Econometrica, Vol. 84, No. 2 (March, 2016), 809–833

IV QUANTILE REGRESSION FOR GROUP-LEVEL TREATMENTS, WITH AN APPLICATION TO THE DISTRIBUTIONAL EFFECTS OF TRADE

BY DENIS CHETVERIKOV, BRADLEY LARSEN, AND CHRISTOPHER PALMER[1]

We present a methodology for estimating the distributional effects of an endogenous treatment that varies at the group level when there are group-level unobservables, a quantile extension of Hausman and Taylor (1981). Because of the presence of group-level unobservables, standard quantile regression techniques are inconsistent in our setting even if the treatment is independent of unobservables. In contrast, our estimation technique is consistent as well as computationally simple, consisting of group-by-group quantile regression followed by two-stage least squares. Using the Bahadur representation of quantile estimators, we derive weak conditions on the growth of the number of observations per group that are sufficient for consistency and asymptotic zero-mean normality of our estimator. As in Hausman and Taylor (1981), micro-level covariates can be used as internal instruments for the endogenous group-level treatment if they satisfy relevance and exogeneity conditions. Our approach applies to a broad range of settings including labor, public finance, industrial organization, urban economics, and development; we illustrate its usefulness with several such examples. Finally, an empirical application of our estimator finds that low-wage earners in the United States from 1990 to 2007 were significantly more affected by increased Chinese import competition than high-wage earners.

KEYWORDS: Quantile regression, instrumental variables, panel data, income inequality, import competition.

1. INTRODUCTION

IN CLASSICAL PANEL DATA MODELS FOR MEAN REGRESSION, fixed effects are commonly used to obtain identification when time-invariant unobservables are correlated with included variables. While this approach yields consistent estimates of the coefficients on time-varying variables, it precludes identification of the coefficients of any time-invariant variables, as these variables are eliminated by the within-group transformation. In an influential paper, Hausman and Taylor (1981) demonstrated that exogenous between variation of time-varying variables can help to identify the coefficients of time-invariant variables after their within variation has been used to identify the coefficients on time-varying variables, thus yielding identification of the whole model without external instruments. Our paper provides a quantile extension of the Hausman and Taylor (1981) classical linear panel estimator.

[1]We thank the editor and four anonymous referees for their detailed comments and feedback; Moshe Buchinsky, Ivan Canay, Brigham Frandsen, Antonio Galvao, Wenshu Guo, Jerry Hausman, Rosa Matzkin, Whitney Newey, and Christopher Taber for helpful comments; and Sam Hughes, Yuqi Song, and Caio Waisman for meticulous research assistance. We are especially grateful to Jin Hahn for many useful discussions.

 DOI: 10.3982/ECTA12121

We present our model in Section 2. To clarify the range of potential applications of our estimator, we depart in the model from the usual panel data terminology and refer to panel units as groups (instead of as individuals; groups might be states, cities, schools, etc.) and to within-group observations as individuals or micro-level observations (instead of as time observations; individuals might be students, families, firms, etc.).[2] The model is of practical significance when the researcher has data on a group-level endogenous treatment and has microdata on the outcome of interest within each group. For example, a researcher may be interested in the effect of a policy which varies across states and years (a "group") on the within-group distribution of micro-level outcomes. In Section 2, we also explain how the problem we solve differs from others in the quantile regression literature, and we demonstrate that, as in Hausman and Taylor (1981), micro-level covariates can be used as internal instruments for the endogenous group-level treatment if they satisfy relevance and exogeneity conditions. This last feature of the model is especially appealing because, in practice, it may be difficult to find external instruments.

In Section 3, we introduce our estimation approach, which we refer to as grouped IV quantile regression. The estimator is computationally simple to implement and consists of two steps: (i) perform quantile regression within each group to estimate effects of micro-level covariates, or, if no micro-level covariates are included, calculate the desired quantile for the outcome within each group; and (ii) regress the estimated group-specific effects on group-level covariates using either 2SLS, if the group-level covariates are endogenous, or OLS, if the group-level covariates are exogenous, either of which cases would render standard quantile regression (e.g., Koenker and Bassett (1978)) inconsistent.[3] Section 3 also discusses Monte Carlo simulations (found in Appendix A of the Supplemental Material (Chetverikov, Larsen, and Palmer (2016))) that demonstrate that our estimator has much lower bias than that of the standard quantile regression estimator when the group-level treatment is endogenous, even in small samples, and at larger sample sizes our estimator outperforms quantile regression even when the treatment is exogenous. Section 3 also highlights additional computational benefits of our estimator.

Section 4 provides a variety of examples illustrating the use of the grouped IV quantile regression estimator. In particular, we use examples from Angrist and Lang (2004), Larsen (2014), Palmer (2011), and Backus (2014) to illustrate applicability of our estimator. In addition to these examples, the grouped

[2]Similar terminology was used, for example, by Altonji and Matzkin (2005).

[3]Even in the absence of endogeneity, the Koenker and Bassett (1978) estimator will be inconsistent in our setting because of group-level unobservables, akin to left-hand-side measurement error; see Section 2 for details on our setting. While posing no problems for linear models, left-hand-side errors-in-variables can bias quantile estimation (see Hausman (2001) and Hausman, Luo, and Palmer (2014)).

quantile approach can apply to a wide range of settings in labor, industrial organization, trade, public finance, development, and other applied fields.

We derive theoretical properties of the estimator in Section 5. The results are based on asymptotics where both the number of groups and the number of observations per group grow to infinity. While linear panel models, including Hausman and Taylor (1981), admit a simple *unbiased* fixed effects estimator and hence do not require asymptotics in the number of observations per group, quantile estimators are *biased in finite samples* leading to inconsistency of our estimator if the number of observations per group remains small as the number of groups increases, and making the estimator inappropriate in the settings with a small number of observations per group and a large number of groups. However, since quantile estimators are *asymptotically unbiased*, we are able to employ Bahadur's representation of quantile estimators to derive weak conditions on the growth of the number of observations per group that are sufficient for the consistency and asymptotic zero-mean normality of our estimator. Importantly, the attractive theoretical properties of the estimator remain valid even if the number of observations per group is relatively small in comparison with the number of groups. We demonstrate that standard errors for the proposed estimator can be obtained using traditional heteroscedasticity-robust variance estimators for 2SLS, making inference particularly simple. In the Supplemental Material, we also discuss clustered standard errors, and we show how to construct confidence bands for the coefficient of interest which hold uniformly over a set of quantiles via multiplier bootstrap procedure.

Section 6 presents an empirical application which studies the effect of trade on the distribution of wages within local labor markets. We build on the work of Autor, Dorn, and Hanson (2013), who studied the effect of Chinese import competition on average wages in local labor markets.

