \in \mathcal{Y}$. Part (iii) imposes that support of the covariates and the unobserved effects is a compact set. This condition is imposed for analytical convenience. Part (iv) can be slightly weakened to Lipschitz continuity with uniformly bounded Lipschitz constant, instead of differentiability. If the panel is balanced, part (vi) can be stated as

$$\frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \tilde{x}_{ij} \tilde{x}'_{ij} \geq c_3 \|\dim \beta\|,$$

where $\tilde{x}_{ij} = x_{ij} - x_i - x_j + x_., x_i = J^{-1} \sum_{j=1}^J x_{ij}, x_j = I^{-1} \sum_{i=1}^I x_{ij},$ and $x._. = (IJ)^{-1} \sum_{i=1}^I \sum_{j=1}^J x_{ij}$.

### 4.1. Asymptotic Distribution of the Uncorrected Estimator.

We require some further notation. Denote the $q$’the derivatives of the cdf $\Lambda(\cdot)$ by $\Lambda^{(q)}(\cdot)$, and define $\Lambda_{ij}^{(q)}(y) = \Lambda^{(q)}(x'_{ij} \beta(y) + \alpha_i(y) + \gamma_j(y))$ and $\Lambda_{ij,k}^{(q)}(y) = \Lambda^{(q)}(z'_{ij,k} \beta(y) + \alpha_i(y) + \gamma_j(y))$ with $z_{ij,k} := (d_{ij}^k, z'_{ij})'$ and $q = 1, 2, \ldots$. For $\ell \in \{1, \ldots, d_x\}$ define the following projections...
of the \( \ell \)'th covariate \( x_{ij}^\ell \),

\[
\left( \alpha^\ell_x(y), \gamma^\ell_x(y) \right) \in \arg \min_{(a,c) \in \mathbb{R}^{I+J}} \left[ \sum_{(i,j) \in D} \Lambda_{ij}^{(1)}(y) \left( x_{ij}^\ell - a_i - c_j \right)^2 \right],
\]

and let \( \alpha_{x,i}(y) \) and \( \gamma_{x,i}(y) \) be the \( d_x \)-vectors with components \( \alpha^\ell_{x,i}(y) \) and \( \gamma^\ell_{x,i}(y) \), where \( \alpha^\ell_{x,i}(y) \) is the \( i \)th component of \( \alpha^\ell_x(y) \) and \( \gamma^\ell_{x,i}(y) \) is the \( j \)th component of \( \gamma^\ell_x(y) \). Also define \( \bar{x}_{ij}(y) = x_{ij} - \alpha_{x,i}(y) - \gamma_{x,j}(y) \) and \( \bar{x}_{ij,k}(y) = x_{ij,k} - \alpha_{x,i}(y) - \gamma_{x,j}(y) \). Notice that \( \bar{x}_{ij,k}(y) \) is defined using projections of \( x_{ij} \) instead of \( x_{ij,k} \). Also, while the locations of \( \alpha_{x,i}(y) \) and \( \gamma_{x,j}(y) \) are not identified, \( \bar{x}_{ij}(y) \) and \( \bar{x}_{ij,k}(y) \) are uniquely defined. Analogous to the projection of \( x_k \) above, we define \( \Psi_{ij,k}(y) = \alpha^\ell_y(y) + \gamma^\ell_y(y) \), where

\[
\left( \alpha^\ell_y(y), \gamma^\ell_y(y) \right) \in \arg \min_{(a,c) \in \mathbb{R}^{I+J}} \left[ \sum_{(i,j) \in D} \Lambda_{ij}^{(1)}(y) \left( \frac{\Lambda^{(1)}_{ij,k}(y)}{\Lambda^{(1)}_{ij}(y)} - a_i - c_j \right)^2 \right].
\]

[SHOULD IT BE \(-\frac{\Lambda^{(1)}_{ij,k}(y)}{\Lambda^{(1)}_{ij}(y)}\)? For example, if \( x_{ij,k} = x_{ij} \), then \( \Psi_{ij,k}(y) = 1 \). Furthermore, we define

\[
W(y) = \frac{1}{n} \sum_{(i,j) \in D} \Lambda_{ij}^{(1)}(y) \bar{x}_{ij}(y) \bar{x}_{ij}(y)', \quad \partial_\beta F_k(y) = \frac{1}{n} \sum_{(i,j) \in D} \Lambda_{ij,k}^{(1)}(y) \bar{x}_{ij,k}(y) \bar{x}_{ij,k}(y)',
\]

and

\[
B^{(F)}_k(y) = \frac{1}{2} \sum_{j=1}^J \sum_{i \in D_j} \frac{\Lambda^{(2)}_{ij,k}(y) - \Lambda^{(2)}_{ij}(y) \Psi_{ij,k}(y)}{\Lambda^{(1)}_{ij}(y)},
\]

\[
D^{(F)}_k(y) = \frac{1}{2} \sum_{j=1}^J \sum_{i \in D_j} \frac{\Lambda^{(2)}_{ij,k}(y) - \Lambda^{(2)}_{ij}(y) \Psi_{ij,k}(y)}{\Lambda^{(1)}_{ij}(y)}.
\]

where \( D_i := \{(i', j') \in D : i' = i\} \) and \( D_j := \{(i', j') \in D : j' = j\} \) are the subsets of observational units that contain the index \( i \) and \( j \), respectively. In the previous expressions, \( \partial_\beta F_k(y) \) is a \( 1 \times d_x \) vector for each \( k \in K \) that we stack in the \( |K| \times d_x \) matrix \( \partial_\beta F(y) = [\partial_\beta F_k(y) : k \in K] \). Similarly, \( F_k(y), B^{(F)}_k(y), D^{(F)}_k(y) \) and \( \Psi_{ij,k}(y) \) are scalars for each \( k \in K \), that we stack in the \( |K| \times 1 \) vectors \( F(y) = [F_k(y) : k \in K], B^{(F)}(y) = [B^{(F)}_k(y) : k \in K], D^{(F)}(y) = [D^{(F)}_k(y) : k \in K], \Psi_{ij}(y) = \Psi_{ij,k}(y) : k \in K] \).
Let $\ell^\infty(\mathcal{Y})$ be the space of real-valued bounded functions on $\mathcal{Y}$ equipped with the sup-norm $\| \cdot \|_\mathcal{Y}$, and $\rightsquigarrow$ denote weak convergence (in distribution). We establish a functional central limit theorem for the fixed effects estimators of $y \mapsto \beta(y)$ and $y \mapsto F(y)$ in $\mathcal{Y}$.