Using the grouped IV quantile regression approach developed here, we find that Chinese import competition reduced the wages of low-wage earners (individuals at the bottom quartile of the conditional wage distribution) more than high-wage earners, particularly for females, heterogeneity which is missed by focusing on traditional 2SLS estimates.

To the best of our knowledge, our paper is the first to present a framework for estimating distributional effects as a function of group-level covariates. There is, however, a large literature studying quantile models for panel data when the researcher wishes to estimate distributional effects of micro-level covariates. See, for example, Koenker (2004), Abrevaya and Dahl (2008), Lamarche (2010), Canay (2011), Galvao (2011), Kato and Galvao (2011), Ponomareva (2011), Kato, Galvao, and Montes-Rojas (2012), Rosen (2012), Arellano and Bonhomme (2013), and Galvao and Wang (2013). Our paper also contributes to the growing literature on IV treatment effects in quantile models, such as Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005, 2006, 2008), Lee (2007), Chesher (2003), and Imbens and Newey (2009).

Our paper differs, however, in that this literature focuses on the case where individual-level unobserved heterogeneity is correlated with an individual-level treatment, whereas we focus on the case where a group-level, additively separable unobservable is correlated with a group-level treatment.

Throughout the paper, we use the following notation. The symbol $\|\cdot\|$ denotes the Euclidean norm. The symbol $\Rightarrow$ signifies weak convergence, and $l^\infty(\mathcal{U})$ represents the set of bounded functions on $\mathcal{U}$. With some abuse of notation, $\ell^\infty(\mathcal{U})$ also denotes the set of component-wise bounded vector-valued functions on $\mathcal{U}$. All equalities and inequalities concerning random variables are implicitly assumed to hold almost surely. All proofs and some extensions of our results are contained in the Supplemental Material.

2. MODEL

We study a panel data quantile regression model for a response variable y_{ig} of individual i in group g. We first present a simple version of the model, which we consider as most appealing in empirical work, and then present the general version of our model, which allows for more flexible distributional effects. Our estimator and theoretical results apply to both the general and simple versions of the model.

In the simple version of the model, we assume that the uth quantile of the conditional distribution of y_{ig} is given by

$$Q_{y_{ig}|\tilde{z}_{ig},x_g,\varepsilon_g}(u) = \tilde{z}_{ig}'\gamma(u) + x_g'\beta(u) + \varepsilon_g(u), \quad u \in \mathcal{U}, \tag{1}$$

where $Q_{y_{ig}|\tilde{z}_{ig},x_g,\varepsilon_g}(u)$ is the uth conditional quantile of y_{ig} given $(\tilde{z}_{ig}, x_g, \varepsilon_g)$, $\tilde{z}_{ig}$ is a (d_z-1)-vector of observable individual-level covariates (which we sometimes refer to as micro-level covariates), x_g is a d_x-vector of observable group-level covariates (x_g contains a constant), $\gamma(u)$ and $\beta(u)$ are (d_z-1)- and d_x-vectors of coefficients, $\varepsilon_g = \{\varepsilon_g(u), u \in \mathcal{U}\}$ is a set of unobservable group-level random scalar shifters,[4] and $\mathcal{U}$ is a set of quantile indices of interest. Here, $\gamma(u)$ and $\beta(u)$ represent the effects of individual- and group-level covariates, respectively. In this paper, we are primarily interested in estimating $\beta(u)$, although we also provide some new results on estimating $\gamma(u)$.

In the more general version of the model, of which (1) is a special case, we assume that the uth quantile of the conditional distribution of y_{ig} is given by

$$Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u) = z_{ig}'\alpha_g(u), \quad u \in \mathcal{U}, \tag{2}$$

$$\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u), \quad u \in \mathcal{U}, \tag{3}$$

[4] One interpretation of the term $\varepsilon_g(u)$ in (1) is that it accounts for all unobservable group-level covariates η_g that affect the distribution of y_{ig} but are not included in x_g. In this case, $\varepsilon_g(u) = \varepsilon(u, \eta_g)$. Note that we do not impose any parametric restrictions on $\varepsilon(u, \eta_g)$, and so we allow for arbitrary nonlinear effects of the group-level unobservable covariates that can affect different quantiles in different ways.

where $Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u)$ is the uth conditional quantile of y_{ig} given (z_{ig}, x_g, α_g), z_{ig} is a d_z-vector of observable individual-level covariates, $\alpha_g = \{\alpha_g(u), u \in \mathcal{U}\}$ is a set of (random) group-specific effects with $\alpha_{g,1}(u)$ being the *first component* of the vector $\alpha_g(u) = (\alpha_{g,1}(u), \dots, \alpha_{g,d_z}(u))'$, and all other notation is the same as above. In this model, we assume that the response variable y_{ig} satisfies the quantile regression model in (2) with group-specific effects $\alpha_g(u)$. We are primarily interested in studying how these effects depend on the group-level covariates x_g, and, without loss of generality, we focus on $\alpha_{g,1}(u)$, the first component of the vector $\alpha_g(u)$. To make the problem operational, we assume that $\alpha_{g,1}(u)$ satisfies the linear regression model (3), in which we are interested in estimating the vector of coefficients $\beta(u)$.

Observe that the model (1) is a special case of the model (2)–(3). Indeed, setting $z_{ig} = (1, \tilde{z}_{ij}')'$ and assuming that $(\alpha_{g,2}(u), \dots, \alpha_{g,d_z}(u))' = \gamma(u)$ for some non-stochastic $(d_z - 1)$-vector $\gamma(u)$ and all $g = 1, \dots, G$ in the model (2)–(3) gives the model (1) after substituting (3) into (2). The model (2)–(3) is more general, however, because it allows all coefficients of individual-level covariates to vary across groups via group-specific effects $\alpha_g(u)$, and it also allows to study not only location shift effects of the group-level covariates x_g but also their interaction effects. Therefore, throughout this paper, we study the model (2)–(3).

As an example of where the above modeling framework is useful, consider a case in which a researcher wishes to model the effects of a policy, contained in x_g, which varies at the state-by-year level (a "group" in this setting) on the distribution of micro-level outcomes (such as individuals' wages within each state-by-year combination), denoted y_{ig}, conditional on micro-level covariates, such as education level, denoted z_{ig}. The framework in (1) would model the location-shift effect of the policy on conditional quantiles of wages within a group, given by $\beta(u)$. The additional flexibility of (2)–(3) would also allow for interaction effects. For example, a policy x_g may have differential effects on lower wage quantiles for the less-educated than for the higher-educated; model (2) would capture this idea by allowing the researcher to specify a linear regression model of the form of (3) for the component of α_g that is the coefficient on education level, allowing the researcher to study how the effect of education level on the wage distribution varies as a function of x_g, the policy.[5]

In many applications, it is likely that the group-level covariates x_g may be endogenous in the sense that $E[x_g \varepsilon_g(u)] \neq 0$, at least for some values of the quantile index $u \in \mathcal{U}$. Therefore, to increase applicability of our results, we assume that there exists a d_w-vector of observable instruments w_g such that

[5]If the researcher is interested in modeling several effects, for example location-shift and some interaction effects, she can specify a linear regression model of the form (3) for each effect.

$E[w_g \varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$, $E[w_g x_g']$ is nonsingular, and y_{ig} is independent of w_g conditional on (z_{ig}, x_g, α_g).[6] The first two conditions are familiar from the classical linear instrumental variable regression analysis, and the third condition requires the distribution of y_{ig} to be independent of w_g once we control for z_{ig}, x_g, and α_g. It implies, in particular, that $Q_{y_{ig}|z_{ig},x_g,\alpha_g,w_g}(u) = z_{ig}'\alpha_g(u)$ for all $u \in \mathcal{U}$.[7]

We assume that a researcher has data on G groups and N_g individuals within group $g = 1, \dots, G$. Thus, the data consist of observations on $\{(z_{ig}, y_{ig}), i = 1, \dots, N_g\}$, x_g, and w_g for $g = 1, \dots, G$. Throughout the paper, we denote $N_G = \min_{1 \le g \le G} N_g$. For our asymptotic theory in Section 5, we will assume that N_G gets large as $G \to \infty$. Specifically, for the asymptotic zero-mean normality of our estimator $\hat{\beta}(u)$ of $\beta(u)$, we will assume that $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$; see Assumption 3 below. Thus, our results are useful when both G and N_G are large, which occurs in many empirical applications, but we also note that our results apply even if the number of observations per group is relatively small in comparison with the number of groups.

We also emphasize that, like in the original panel data *mean* regression model of Hausman and Taylor (1981), an important feature of our panel data *quantile* regression model is that it allows for *internal* instruments. Specifically, if some component of the vector z_{ig}, say $z_{ig,k}$, is exogenous in the sense that $E[z_{ig,k}\varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$, we can use, for example, $N_g^{-1/2}\sum_{i=1}^{N_g} z_{ig,k}$ as an additional instrument provided it is correlated with x_g, including it into the vector w_g. Since in practice it is often difficult to find an appropriate external instrument, allowing for internal instruments greatly increases the applicability of our results.

[6]The assumption that $E[w_g\varepsilon_g(u)] = 0$ holds jointly for all $u \in \mathcal{U}$ should not be confused with requiring quantile crossing. To understand it, assume, for example, that $\varepsilon_g(u) = \varepsilon(u, \eta_g)$ where η_g is a vector of group-level omitted variables in regression (3). Then a sufficient condition for the assumption $E[w_g\varepsilon_g(u)] = E[w_g\varepsilon(u, \eta_g)] = 0$ is that $E[\varepsilon(u, \eta_g)|w_g] = 0$. In turn, the restriction of the condition $E[\varepsilon(u, \eta_g)|w_g] = 0$ is that $E[\varepsilon(u, \eta_g)|w_g]$ does not depend on w_g, which occurs (for example) if η_g is independent of w_g. Once we assume that $E[\varepsilon(u, \eta_g)|w_g]$ does not depend on w_g, the further restriction that $E[\varepsilon(u, \eta_g)|w_g] = 0$ is a normalization of the component of the vector $\beta(u)$ corresponding to the constant in the vector x_g.

[7]The setting we model differs from other IV quantile settings, such as Chernozhukov and Hansen (2005, 2006, 2008). Consider, for simplicity, our model (1) and assume that $\mathcal{U} = [0, 1]$. Then the Skorohod representation implies that $y_{ig} = \tilde{z}_{ig}'\gamma(u_{ig}) + x_g'\beta(u_{ig}) + \varepsilon_g(u_{ig})$ where u_{ig} is a random variable that is distributed uniformly on $[0, 1]$ and is independent of $(\tilde{z}_{ig}, x_g, \varepsilon_g)$. Here, one can think of u_{ig} as unobserved individual-level heterogeneity. In this model, the unobserved group-level component $\varepsilon_g(\cdot)$ is modeled as an additively separable term. In contrast, the model in Chernozhukov and Hansen (2005, 2006, 2008) assumes that $\varepsilon_g(u) = 0$ for all $u \in [0, 1]$ and instead assumes that u_{ig} is not independent of $(\tilde{z}_{ig}, x_g)$. Thus, these two models are different and require different analysis.

Our problem in this paper is different from that studied in Koenker (2004), Kato, Galvao, and Montes-Rojas (2012), and Kato and Galvao (2011).[8] Specifically, they considered the panel data quantile regression model

$$(4) \qquad Q_{y_{ig}|z_{ig},\alpha_g}(u) = z'_{ig}\gamma(u) + \alpha_g(u), \quad u \in \mathcal{U},$$

and developed estimators of $\gamma(u)$. Building on Koenker (2004), Kato, Galvao, and Montes-Rojas (2012) suggested estimating $\gamma(u)$ in this model by running a quantile regression estimator of Koenker and Bassett (1978) on the pooled data, treating $\{\alpha_g(u), g = 1, \dots, G\}$ as a set of parameters to be estimated jointly with the vector of parameters $\gamma(u)$ (the same technique can be used to estimate $\gamma(u)$ in our model (1) by setting $\alpha_g(u) = x'_g\beta(u) + \varepsilon_g(u)$). They showed that their estimator is asymptotically zero-mean normal if $G^2(\log G)^3/N_G \to 0$ as $G \to \infty$. Making further progress, Kato and Galvao (2011) suggested an interesting smoothed quantile regression estimator of $\gamma(u)$ that is asymptotically zero-mean normal if $G/N_G \to 0$.[9] These papers do not provide a model for our estimator of $\beta(u)$, our primary object of interest, but instead focus solely on $\gamma(u)$.

Our model is also different from that studied in Hahn and Meinecke (2005), who considered an extension of Hausman and Taylor (1981) to cover nonlinear panel data models. Formally, they considered a nonlinear panel data model defined by the following equation:

$$E\big[\varphi\big(y_{ig}, z'_{ig}\gamma + x'_g\beta + \varepsilon_g\big)\big] = 0,$$

[8]Our paper is also related to but different from Graham and Powell (2012), who studied the model that in our notation would take the form $y_{ig} = z'_{ig}\alpha_g(u_{ig})$ where u_{ig} represents (potentially multidimensional) random unobserved heterogeneity, and developed an interesting identification and estimation strategy for the parameter $E[\alpha_g(u_{ig})]$, achieving identification when the number of observations per group remains small as the number of groups gets large and, under certain conditions, allowing $\alpha_g(\cdot) = \alpha_{ig}(\cdot)$ to depend on i.