**Theorem 1** (FCLT for Fixed Effects DR Estimators). Let Assumption 1 hold. For all $y_1, y_2 \in \mathcal{Y}$ with $y_1 \geq y_2$ we assume the existence of

$$\bar{V}(y_1, y_2) = \operatorname{plim}_{n \to \infty} \frac{1}{n} \sum_{(i,j) \in \mathcal{D}} \Lambda_{ij}(y_1) \left[ 1 - \Lambda_{ij}(y_2) \right] \tilde{x}_{ij}(y_1) \tilde{x}_{ij}(y_2)' ,$$

$$\bar{\Omega}(y_1, y_2) = \operatorname{plim}_{n \to \infty} \frac{1}{n} \sum_{(i,j) \in \mathcal{D}} \Lambda_{ij}(y_1) \left[ 1 - \Lambda_{ij}(y_2) \right] \Xi_{ij}(y_1)\Xi_{ij}(y_2)' ,$$

where $\Xi_{ij}(y) = \Psi_{ij}(y) + \partial_\beta F(y) W^{-1}(y) \tilde{x}_{ij}(y)$. Let $\bar{V}(y_2, y_1) := \bar{V}(y_1, y_2)'$, $\bar{\Omega}(y_2, y_1) := \bar{\Omega}(y_1, y_2)'$, and $\bar{W}(y_1) := \bar{V}(y_1, y_1)$. Then, in the metric space $\ell^\infty(\mathcal{Y})^d$,

$$\sqrt{n} \left[ \hat{\beta}(y) - \beta(y) - \frac{I}{n} D(\beta)(y) - \frac{J}{n} D(\beta)(y) \right] \rightsquigarrow Z^{(\beta)}(y),$$

and, in the metric space $\ell^\infty(\mathcal{Y})^{|\mathcal{X}|}$,

$$\sqrt{n} \left\{ \hat{F}(y) - F(y) - \frac{I}{n} \left[ B(\hat{F})(y) + (\partial_\beta F(y)) B(\beta)(y) \right] - \frac{J}{n} \left[ D(\hat{F})(y) + (\partial_\beta F(y)) D(\beta)(y) \right] \right\} \rightsquigarrow Z^{(\hat{F})}(y),$$

as stochastic processes indexed by $y \in \mathcal{Y}$, where $y \mapsto Z^{(\beta)}(y)$ and $y \mapsto Z^{(\hat{F})}(y)$ are tight zero-mean Gaussian processes with covariance functions $(y_1, y_2) \mapsto \bar{W}^{-1}(y_1) \bar{V}(y_1, y_2) \bar{W}^{-1}(y_2)$ and $(y_1, y_2) \mapsto \bar{\Omega}(y_1, y_2)$, respectively.

**Remark 4.** Assumption 1(vi) guarantees the invertibility of $\bar{W}(y)$ and $\bar{W}(y)$. Notice that $\bar{W}(y)$ is equal to the limit of $W(y)$ because $\Lambda_{ij}^{(1)}(y) = \Lambda_{ij}(y) (1 - \Lambda_{ij}(y))$ by the properties of the logistic distribution. This information equality follows by the correct specification condition in Assumption 1(ii). By Assumption 1(v), we could have used $\sqrt{TJ}$ instead of $\sqrt{n}$, $1/J$ instead of $I/n$, and $1/I$ instead of $J/n$. We prefer the expressions in the theorem, because they might provide a more accurate finite-sample approximation.

**Remark 5** (Comparison with binary response models). Fernandez-Val and Weidner (2016) derived the properties of fixed effects estimators of coefficients and APEs in binary regressions with two-way effects. We generalized their results to multiple binary regressions where our quantities of interest are function-valued, for example $y \mapsto \beta(y), y \in \mathcal{Y}$, and therefore the corresponding estimators are stochastic processes.

**Remark 6** (Case $\pi_{ij,k} = x_{ij}$). If $\pi_{ij,k} = x_{ij}$, the asymptotic bias of $\hat{F}_k$ is zero because $\hat{F}_k$ is equal to the empirical distribution function, namely

$$\hat{F}_k(y) = \frac{1}{n} \sum_{(i,j) \in \mathcal{D}} \Lambda(x_{ij}' \beta_{ij} + \tilde{\alpha}_i(y) + \tilde{\gamma}_j(y)) = \frac{1}{n} \sum_{(i,j) \in \mathcal{D}} 1\{y_{ij} \leq y\} ,$$

where $(x_{ij}, \beta_{ij}) = (x_{ij}, \beta_t)$. This implies

$$\sqrt{n} \left\{ \hat{F}(y) - F(y) - \frac{I}{n} \left[ B(\hat{F})(y) + (\partial_\beta F(y)) B(\beta)(y) \right] - \frac{J}{n} \left[ D(\hat{F})(y) + (\partial_\beta F(y)) D(\beta)(y) \right] \right\} \rightsquigarrow Z^{(\hat{F})}(y) .$$
by the first order conditions of the fixed effects logit DR estimator when \( x_{ij} \) includes a constant term. This property provides another appealing feature to choose the logistic distribution.