[9]To clarify the difference between the growth condition in our paper, which is $G^{2/3}(\log N_G)/N_G \to 0$, and the growth condition, for example, in Kato, Galvao, and Montes-Rojas (2012), which is $G^2(\log G)^3/N_G \to 0$, assume, for simplicity, that $d_x = 1$, $d_z = 2$, and x_g and the second component of z_{ig} are constants, that is, $x_g = 1$ and $z_{ig} = (\tilde{z}'_{ig}, 1)'$. Then our model (2)–(3) reduces to $Q_{y_{ig}|\tilde{z}_{ig},\varepsilon_g,\alpha_g}(u) = \tilde{z}_{ig}(\beta(u) + \varepsilon_g(u)) + \alpha_g(u)$, which is similar to the model (4) studied in Kato, Galvao, and Montes-Rojas (2012) with the exception that we allow for additional group-specific random shifter $\varepsilon_g(u)$. When $\varepsilon_g(u)$ is present, our estimator $\hat{\beta}(u)$ of $\beta(u)$ satisfies $G^{1/2}(\hat{\beta}(u) - \beta(u)) \Rightarrow N(0, V_1)$ for some non-vanishing variance V_1; see Section 5. When $\varepsilon_g(u)$ is set to zero, however, V_1 vanishes, making the limiting distribution degenerate and leading to faster convergence rate of the estimator $\hat{\beta}(u)$. In fact, when V_1 vanishes, one obtains $(GN_G)^{1/2}(\hat{\beta}(u) - \beta(u)) \Rightarrow N(0, V_2)$ for some non-vanishing variance V_2. An additional $N_G^{1/2}$ factor, in turn, appears in the residual terms of the Bahadur representation of the estimator $\hat{\beta}(u)$, which eventually lead to stronger requirements on the growth of the number of observations per group N_G relative to the number of groups, explaining the difference between the growth condition in Kato, Galvao, and Montes-Rojas (2012) and our growth condition.

where $\varphi(\cdot,\cdot)$ is a vector of moment functions and $x_g'\beta+\varepsilon_g$ is the group-specific effect. As in this paper, the authors were interested in estimating the effect of group-level covariates (coefficient β) without assuming that ε_g is independent (or mean-independent) of x_g but assuming instead that there exists an instrument w_g satisfying $E[w_g\varepsilon_g]=0$. Importantly, however, they assumed that $\varphi(\cdot,\cdot)$ is a vector of *smooth* functions, so that their results do not apply immediately to our model. In addition, Hahn and Meinecke (2005) required that $N_G/G>c$ for some $c>0$ uniformly over all G to prove that their estimator is asymptotically zero-mean normal. In contrast, as emphasized above, we only require that $G^{2/3}(\log N_G)/N_G\to 0$ as $G\to\infty$, with the improvement coming from a better control of the residuals in the Bahadur representation.

3. ESTIMATOR

In this section, we develop our estimator, which we refer as grouped IV quantile regression. Our main emphasis is to derive a computationally simple, yet consistent, estimator. The estimator consists of the following two stages.

Stage 1: For each group g and each quantile index u from the set $\mathcal{U}$ of indices of interest, estimate uth quantile regression of y_{ig} on z_{ig} using the data $\{(y_{ig},z_{ig}):i=1,\dots,N_g\}$ by the classical quantile regression estimator of Koenker and Bassett (1978):

$$\hat\alpha_g(u)=\arg\min_{a\in\mathbb{R}^{d_z}}\sum_{i=1}^{N_g}\rho_u\bigl(y_{ig}-z_{ig}'a\bigr),$$

where $\rho_u(x)=(u-1\{x<0\})x$ for $x\in\mathbb{R}$. Denote $\hat\alpha_g(u)=(\hat\alpha_{g,1}(u),\dots,\hat\alpha_{g,d_z})'$.

Stage 2: Estimate a 2SLS regression of $\hat\alpha_{g,1}(u)$ on x_g using w_g as an instrument to get an estimator $\hat\beta(u)$ of $\beta(u)$, that is,

$$\hat\beta(u)=\bigl(X'P_WX\bigr)^{-1}\bigl(X'P_W\hat A(u)\bigr),$$

where $X=(x_1,\dots,x_G)'$, $W=(w_1,\dots,w_G)'$, $\hat A(u)=(\hat\alpha_{1,1}(u),\dots,\hat\alpha_{G,1}(u))'$, and $P_W=W(W'W)^{-1}W'$.[10]

Intuitively, as the number of observations per group increases, $\hat\alpha_{g,1}-\alpha_{g,1}$ shrinks to zero uniformly over $g=1,\dots,G$, and we obtain a classical instrumental variables problem. The theory presented below provides a mild condition on the growth of the number of observations per group that is sufficient to achieve consistency and asymptotic zero-mean normality of $\hat\beta(u)$.

[10]The use of a 2SLS regression on the second stage of our estimator is dictated by our assumption that $\varepsilon_g(u)$ is (mean)-uncorrelated with w_g: $E[w_g\varepsilon_g(u)]=0$. If, instead, we assumed that $\varepsilon_g(u)$ is median-uncorrelated with w_g, a concept developed in Komarova, Severini, and Tamer (2012), the second stage of our estimator would be an IV quantile regression developed in Chernozhukov and Hansen (2006). In this case, our method would be a quantile-after-quantile estimator.

Several special cases of our estimator are worth noting. First, when the model is given by equation (1), the steps of our estimator consist of (i) group-by-group quantile regression of y_{ig} on $\tilde{z}_{ig}$ and on a constant, saving the estimated coefficient $\hat{\alpha}_{g,1}(u)$ corresponding to the constant, $\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u)$, in each group; and (ii) regressing those saved coefficients $\hat{\alpha}_{g,1}(u)$ on x_g via 2SLS using w_g as instruments. Second, if z_{ig} contains *only* a constant, the first stage simplifies to selecting the uth quantile of the outcome variable y_{ig} within each group. Third, if x_g is exogenous, that is, $E[x_g\varepsilon_g(u)] = 0$, OLS of $\hat{\alpha}_{g,1}(u)$ on x_g may be used rather than 2SLS in the second stage. In this latter case, the grouped quantile estimation approach provides the advantage of handling group-level unobservables (or, alternatively, left-hand-side measurement error), which would bias the traditional Koenker and Bassett (1978) estimator. When z_{ig} only includes a constant and x_g is exogenous, the grouped IV quantile regression estimator $\hat{\beta}(u)$ simplifies to the minimum distance estimator described in Chamberlain (1994) (see also Angrist, Chernozhukov, and Fernandez-Val (2006)).

This estimator has several computational benefits relative to alternative methods. First, note that when the model is given by equation (1), another approach to perform the first stage of our estimator would be to denote $\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u)$ and estimate parameters $\gamma(u)$ and $\{\alpha_{g,1}(u), g = 1, \dots, G\}$ jointly from the pooled data set as in Kato, Galvao, and Montes-Rojas (2012). This would provide an efficiency gain given that in this case, individual-level effects $\gamma(u)$ are group-independent. Although the method we use is less efficient, it is computationally much less demanding since only few parameters are estimated in each regression, which can greatly reduce computation times in large data sets with many fixed effects.[11] Second, even if no group-level unobservables exist (consider model (1) with $\varepsilon_g(u) = 0$ for all $g = 1, \dots, G$), the grouped estimation approach can be considerably faster than the traditional Koenker and Bassett (1978) estimator (though both estimators will be consistent). This computational advantage occurs when the dimension of x_g is large: standard quantile regression estimates $\beta(u)$ in a single, nonlinear step, whereas the grouped quantile approach estimates $\beta(u)$ in a linear second stage.[12]

Monte Carlo simulations in Appendix A of the Supplemental Material highlight the performance of our estimator for $\beta(u)$ in (1) relative to the traditional Koenker and Bassett (1978) estimator (which ignores endogeneity of x_g as well

[11]In Monte Carlo experiments in Appendix A of the Supplemental Material, we find that jointly estimating group-level effects can take over 150 times as long as the grouped quantile approach when $G = 200$. With $G > 200$, the computation time ratio drastically increases further, with standard optimization packages often failing to converge appropriately.

[12]One such example would be a case where a group is a state-by-year combination, and x_g contains many state and year fixed effects, in addition to the treatment of interest, as in Example 2 of Section 4.

as the existence of $\varepsilon_g(u)$). Even when N_G and G are both small, the grouped IV quantile approach has lower bias than traditional quantile regression when x_g is endogenous. When x_g is exogenous but group-level unobservables $\varepsilon_g(u)$ are still present, the bias of the grouped quantile approach shrinks quickly to zero as N_G grows but the bias of traditional quantile estimator does not. When no group-level unobservables are present, and hence both the grouped estimation approach and traditional quantile regression should be consistent, our estimator still has small bias, although traditional quantile regression outperforms our method in this case.

As we demonstrate below, standard errors for our estimator $\hat{\beta}(u)$ may be obtained using standard heteroscedasticity-robust (Section 5) or clustering (Appendix E of the Supplemental Material) approaches for 2SLS or OLS as if there were no first stage. Note that clustering in the second stage refers to dependence *across* groups, not within groups. For example, if a group is a state-by-year combination, the researcher may wish to use standard errors which are clustered at the state level.

4. EXAMPLES OF GROUPED IV QUANTILE REGRESSION

To help the reader envision applications of our estimator, in this section, we provide several motivating examples of settings for which our estimator may be useful. Each of the following examples involves estimation of a treatment effect that varies at the group level with all endogeneity concerns also existing only at the group level.[13]

EXAMPLE 1—Peer Effects of School Integration: Angrist and Lang (2004) studied how suburban student test scores were affected by the reassignment of participating urban students to suburban schools through Boston's Metco program. Before estimating their main instrumental variables model, the authors tested for a relationship between the presence of urban students in the classroom and the second decile of student test scores by estimating

$$Q_{y_{igjt}|x_{gjt}}(0.2) = \alpha_g(0.2) + \beta_j(0.2) + \gamma_t(0.2) + \delta(0.2)m_{gjt} + \lambda(0.2)s_{gjt} + \xi_{gjt}(0.2), \tag{5}$$

where the left-hand side represents the second decile of student test scores within a group, $x_{gjt} = (m_{gjt}, s_{gjt}, \xi_{gjt}, \alpha_g, \beta_j, \gamma_t)$, and a group is a grade $g \times$ school $j \times$ year t cell. The variables s_{gjt} and m_{gjt} denote the class size and the

[13]This is in contrast to settings where the endogeneity exists at the individual level, that is, when the individual unobserved heterogeneity is correlated with treatment. Such situations require a different approach than the one presented here, for example, Chernozhukov and Hansen (2005), Abadie, Angrist, and Imbens (2002), or the other approaches referenced in Section 1.

fraction of Metco students within each $g \times j \times t$ cell, and α_g, β_j, and γ_t represent grade, school, and year effects, respectively. The unobserved component ξ_{gjt} is analogous to $\varepsilon_g(0.2)$ in our model (1).

Angrist and Lang (2004) estimated equation (5) by OLS, which is equivalent to the non-IV application of our estimator with no micro-level covariates. Similarly to their OLS results on average test scores, they found that classrooms with higher proportions of urban students have lower second decile test scores. Once they instrumented for a classroom's level of Metco exposure, the authors found no effect on *average* test scores. However, by not estimating model (5) by 2SLS, they were unable to address the causal *distributional* effects of Metco exposure.

In estimating (5), Angrist and Lang (2004) used heteroscedasticity-robust standard errors, which we demonstrate in Section 5 is valid. The extension in Appendix E of the Supplemental Material implies that the authors could have instead allowed for clustering across groups in computing standard errors (e.g., clustering at the school level given a sufficient number of schools).

EXAMPLE 2—Occupational Licensing and Quality: Larsen (2014) applied the estimator developed in this paper to study the effects of occupational licensing laws on the distribution of quality within the teaching profession. Similarly to Example 1, the explanatory variable of interest is treated as exogenous and the researcher is concerned that there may be unobserved group-level disturbances. In this application, a group is a state-year combination (s, t), and micro-level data consist of teachers within a particular state in a given year. The conditional uth quantile of teacher quality among teachers who began teaching in state s in year t is modeled as

$$Q_{q_{ist}|Law_{st},\varepsilon_{st}}(u) = \gamma_s(u) + \lambda_t(u) + Law'_{st}\delta(u) + \varepsilon_{st}(u), \tag{6}$$

where Law_{st} is a vector of dummies capturing the type of certification tests required for licensure in state s and year t, $\gamma_s(u)$ and $\lambda_t(u)$ are state and year effects, and $\varepsilon_{st}(u)$ represents group-level unobservables.

Because no micro-level covariates are included, the first stage of the grouped quantile estimator is obtained by simply selecting the uth quantile of quality in a given state-year cell. The second stage is obtained via OLS. Larsen (2014) found that, for first-year teachers, occupational licensing laws requiring teachers to pass a subject test lead to a small but significant decrease in the upper tail of quality, suggestive that these laws may drive some highly qualified candidates from the occupation.