4.2. Bias Corrections. Theorem 1 shows that the fixed effects DR estimators have asymptotic biases of the same order as the asymptotic standard deviations under the asymptotic approximation that we consider. The finite-sample implications are that these estimators can have substantial biases and that confidence intervals constructed around them can have severe undercoverage. We deal with these problems by reducing the bias of the estimators.

We estimate the bias components using the plug-in rule. Define \( \hat{\Lambda}_{ij}^{(q)}(y) = \Lambda^{(q)}(x'_{ij} \hat{\beta}(y) + \hat{\alpha}_i(y) + \hat{\gamma}_j(y)) \) and \( \hat{\Lambda}_{ij,k}(y) = \hat{\Lambda}^{(q)}(x'_{ij,k} \hat{\beta}(y) + \hat{\alpha}_i(y) + \hat{\gamma}_j(y)) \). Replacing \( \Lambda_{ij}^{(1)}(y) \) and \( \Lambda_{ij,k}^{(1)}(y) \) by \( \hat{\Lambda}_{ij}^{(1)}(y) \) and \( \hat{\Lambda}_{ij,k}^{(1)}(y) \) in the definitions of \( \alpha^F_{x}(y) \), \( \gamma^F_{x}(y) \), and \( \gamma^F(y) \) yields the corresponding estimators. We plug-in these estimators to obtain \( \hat{x}_{ij}(y) = x_{ij} - \hat{\alpha}_{x,i}(y) - \hat{\gamma}_{x,j}(y), \hat{\Psi}_{ij,k}(y) = x_{ij,k} - \hat{\alpha}_{x,i}(y) - \hat{\gamma}_{x,j}(y), \) and \( \hat{\Psi}_{ij,k}(y) = \alpha^F_{x}(y) + \hat{\gamma}^F_{x}(y) \). Then we construct

\[
\hat{W}(y) = \frac{1}{n} \sum_{(i,j) \in D} \hat{\Lambda}_{ij}^{(1)}(y) \hat{x}_{ij}(y) \hat{x}'_{ij}(y), \quad \partial_{\beta} \hat{F}_k(y) = \frac{1}{n} \sum_{(i,j) \in D} \hat{\Lambda}_{ij,k}^{(1)}(y) \hat{\Xi}_{ij,k}(y),
\]

and

\[
\hat{B}(\beta)(y) = -\frac{1}{2} \hat{W}^{-1}(y) \left[ \frac{1}{J} \sum_{j=1}^{J} \frac{\sum_{i \in D_j} \hat{\Lambda}_{ij}^{(2)}(y) \hat{x}_{ij}(y)}{\sum_{i \in D_j} \hat{\Lambda}_{ij}^{(1)}(y)} \right],
\]

\[
\hat{D}(\beta)(y) = -\frac{1}{2} \hat{W}^{-1}(y) \left[ \frac{1}{J} \sum_{j=1}^{J} \frac{\sum_{i \in D_j} \hat{\Lambda}_{ij}^{(2)}(y) \hat{x}_{ij}(y)}{\sum_{i \in D_j} \hat{\Lambda}_{ij}^{(1)}(y)} \right],
\]

\[
\hat{B}_k^{(F)}(y) = \frac{1}{2J} \sum_{j=1}^{J} \sum_{i \in D_j} \frac{\hat{\Lambda}_{ij,k}^{(2)}(y) - \hat{\Lambda}_{ij}^{(2)}(y) \hat{\Psi}_{ij,k}(y)}{\sum_{j \in D_j} \hat{\Lambda}_{ij}^{(1)}(y)},
\]

\[
\hat{D}_k^{(F)}(y) = \frac{1}{2J} \sum_{j=1}^{J} \sum_{i \in D_j} \frac{\hat{\Lambda}_{ij,k}^{(2)}(y) - \hat{\Lambda}_{ij}^{(2)}(y) \hat{\Psi}_{ij,k}(y)}{\sum_{j \in D_j} \hat{\Lambda}_{ij}^{(1)}(y)}.
\]

We also define the \( |K| \times d_x \) matrix \( \partial_{\beta} \hat{F}(y) = [(\partial_{\beta} \hat{F}_k(y)) : k \in K] \), and the \( |K| \times 1 \) vectors \( \hat{B}_k^{(F)}(y) = [\hat{B}_k^{(F)}(y) : k \in K], \hat{D}_k^{(F)}(y) = [\hat{D}_k^{(F)}(y) : k \in K], \hat{\Psi}_{ij}(y) = [\hat{\Psi}_{ij,k}(y) : k \in K]. \)

Finally, we also construct the estimator of the asymptotic variance of \( \hat{F}(y) \)

\[
\hat{\Omega}(y) = \frac{1}{n} \sum_{(i,j) \in D} \hat{\Lambda}_{ij}^{(1)}(y) \hat{\Xi}(y) \hat{\Xi}'(y),
\]

where \( \hat{\Xi}(y) = \hat{\Psi}_{ij}(y) + (\partial_{\beta} \hat{F}(y)) \hat{W}^{-1}(y) \hat{x}_{ij}(y) \).
The following theorem shows that the estimators of the asymptotic bias and variance are consistent, uniformly in \( y \in \mathcal{Y} \).