In this setting, if micro-level covariates, z_{ist}, were included in the first stage of estimation, the researcher could also estimate *interaction* effects of the group-level treatment and a micro-level covariate, such as the percent of minority students at the teacher's school. This would be done by (i) estimating quantile regression of q_{ist} on a vector z_{ist} (which would include a measure of the percent minority students) separately for each (s, t) group and saving each group-level estimate for the coefficient corresponding to the percent minority variable; and (ii) estimating a linear regression of these coefficients on Law_{st} and on the state and year fixed effects.

This example highlights another useful feature of grouped IV quantile regression. Including many variables in a standard quantile regression can drastically increase the computational time (see Koenker (2004), Lamarche (2010), Galvao and Wang (2013), and Galvao (2011) for further discussion) and, in our experience, can often lead standard optimization packages to fail to converge. The grouped quantile approach, on the other hand, can handle large numbers of variables easily when these variables happen to be constant within group, as in the case of state and year fixed effects in this example, because the coefficients corresponding to these variables can be estimated in the second-stage linear model, greatly reducing the number of parameters to be estimated in the nonlinear first stage and hence reducing the computational burden significantly.[14]

EXAMPLE 3—Distributional Effects of Suburbanization: Palmer (2011) applied the grouped quantile estimator to study the effects of suburbanization on resident outcomes. This application illustrates the use of our estimator in an IV setting. In this application, a group is a metropolitan statistical area (MSA), and individuals are MSA residents. As an identification strategy, Palmer (2011) used the results of Baum-Snow (2007) in instrumenting suburbanization with planned highways.[15]

The model is

$$\Delta Q_{y_{igt}|x_g,s_g,\varepsilon_g}(u) = \beta(u) \cdot \textit{suburbanization}_g + x_g'\gamma_1(u) + \varepsilon_g(u),$$
$$\textit{suburbanization}_g = \pi(u) \cdot \textit{planned highway rays}_g + x_g'\gamma_2(u) + v_g(u),$$

where $\Delta Q_{y|x_g,s_g,\varepsilon_g}(u)$ is the change in the uth quantile of log wages y_{igt} within an MSA between 1950 and 1990 and x_g is a vector of controls (including a constant) conditional upon which *planned highway rays*$_g$ is uncorrelated with $\varepsilon_g(u)$ and $v_g(u)$. The variable *suburbanization*$_g$ is a proxy measure of population decentralization, such as the amount of decline of central-city population density. $\beta(u)$ is the coefficient of interest, capturing the effect of suburbanization on

[14]Note also that this specific computational advantage of the grouped quantile regression estimator exists even in cases where both standard quantile regression and the grouped approach are valid (i.e., when no group-level unobservables are present). Larsen (2014) found that estimating (6) using the grouped approach was significantly faster than estimating (6) in a single standard quantile regression. See also Appendix A of the Supplemental Material for further discussion of computational advantages of the grouped quantile approach.

[15]Baum-Snow (2007) instrumented for actual constructed highways with planned highways and estimated that each highway ray emanating out of a city caused an 18% decline in central-city population.

the within-MSA conditional wage distribution. For example, if the process of suburbanization had particularly acute effects on the prospects of low-wage workers, we may expect $\beta(u)$ to be negative for $u = 0.1$. For a given u, the grouped IV quantile approach estimates $\beta(u)$ through a 2SLS regression.

EXAMPLE 4—The Relationship Between Productivity and Competition: Backus (2014) studied the relationship between competition and productivity in the ready-mix concrete industry. The author discussed the fact that competition and productivity are positively correlated, and studied whether this relationship is similar for firms of all productivity levels (e.g., through encouraging better monitoring of firm managers or better investments), or whether increased competition primarily affects the lower tail of the productivity distribution (driving out less productive firms).

Let ρ_{imt} represent a measure of productivity of firm i in market m and time period t. Using our notation, define a group as a pair $m \times t$. The author assumes that ρ_{imt} satisfies the following quantile regression model:

$$Q_{\rho_{imt}|c_{mt},n_{mt},\varepsilon_{mt}}(u) = \beta_t(u) + c_{mt}\beta_c(u) + g(n_{mt}, u) + \varepsilon_{mt}(u), \tag{7}$$

where c_{mt} is a group-level measure of competition, n_{mt} is the number of firms in the group, $g(n_{mt}, u)$ is the third-order polynomial of n_{mt}, and ε_{mt} is an unobserved group-level disturbance, which is possibly correlated with c_{mt}.

Backus (2014) instrumented for c_{mt} using group-level measures which shift the demand for concrete. Thus, the IV regression in (7) represents an application of our estimator when group-level shocks are endogenous and no micro-level covariates are present. The author found some evidence that the effect of competition on the left tail of the productivity distribution may be more positive than at some quantiles in the middle of the distribution (consistent with selection of low-productivity firms out of the industry), but was unable to reject the hypothesis of a constant effect. Backus (2014) reported standard errors clustered at the market level, which we demonstrate are valid in Appendix E of the Supplemental Material.

5. ASYMPTOTIC THEORY

In this section, we formulate our assumptions and present our main theoretical results.

5.1. *Assumptions*

Let c_M, c_f, C_M, C_f, C_L be strictly positive constants whose values are fixed throughout the paper. Recall that $N_G = \min_{g=1,\dots,G} N_g$. We start with specifying our main assumptions.

ASSUMPTION 1—Design: (i) *Observations are independent across groups*. (ii) *For all* $g = 1, \dots, G$, *the pairs* (z_{ig}, y_{ig}) *are i.i.d. across* $i = 1, \dots, N_g$ *conditional on* (x_g, α_g).

ASSUMPTION 2—Instruments: (i) *For all* $u \in \mathcal{U}$ *and* $g = 1, \dots, G$, $E[w_g \varepsilon_g(u)] = 0$. (ii) *As* $G \to \infty$, $G^{-1} \sum_{g=1}^{G} E[x_g w_g'] \to Q_{xw}$ *and* $G^{-1} \times \sum_{g=1}^{G} E[w_g w_g'] \to Q_{ww}$ *where* Q_{xw} *and* Q_{ww} *are matrices with singular values bounded in absolute value from below by* c_M *and from above by* C_M. (iii) *For all* $g = 1, \dots, G$ *and* $i = 1, \dots, N_g$, y_{ig} *is independent of* w_g *conditional on* (z_{ig}, x_g, α_g). (iv) *For all* $g = 1, \dots, G$, $E[\|w_g\|^{4+c_M}] \le C_M$.

ASSUMPTION 3—Growth Condition: *As* $G \to \infty$, *we have* $G^{2/3}(\log N_G)/N_G \to 0$.

Assumption 1(i) holds, for example, if groups are sampled randomly from some population of groups. This assumption precludes the possibility of clustering across groups (e.g., if a group is a state-by-year combination, there may be clustering on the state level). Since clustered standard errors are important in practice, however, we derive an extension of our results relaxing the independence across groups condition and allowing for clustering in Appendix E of the Supplemental Material. Assumption 1(ii) allows for interdependence (clustering) within groups but imposes the restriction that the interdependence between observations within the group g is fully controlled for by the group-level covariates x_g and the group-specific effect α_g. Assumption 2 is our main identification condition. Note that Assumption 2 allows for internal instruments. In particular, if $w_g = N_g^{-1/2} \sum_{i=1}^{N_g} z_{ig,k}$ for some k, then Assumption 2(iii) automatically follows from Assumption 1(ii). Assumption 3 implies that the number of observations per group grows sufficiently fast as G gets large, and gives a particular growth rate that suffices for our results. Note that our growth condition is rather weak and, most importantly, allows for the case when the number of observations per group is small relative to the number of groups.[16]

Next, we specify technical conditions that are required for our analysis. Let $E_g[\cdot] = E[\cdot | x_g, \alpha_g]$, and let $f_g(\cdot)$ denote the conditional density function of y_{1g} given (z_{1g}, x_g, α_g) (dependence of $f_g(\cdot)$ on (z_{1g}, x_g, α_g) is not shown explicitly for brevity of notation). Also denote $B_g(u, c) = (z_{1g}'\alpha_g(u) - c, z_{1g}'\alpha_g(u) + c)$ for $c > 0$. We will assume the following regularity conditions:

ASSUMPTION 4—Covariates: (i) *For all* $g = 1, \dots, G$ *and* $i = 1, \dots, N_g$, *random vectors* z_{ig} *and* x_g *satisfy* $\|z_{ig}\| \le C_M$ *and* $\|x_g\| \le C_M$. (ii) *For all* $g = 1, \dots, G$, *all eigenvalues of* $E_g[z_{1g} z_{1g}']$ *are bounded from below by* c_M.

[16]Using the more common notation of panel data models, where N is the number of individuals (groups) and T is the number of time periods (individuals within the group), Assumption 3 would take the form: $N^{2/3}(\log T)/T \to 0$ as $N \to \infty$.

ASSUMPTION 5—Coefficients: *For all* $u_1, u_2 \in \mathcal{U}$ *and* $g = 1, \dots, G$, $\|\alpha_g(u_2) - \alpha_g(u_1)\| \leq C_L|u_2 - u_1|$.

ASSUMPTION 6—Noise: (i) *For all* $g = 1, \dots, G$, $E[\sup_{u\in\mathcal{U}} |\varepsilon_g(u)|^{4+c_M}] \leq C_M$. (ii) *For some (matrix-valued) function* $J : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^{d_w\times d_w}$, $G^{-1} \times \sum_{g=1}^{G} E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_g w_g'] \to J(u_1, u_2)$ *uniformly over* $u_1, u_2 \in \mathcal{U}$. (iii) *For all* $u_1, u_2 \in \mathcal{U}$, $|\varepsilon_g(u_2) - \varepsilon_g(u_1)| \leq C_L|u_2 - u_1|$.

ASSUMPTION 7—Density: (i) *For all* $u \in \mathcal{U}$ *and* $g = 1, \dots, G$, *the conditional density function* $f_g(\cdot)$ *is continuously differentiable on* $B_g(u, c_f)$ *with the derivative* $f_g'(\cdot)$ *satisfying* $|f_g'(y)| \leq C_f$ *for all* $y \in B_g(u, c_f)$ *and* $|f_g'(z_{1g}'\alpha_g(u))| \geq c_f$. (ii) *For all* $u \in \mathcal{U}$ *and* $g = 1, \dots, G$, $f_g(y) \leq C_f$ *for all* $y \in B_g(u, c_f)$ *and* $f_g(z_{1g}'\alpha_g(u)) \geq c_f$.

ASSUMPTION 8—Quantile Indices: *The set of quantile indices* $\mathcal{U}$ *is a compact set included in* $(0, 1)$.

Assumption 4(i) requires that both individual- and group-level observable covariates z_{ig} and x_g are bounded. Assumption 4(ii) is a familiar identification condition in regression analysis. Assumption 5 is a mild continuity condition. Assumption 6(i) requires sufficient integrability of the noise $\varepsilon_g(u)$, which is a mild regularity condition. In fact, under Assumption 6(iii), which is also a mild continuity condition, Assumption 6(i) is satisfied as long as $E[|\varepsilon_g(u)|^{4+c_M}] \leq C_M$ for some $u \in \mathcal{U}$ (with a possibly different constant C_M). Assumption 6(ii) is trivially satisfied if the pairs (w_g, ε_g) are i.i.d. across g. Assumption 7 is a mild regularity condition that is typically imposed in the quantile regression analysis. Finally, Assumption 8 excludes quantile indices that are too close to either 0 or 1 (when the quantile index u is close to either 0 or 1, one obtains a so-called extremal quantile model, which requires a rather different analysis; see, e.g., Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011)).

5.2. *Results*

We now present our main results. In Theorem 1, we derive the asymptotic distribution of our estimator. In Theorem 2, we show how to estimate the asymptotic covariance of our estimator. For brevity of the paper, further results are relegated to Appendices C–E of the Supplemental Material. In particular, in Appendix C, we describe a multiplier bootstrap method for constructing uniform over $u \in \mathcal{U}$ confidence intervals for $\beta(u)$ and prove its validity relying on results from Chernozhukov, Chetverikov, and Kato (2013). In Appendix D, we present an approach for uniform inference on $\{\alpha_{g,1}(u), g = 1, \dots, G\}$ in the model (2)–(3) by constructing the confidence bands $[\hat{\alpha}^l_{g,1}(u), \hat{\alpha}^r_{g,1}(u)]$ that cover the true group-specific effects $\alpha_{g,1}(u)$ for all $g = 1, \dots, G$ simultaneously with probability approximately $1 - \alpha$. In Appendix E, we consider clustered standard errors.

The first theorem derives the asymptotic distribution of our estimator.

THEOREM 1—Asymptotic Distribution: *Let Assumptions 1–8 hold. Then*

$$\sqrt{G}\big(\hat{\beta}(\cdot) - \beta(\cdot)\big) \Rightarrow \mathbb{G}(\cdot), \quad \textit{in} \quad \ell^\infty(\mathcal{U}),$$

where $\mathbb{G}(\cdot)$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $\mathcal{C}(u_1, u_2) = SJ(u_1, u_2)S'$, where $S = (Q_{xw}Q_{ww}^{-1}Q'_{xw})^{-1}Q_{xw}Q_{ww}^{-1}$, Q_{xw} and Q_{ww} appear in Assumption 2, and $J(u_1, u_2)$ in Assumption 6.