**Theorem 2** *(Uniform Consistency of Estimators of Bias and Variance Components).* Let Assumption 1 hold. Then,

\[
\sup_{y \in \mathcal{Y}} \left\| \hat{W}(y) - \bar{W}(y) \right\| = o_P(1), \quad \sup_{y \in \mathcal{Y}} \left\| \partial_\beta \hat{F}(y) - \partial_\beta F(y) \right\| = o_P(1),
\]

\[
\sup_{y \in \mathcal{Y}} \left\| \hat{B}(y) - B(y) \right\| = o_P(1), \quad \sup_{y \in \mathcal{Y}} \left\| \hat{D}(y) - D(y) \right\| = o_P(1),
\]

\[
\sup_{y \in \mathcal{Y}} \left\| \hat{\bar{\Omega}}(y) - \bar{\Omega}(y) \right\| = o_P(1).
\]

Bias corrected estimators for \( \beta(y) \) and \( F(y) \) are formed as

\[
\tilde{\beta}(y) = \hat{\beta}(y) - \frac{I}{n} \hat{B}(\beta)(y) - \frac{J}{n} \hat{D}(\beta)(y),
\]

and

\[
\tilde{F}(y) = \hat{F}(y) - \frac{I}{n} \left[ \hat{B}(F)(y) + (\partial_\beta \hat{F}(y)) \hat{B}(\beta)(y) \right] - \frac{J}{n} \left[ \hat{D}(F)(y) + (\partial_\beta \hat{F}(y)) \hat{D}(\beta)(y) \right].
\]

Alternatively, we could define the bias corrected version of \( \hat{F}(y) \) as

\[
\tilde{F}_k^*(y) = \left[ \frac{1}{n} \sum_{(i,j) \in D} \Lambda \left( x_{i,j,k} \tilde{\beta}(y) + \tilde{\alpha}_i(y) + \tilde{\gamma}_j(y) \right) \right] - \frac{I}{n} \hat{B}_k^*(y) - \frac{J}{n} \hat{D}_k^*(y),
\]

where \( \tilde{\xi}(y) := (\tilde{\alpha}_1(y), \ldots, \tilde{\alpha}_I(y), \tilde{\gamma}_1(y), \ldots, \tilde{\gamma}_J(y)) \) is the solution to

\[
\max_{\xi \in \mathbb{R}^{I+J}} \sum_{(i,j) \in D} \left( 1 \{ y_{ij} \leq y \} \log \Lambda(x_{ij} \tilde{\beta}(y) + \alpha_i + \gamma_j) + 1 \{ y_{ij} > y \} \log[1 - \Lambda(x_{ij} \tilde{\beta}(y) + \alpha_i + \gamma_j)] \right).
\]

It can be shown that \( \sup_{y \in \bar{\mathcal{Y}}} \sqrt{n} \left| \tilde{F}_k^*(y) - \tilde{F}_k(y) \right| = o_P(1) \), that is, the difference between those alternative bias corrected estimators is asymptotically negligible. There is no obvious reason to prefer one over the other, and we present result for \( \tilde{F}_k \) in the following, which equivalently hold for \( \tilde{F}_k^* \).

Combining Theorem 1 and 2 we obtain the following functional central limit theorem for the bias corrected estimators.

\[\text{We use the estimator } \tilde{F}_k^* \text{ in the numerical examples for computational convenience as the bias correction involves estimating less terms.}\]
Corollary 1 (FCLT for Bias Corrected Fixed Effects DR Estimators). Let Assumption 1 hold. Then, in the metric space $\ell^\infty(Y)^d$,
\[
\sqrt{n} \left[ \hat{\beta}(y) - \beta(y) \right] \rightsquigarrow Z^{(\beta)}(y),
\]
and, in the metric space $\ell^\infty(Y)^{|K|}$,
\[
\sqrt{n} \left[ \hat{F}(y) - F(y) \right] \rightsquigarrow Z^{(F)}(y),
\]
as stochastic processes indexed by $y \in Y$, where $Z^{(\beta)}(y)$ and $Z^{(F)}(y)$ are the same Gaussian processes that appear in Theorem 1.

4.3. Uniform Confidence Bands and Bootstrap. We show how to construct pointwise and uniform confidence bands for $y \mapsto \beta(y)$ and $y \mapsto F(y)$ on $\bar{Y}$ using Corollary 1. The uniform bands for $F$ can be used as inputs in Lemmas 1 and 2 to construct uniform bands for the QFs $\tau \mapsto Q_k(\tau) = F_k^{-1}(\tau)$, $k \in K$, and the QEF $\tau \mapsto \Delta(\tau)$ on $T$.

Let $B \subseteq \{1, \ldots, d_x\}$ be the set of indexes for the coefficients of interest. For given $y \in \bar{Y}$, $\ell \in B$, $k \in K$, and $p \in (0,1)$, a pointwise $p$-confidence interval for $\beta_\ell(y)$, the $\ell$th component of $\beta(y)$, is
\[
[\hat{\beta}_\ell(y) \pm \Phi^{-1}(1-p/2)\hat{\sigma}_\beta(y)],
\]
and a pointwise $p$-confidence intervals for $F_k(y)$ is
\[
[\hat{F}_k(y) \pm \Phi^{-1}(1-p/2)\hat{\sigma}_F(y)],
\]
where $\Phi$ denotes the cdf of the standard normal distribution, $\hat{\sigma}_\beta(y)$ is the standard error of $\hat{\beta}(y)$ given in (4.3), and $\hat{\sigma}_F(y)$ is the standard error of $\hat{F}(y)$ given in (4.6). These intervals have coverage $p$ in large samples by Corollary 1 and Theorem 2.

We construct joint uniform bands for the coefficients and distributions using Kolmogorov-Smirnov type critical values, instead of quantiles from the normal distribution. A uniform $p$-confidence band joint for the vector of functions $\{\beta_\ell(y) : \ell \in B, y \in \bar{Y}\}$ is
\[
I_\beta = \{[\hat{\beta}_\ell(y) \pm t^{(\beta)}_{B,\bar{Y}}(p)\hat{\sigma}_\beta(y)] : \ell \in B, y \in \bar{Y}\},
\]
where $t^{(\beta)}_{B,\bar{Y}}(p)$ is the $p$-quantile of the maximal $t$-statistic
\[
t^{(\beta)}_{B,\bar{Y}} = \sup_{y \in \bar{Y}, \ell \in B} \frac{|Z^{(\beta)}(y)|}{\hat{\sigma}^{(\beta)}_\ell(y)}.
\]
Similarly, a uniform $p$-confidence band joint for the set of distribution functions $\{F_k(y) : k \in K, y \in \bar{Y}\}$ is
\[
I_F = \{[\hat{F}_k(y) \pm t^{(F)}_{K,\bar{Y}}(p)\hat{\sigma}_F(y)] : k \in K, y \in \bar{Y}\},
\]
where $t^{(F)}_{B,Y}(p)$ is the $p$-quantile of the maximal $t$-statistic

$$t^{(F)}_{B,Y} = \sup_{y \in \bar{Y}, k \in K} \frac{|Z^{(F)}_{k}(y)|}{\hat{\sigma}^{(F)}_{k}(y)}.$$

The previous confidence bands also have coverage $p$ in large samples by Corollary 1 and Theorem 2.

The maximal $t$-statistics used to construct the bands $I_{\beta}$ and $I_{F}$ are not pivotal, but their distributions can approximated by simulation after replacing the variance functions of the limit processes by uniformly consistent estimators. In practice, however, we find more convenient to use resampling methods. We consider a multiplier bootstrap scheme that resamples the efficient scores or influence functions of the fixed effects estimators $\hat{\beta}(y)$ and $\hat{F}(y)$. This scheme is computationally convenient because it does not need to solve the high dimensional nonlinear fixed effects conditional maximum likelihood program (3.1) or making any bias correction in each bootstrap replication. In these constructions we rely on the uncorrected fixed effects estimators instead of the bias corrected estimators, because they have the same asymptotic variance and the uncorrected estimators are consistent under the asymptotic approximation that we consider.

To describe the standard errors and multiplier bootstrap we need to introduce some notation for the influence functions of $\hat{\theta}(y)$ and $\hat{F}(y)$. Let $\theta = (\beta, \alpha_{1}, \ldots, \alpha_{I}, \gamma_{1}, \ldots, \gamma_{J})$ be a generic value for the parameter $\theta(y)$, the influence function for $\hat{\theta}(y)$ is the $(d_{x} + I + J)$-vector $\psi^{y}_{ij}(\theta(y))$, where

$$\psi^{y}_{ij}(\theta) = H(\theta)^{-1}[1\{y_{ij} \leq y\} - \Lambda(x'_{ij}\beta + \alpha_{i} + \gamma_{j})]w_{ij}, \quad w_{ij} = (x_{ij}, e_{i,I}, e_{j,J}), \quad y \in \bar{Y},$$

$e_{i,I}$ is a unit vector of dimension $I$ with a one in the position $i$, $e_{j,J}$ is defined analogously, and

$$H(\theta) = \frac{1}{n} \sum_{(i,j) \in D} \lambda(x'_{ij}\beta + \alpha_{i} + \gamma_{j})w_{ij}w'_{ij}, \quad \lambda(u) = \Lambda(u)\Lambda(-u),$$

is minus the Hessian of the log-likelihood with respect to $\theta$, which does not depend on $y$ in the case of the logistic distribution. The influence function for $\hat{F}(y)$ is $\varphi^{y}_{ij,k}(\theta(y))$, where

$$\varphi^{y}_{ij,k}(\theta) = J_{k}(\theta)^{t}\psi^{y}_{ij}(\theta),$$

and

$$J_{k}(\theta) = \frac{1}{n} \sum_{(i,j) \in D} \lambda(x'_{ij,k}\beta + \alpha_{i} + \gamma_{j})w_{ij,k}, \quad w_{ij,k} = (x_{ij,k}, e_{i,I}, e_{j,J}).$$
The standard error of $\tilde{\beta}_k(y)$ is constructed as

$$
\tilde{\sigma}_{\beta_k}(y) = n^{-1} \left[ \sum_{(i,j) \in D} \psi^y_{ij}(\hat{\theta}(y)) \psi^y_{ij}(\hat{\theta}(y))' \right]^{1/2},
$$

the square root of the $(\ell, \ell)$ element of the sandwich matrix $n^{-2} \sum_{(i,j) \in D} \psi^y_{ij}(\hat{\theta}(y)) \psi^y_{ij}(\hat{\theta}(y))'$. Similarly, the standard error of $\tilde{F}_k(y)$ is constructed as

$$
\tilde{\sigma}_{F_k}(y) = n^{-1} \left[ \sum_{(i,j) \in D} \phi^y_{ij,k}(\hat{\theta}(y))^2 \right]^{1/2}.
$$

The following algorithm describes a multiplier bootstrap-t scheme to obtain the critical values for a set of parameters indexed by $\ell \in B \subseteq \{1, \ldots, d_x\}$ and a set of distributions indexed by $k \in K \subseteq \{0, 1\}$. This scheme is based on perturbing the first order conditions of the fixed effects estimators with random multipliers independent from the data.

**Algorithm 1** (Multiplier Bootstrap). (1) Draw the bootstrap multipliers $\{\omega^m_{ij} : (i, j) \in D\}$ i.i.d. from some distribution with zero mean and unit variance, independently from the data. For example, $\omega^m_{ij} = \tilde{\omega}^m_{ij} = \sum_{(i,j) \in D} \tilde{\omega}^m_{ij} / n$, $\tilde{\omega}^m_{ij} \sim$ i.i.d. $\mathcal{N}(0,1)$. Here we have normalized the multipliers to have zero mean as a finite-sample adjustment. (2) For each $y \in \tilde{Y}$, obtain the bootstrap draws of the estimator of parameter $\hat{\theta}^m(y) = \hat{\theta}(y) + n^{-1} \sum_{(i,j) \in D} \omega^m_{ij} \psi^y_{ij}(\hat{\theta}(y))$, and of the estimator of the distribution functions $\tilde{F}^m_k(y) = \tilde{F}_k(y) + n^{-1} \sum_{(i,j) \in D} \omega^m_{ij} \phi^y_{ij,k}(\hat{\theta}(y)), k \in K$. (3) Construct the bootstrap draw of the maximal $t$-statistic for the parameters, $t^{(\beta),m}_{B,\tilde{Y}} = \max_{y \in \tilde{Y}, \ell \in B} |\hat{\beta}_{\ell}(y) - \beta_{\ell}(y)| / \tilde{\sigma}_{\beta_{\ell}}(y)$, where $\tilde{\sigma}_{\beta_{\ell}}(y)$ is the multiplier bootstrap standard error of $\hat{\beta}_{\ell}(y)$, that is $\tilde{\sigma}_{\beta_{\ell}}^2(y) = n^{-2} \sum_{(i,j) \in D} [\omega^m_{ij} \phi_{ij,\ell}(\hat{\theta}(y))]^2$, and $\phi_{ij,\ell}(\theta)$ is the component of $\psi^y_{ij}(\theta)$ corresponding to $\beta_{\ell}$. Similarly, construct the bootstrap draw of the maximal $t$-statistic for the distributions, $t^{(F),m}_{K,\tilde{Y}} = \max_{y \in \tilde{Y}, k \in K} |\tilde{F}^m_k(y) - \tilde{F}_k(y)| / \tilde{\sigma}_{F_k}^m(y)$, where $\tilde{\sigma}_{F_k}^m(y)$ is the multiplier bootstrap standard error of $\tilde{F}_k^m(y)$, that is $\tilde{\sigma}_{F_k}^2(y) = n^{-2} \sum_{(i,j) \in D} [\omega^m_{ij} \varphi_{ij,k}(\hat{\theta}(y))]^2$. (4) Repeat steps (1)–(3) $M$ times and index the bootstrap draws by $m \in \{1, \ldots, M\}$. In the numerical examples we set $M = 500$. (5) Obtain the bootstrap estimators of the critical values as

$$
\tilde{t}^{(\beta),m}_{B,\tilde{Y}}(p) = p - \text{quantile of } \{t^{(\beta),m}_{B,\tilde{Y}} : 1 \leq m \leq M\},
$$

$$
\tilde{t}^{(F),m}_{K,\tilde{Y}}(p) = p - \text{quantile of } \{t^{(F),m}_{K,\tilde{Y}} : 1 \leq m \leq M\}.
$$

### 4.4. Pairwise Clustering Dependence or Reciprocity

The conditional independence of Assumption 1(i) can be relaxed to allow for some forms of conditional weak dependence.
A form of dependence that is particularly relevant for network data is pairwise clustering or reciprocity where the observational units with symmetric indexes \((i,j)\) and \((j,i)\) might be dependent due to unobservable factors not accounted by unobserved effects. In the trade application, for example, these factors might include distributional channels or multinational firms operating in both countries.

The presence of reciprocity does not change the bias of the fixed effects estimators, but affects the standard errors and the implementation of the multiplier bootstrap. The standard error of \(\hat{\beta}_t(y)\) becomes

\[
\hat{\sigma}_{\hat{\beta}_t}(y) = n^{-1} \left[ \sum_{(i,j) \in D} \left\{ \psi_{ij}^y(\hat{\theta}(y)) + \psi_{ji}^y(\hat{\theta}(y)) \right\} \psi_{ij}^y(\hat{\theta}(y))^\prime \right]^{1/2}. \tag{4.5}
\]

Similarly, the standard error of \(\hat{F}_k(y)\) needs to be adjusted to

\[
\hat{\sigma}_{\hat{F}_k}(y) = n^{-1} \left[ \sum_{(i,j) \in D} \left\{ \varphi_{ij,k}^y(\hat{\theta}(y)) + \varphi_{ji,k}^y(\hat{\theta}(y)) \right\} \varphi_{ij,k}^y(\hat{\theta}(y))^\prime \right]^{1/2}. \tag{4.6}
\]

In the previous expressions we assume that if \((i,j) \in D\) then \((j,i) \in D\) to simplify the notation. The modified multiplier bootstrap algorithm becomes:

**Algorithm 2** (Multiplier Bootstrap with Pairwise Clustering). 1. Draw the bootstrap multipliers \(\{\omega_{ij}^m : (i,j) \in D\}\) i.i.d. from some distribution with zero mean and unit variance, independently from the data such that \(\omega_{ji}^m = \omega_{ij}^m\). For example, \(\omega_{ij}^m = \tilde{\omega}_{ij}^m - \sum_{(i,j) \in D} \tilde{\omega}_{ij}^m / n, \quad \tilde{\omega}_{ij}^m \sim i.i.d. N(0,1), \quad i \leq j\). 2. For each \(y \in \tilde{Y}\), obtain the bootstrap draws of the estimator of parameter \(\hat{\theta}_m(y) = \hat{\theta}(y) + n^{-1} \sum_{(i,j) \in D} \omega_{ij}^m \psi_{ij}^y(\hat{\theta}(y))\), and of the estimator of the distribution functions \(\hat{F}_{km}(y) = \hat{F}_k(y) + n^{-1} \sum_{(i,j) \in D} \omega_{ij}^m \varphi_{ij,k}^y(\hat{\theta}(y))\), \(k \in K\). 3. Construct the bootstrap draw of the maximal t-statistic for the parameters, \(t_{B,Y}^{(\beta),m} = \max_{y \in \tilde{Y}, t \in B} |\hat{\beta}_t^m(y) - \hat{\beta}_t(y)| / \hat{\sigma}_{\hat{\beta}_t}(y)\), where \(\hat{\sigma}_{\hat{\beta}_t}(y)\) is the multiplier bootstrap standard error of \(\hat{\beta}_t^m(y)\), that is \(\hat{\sigma}_{\hat{\beta}_t}(y) = n^{-2} \sum_{(i,j) \in D} (\omega_{ij}^m)^2 \left\{ \psi_{ij,t}^y(\hat{\theta}(y)) + \psi_{ji,t}^y(\hat{\theta}(y)) \right\} \psi_{ij,t}^y(\hat{\theta}(y))\), and \(\psi_{ij,t}^y(\hat{\theta})\) is the component of \(\psi_{ij}^y(\hat{\theta})\) corresponding to \(\beta_t\). Similarly, construct the bootstrap draw of the maximal t-statistic for the distributions, \(t_{K,Y}^{(F),m} = \max_{y \in \tilde{Y}, k \in K} |\hat{F}_{km}(y) - \hat{F}_k(y)| / \hat{\sigma}_{\hat{F}_k}(y)\), where \(\hat{\sigma}_{\hat{F}_k}(y)\) is the multiplier bootstrap standard error of \(\hat{F}_{km}(y)\), that is \(\hat{\sigma}_{\hat{F}_k}(y) = n^{-2} \sum_{(i,j) \in D} (\omega_{ij}^m)^2 \left\{ \varphi_{ij,k}(\hat{\theta}(y)) + \varphi_{ji,k}(\hat{\theta}(y)) \right\} \varphi_{ij,k}(\hat{\theta}(y))\). 4. Repeat steps (1)–(3) \(M\) times and index the bootstrap draws by \(m \in \{1, \ldots, M\}\). In the numerical examples we set \(M = 500\). 5. Obtain the bootstrap estimators of the critical values as

\[
\hat{t}_{B,Y}^{(\beta)}(p) = p - \text{quantile of } \{t_{B,Y}^{(\beta),m} : 1 \leq m \leq M\},
\]

\[
\hat{t}_{K,Y}^{(F)}(p) = p - \text{quantile of } \{t_{K,Y}^{(F),m} : 1 \leq m \leq M\}.
\]
The clustered multiplier bootstrap preserves the dependence in the symmetric pairs \((i, j)\) and \((j, i)\) by assigning the same multiplier to each of these pairs.

5. **Quantile Effects in Gravity Equations for International Trade**

We consider an empirical application to gravity equations for bilateral trade between countries. We use data from Helpman et al. (2008), extracted from the Feenstra’s World Trade Flows, CIA’s World Factbook and Andrew Rose’s web site. These data contain information on bilateral trade flows and other trade-related variables for 157 countries in 1986. The data set contains network data where both \(i\) and \(j\) index countries as senders (exporters) and receivers (importers), and therefore \(I = J = 157\). The outcome \(y_{ij}\) is the volume of trade in thousands of constant 2000 US dollars from country \(i\) to country \(j\), and the covariates \(x_{ij}\) include determinants of bilateral trade flows such as the logarithm of the distance in kilometers between country \(i\)’s capital and country \(j\)’s capital and indicators for common colonial ties, currency union, regional free trade area (FTA), border, legal system, language, and religion. Following Anderson and van Wincoop (2003), we include unobserved importer and exporter country effects. These effects control for other country specific characteristics that may affect trade such as GDP, tariffs, population, institutions, infrastructures or natural resources. We allow for these characteristics to affect differently the imports and exports of each country, and be arbitrarily related with the observed covariates. Table 1 reports descriptive statistics of the variables used in the analysis. There are \(157 \times 156 = 24,492\) observations corresponding to different pairs of countries. The observations with \(i = j\) are missing because we do not observe trade flows from a country to the same country. The trade variable in the first row is an indicator for positive volume of trade. Here we finds that there are no trade flows for 55% of the country pairs.

The previous literature estimated nonlinear parametric models such as Poisson, Negative Binomial, Tobit and Heckman-selection models to deal with the large number of zeros in the volume of trade (e.g., Eaton and Kortum, 2001, Santos Silva and Tenreyro, 2006, and Helpman et al., 2008). These models impose strong conditions on the process that generates the zeros and/or on the conditional heteroskedasticity of the volume of trade. The DR model deals with zeros and any other fixed censoring points in a very flexible and natural fashion as it specifies the conditional distribution separately at each point. In particular, the model coefficients at zero can be arbitrarily different from the model coefficients at other values of the volume of trade. Moreover, the DR model can also accommodate very flexible patterns of conditional heteroskedasticity.

---

6 The original data set includes 158 countries. We exclude Congo because it did not export to any other country in 1986.
7 See Harrigan (1994) for an earlier empirical international trade application that includes unobserved country effects.
8 See Head and Mayer (2014) for a recent survey on gravity equations in international trade.
Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>Trade Volume</td>
<td>84,542</td>
<td>1,082,219</td>
</tr>
<tr>
<td>Log Distance</td>
<td>4.18</td>
<td>0.78</td>
</tr>
<tr>
<td>Legal</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>Language</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>Religion</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>Border</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>Currency</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>FTA</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Colony</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Country Pairs</td>
<td>24,492</td>
<td></td>
</tr>
</tbody>
</table>

Source: Helpman, Melitz and Rubinstein (08)

Figure 1 shows estimates and 95% pointwise confidence intervals for the DR coefficients of log distance and common legal system plotted against the quantile indexes of the volume of trade. We report uncorrected and bias corrected fixed effects estimates obtained from (3.1) and (4.1), respectively. The confidence intervals are constructed using (4.2). The x-axis starts at .54, the maximum quantile index corresponding to zero volume of trade. As predicted by the theory in Section 4, the difference between the uncorrected and bias corrected estimates is the same order of magnitude as the width of the confidence intervals, specially for the coefficient of log distance. We find the largest estimated biases for both coefficients at highest quantiles of the volume of trade, where the indicators $1\{y_{ij} \leq y\}$ take on many ones. The signs of the DR coefficients indicate that increasing distance has a negative effect and having a common legal system has a positive effect on the volume of trade throughout the distribution. Recall that the sign of the effect in terms of volume of trade, $y_{ij}$, is the opposite to the sign of the DR coefficient.

Figures 2 and 3 show estimates and 95% uniform confidence bands for distribution and quantile functions of the volume of trade at different values of the log of distance and the common legal system. The left panels plot the functions when distance takes the observed levels (dist) and two times the observed values (2*dist), i.e. when we counterfactually double all the distances between the countries. The right panels plot the functions when all the countries have the same legal system (legal=1) and different systems (legal=0). All the functions are plotted against the quantile indexes of the volume of trade. The confidence bands for the distribution are obtained by Algorithm 1 with 500 bootstrap replications and standard normal multipliers, and a grid of values $\bar{Y}$ that includes the sample quantiles of the volume of trade with indexes $\{ 0.54, 0.55, \ldots, 0.95 \}$. The bands are joint for the two functions displayed in each panel. The confidence bands for the quantile functions are
Figure 1. Estimates and 95% pointwise confidence intervals for the DR coefficients of log distance and common legal system.

obtained by inverting and rotating the bands for the corresponding distribution functions using Lemma 1.

Figure 4 displays estimates and 95% uniform confidence bands for the quantile effects of the log of distance and the common legal system on the volume of trade, constructed using Lemma 2. For comparison, we also include estimates from a Poisson model. Here, we replace the DR estimators of the distributions by

$$
\hat{F}_k(y) = \frac{1}{n} \sum_{(ij) \in D} \exp \lambda_{ij,k} \sum_{y=0}^{|y|} \frac{\lambda_{ij,k} y}{y!}, \quad k \in K,
$$

where $|y|$ is the integer part of $y$, $\lambda_{ij,k} = \exp(x'_{ij,k} \beta + \alpha_i + \gamma_j)$, and $\hat{\theta} = (\hat{\beta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_I, \hat{\gamma}_1, \ldots, \hat{\gamma}_J)$ is the Poisson fixed effects maximum likelihood estimator

$$
\hat{\theta} \in \arg \max_{\theta \in \mathbb{R}^{d_x + I + J}} \sum_{(ij) \in B} [y_{ij}(x'_{ij} \beta + \alpha_i + \gamma_j) - \exp(x'_{ij} \beta + \alpha_i + \gamma_j)].
$$

We find that distance and common legal system have heterogenously increasing effects along the distribution. For example, the negative effects of doubling the distance grows more than proportionally as we move up to the upper tail of the distribution of volume of trade. Putting all the countries under the same legal system has little effects in the extensive margin of trade, but has a strong positive effect at the upper tail of the distribution. The Poisson estimates lie outside the DR confidence bands reflecting overdispersion in the
conditional distribution of the volume of trade that is missed by the Poisson model.\footnote{This overdispersion problem with the Poisson model is well-known in the international trade literature. The Poisson estimator is treated as a quasilikelihood estimator and standard errors robust to misspecification are reported (Santos Silva and Tenreyro, 2006).}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Figure 2. Estimates and 95\% uniform confidence bands for distribution functions of the volume of trade.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{Figure 3. Estimates and 95\% uniform confidence bands for quantile functions of the volume of trade.}
\end{figure}
5 shows confidence bands of the quantile effects that account for pairwise clustering. The bands are constructed from confidence bands from the distributions using Algorithm 2 with 500 bootstrap draws and standard normal multipliers. Accounting for unobservables that affect symmetrically to the country pairs increases the width of the bands, specially for high quantiles in the effect of the legal system.

\[ y_{ij}^s = \sup\{ y \in \mathcal{Y} : x_{ij}' \hat{\beta}(y) + \hat{\alpha}_i(y) + \hat{\gamma}_j(y) \leq \Lambda^{-1}(u_{ij}^s) \}, \quad u_{ij}^s \sim \text{i.i.d } \mathcal{U}(0, 1), \quad (i, j) \in \mathcal{D}, \]

where \( \{[\hat{\beta}(y), \hat{\alpha}_1(y), \ldots, \hat{\alpha}_I(y), \hat{\gamma}_1(y), \ldots, \hat{\gamma}_J(y)] : y \in \mathcal{Y} \} \) are the fixed effect estimates of the parameters, \( \mathcal{Y} \) is a set that includes the sample quantiles of the volume of trade with indexes \( \{.54, .55, \ldots, .99\} \) in the trade dataset, \( \mathcal{D} = \{(i, j) : 1 \leq i, j \leq 157, i \neq j\} \), and \( x_{ij} \) are the value of the covariates for the observational unit \((i, j)\) in the trade data set. This yields a simulated dataset \( \{(y_{ij}^s, x_{ij}) : (i, j) \in \mathcal{D}\} \). All the results are based on 200 simulations.

Figure 6 reports the bias, standard deviation and root mean square error for the fixed effects estimator of the DR coefficient of log-distance as a function of the quantiles of
Figure 5. Estimates and 95% uniform confidence bands for the quantile effects of log distance and common legal system on the volume of trade.

\[ y_{ij} \]. All the results are in percentage of the true value of the parameter. As predicted by the large sample theory, the fixed effects estimator displays a bias of the same order of magnitude as the standard deviation. The bias correction removes most of the bias and does not increase the standard deviation, yielding reductions in the rmse of more than 6% at high quantile indexes.

[APPENDIX TO BE ADDED]

References

Figure 6. Bias, standard deviation and root mean squared error for the DR coefficients of log-distance and legal.