REMARK 1: (i) This is our main convergence result that establishes the asymptotic behavior of our estimator. Note that we provide the *joint* asymptotic distribution of our estimator for all $u \in \mathcal{U}$. In addition, Theorem 1 implies that, for any $u \in \mathcal{U}$,

$$\sqrt{G}\big(\hat{\beta}(u) - \beta(u)\big) \Rightarrow N(0, V),$$

where $V = SJ(u, u)S'$, which is the asymptotic distribution of the classical 2SLS estimator.

(ii) In order to establish the joint asymptotic distribution of our estimator for all $u \in \mathcal{U}$, we have to deal with G independent quantile processes $\{\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u), u \in \mathcal{U}\}$. Since $G \to \infty$, classical functional central limit theorems do not apply. Therefore, we employ a nonstandard but powerful Bracketing by Gaussian Hypotheses Theorem; see Theorem 2.11.11 in Van der Vaart and Wellner (1996).

(iii) Since quantile regression estimators are biased in finite samples, our estimator $\hat{\alpha}_{g,1}(u)$ of $\alpha_{g,1}(u)$ does not necessarily satisfy $E[(\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u))w_g] = 0$. For this reason, our estimator $\hat{\beta}(u)$ of $\beta(u)$ is not consistent if N_g is bounded from above uniformly over $g = 1, \dots, G$ and $G \geq 2$. We note, however, that quantile estimators are *asymptotically* unbiased, and so we use the Bahadur representation of quantile estimators to derive weak condition on the growth of $N_G = \min_{1 \leq g \leq G} N_g$ relative to G, so that consistent estimation of $\beta(u)$ is indeed possible. Specifically, we prove consistency and asymptotic zero-mean normality under Assumption 3 that states that $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$, which is a mild growth condition. In principle, it is also possible to consider bias correction of the quantile regression estimators. This would further relax the growth condition on N_G relative to G at the expense of stronger side assumptions and more complicated estimation procedures.

(iv) The requirement that $N_G \to \infty$ as $G \to \infty$ is in contrast with the classical results of Hausman and Taylor (1981) on estimation of panel data mean regression model. The main difference is that the fixed effect estimator in the panel data mean regression model is unbiased even in finite samples leading to consistent estimators of the effects of group-level covariates with the number of observations per group being fixed.

The result in Theorem 1 derives asymptotic behavior of our estimator. In order to perform inference, we also need an estimator of the asymptotic covariance function. We suggest using an estimator $\hat{\mathcal{C}}(\cdot,\cdot)$ that is defined for all $u_1, u_2 \in \mathcal{U}$ as

$$\hat{\mathcal{C}}(u_1, u_2) = \hat{S}\hat{J}(u_1, u_2)\hat{S}',$$

where

$$\hat{J}(u_1, u_2) = \frac{1}{G}\sum_{g=1}^{G}\big(\big(\hat{\alpha}_{g,1}(u_1) - x_g'\hat{\beta}(u_1)\big)\big(\hat{\alpha}_{g,1}(u_2) - x_g'\hat{\beta}(u_2)\big)w_g w_g'\big),$$

$\hat{S} = (\hat{Q}_{xw}\hat{Q}_{ww}^{-1}\hat{Q}_{xw}')^{-1}\hat{Q}_{xw}\hat{Q}_{ww}^{-1}$, $\hat{Q}_{xw} = X'W/G$, and $\hat{Q}_{ww} = W'W/G$. In the theorem below, we show that $\hat{\mathcal{C}}(u_1, u_2)$ is consistent for $\mathcal{C}(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$.

THEOREM 2—Estimating $\mathcal{C}$: *Let Assumptions 1–8 hold. Then* $\|\hat{\mathcal{C}}(u_1, u_2) - \mathcal{C}(u_1, u_2)\| = o_p(1)$ *uniformly over* $u_1, u_2 \in \mathcal{U}$.

REMARK 2: Theorems 1 and 2 can be used for hypothesis testing concerning $\beta(u)$ for a given quantile index $u \in \mathcal{U}$. In particular, we have that

$$\sqrt{G}\hat{\mathcal{C}}(u, u)^{-1/2}\big(\hat{\beta}(u) - \beta(u)\big) \Rightarrow N(0, 1). \tag{8}$$

Importantly for applied researchers, Theorems 1 and 2 demonstrate that heteroscedasticity-robust standard errors for our estimator can be obtained by the traditional White (1980) standard errors where we proceed as if $\hat{\alpha}_{g,1}(u)$ were equal to $\alpha_{g,1}(u)$, that is, as if there were no first-stage estimation error. Traditional approaches to clustered standard errors are also valid in this setting; extending Theorems 1 and 2 to apply to settings with clustering is straightforward, but requires additional notation, and therefore we present these results in Appendix E of the Supplemental Material. As highlighted above, clustering in this context refers to clustering *across* groups. For example, if a group is state-by-year cell, the researcher could cluster at the state level.

6. THE EFFECT OF CHINESE IMPORT COMPETITION ON THE LOCAL WAGE DISTRIBUTION

6.1. *Background on Wage Inequality*

Over the past 40 years, wage inequality within the United States has increased drastically.[17] Economists have engaged in heated debates about the primary causes of the rising wage inequality—such as globalization, skill-biased technological change, or the declining real minimum wage—and how the importance of these factors has changed over the years.[18] Recent work in Autor, Dorn, and Hanson (2013) (hereafter ADH) focused on import competition and its effects on wages and employment in U.S. local labor markets. ADH studied the period 1990–2007, when the share of U.S. spending on Chinese imports increased dramatically from 0.6% to 4.6%. For identification, the authors used spatial variation in manufacturing concentration, showing that localized U.S. labor markets that specialize in manufacturing were more affected by increased import competition from China. The authors found that those markets which were more exposed to increased import competition in turn had lower employment and lower wages.

We contribute to this debate by studying the effect of increased trade, in the form of increased import competition, on the distribution of local wages (rather than on the average local wages as in ADH). Given that we exploit the same variation in import competition as in ADH, we first describe the ADH framework below and then present our results.

6.2. *Framework of Autor, Dorn, and Hanson (2013)*

To study the effect of Chinese import competition on average domestic wages, ADH used Census microdata to calculate the mean wage within each Commuting Zone (CZ) in the United States.[19] The authors then estimated the following regression:

$$\Delta \overline{\ln w}_g = \beta_1 \Delta IPW^U_g + X'_g \beta_2 + \varepsilon_g, \tag{9}$$

where $\Delta \overline{\ln w}_g$ is the change in average individual log weekly wage in a given CZ in a given decade, X_g are characteristics of the CZ and decade, including indicator variables for each decade. Note that we have changed the notation

[17] Autor, Katz, and Kearney (2008) documented that, from 1963 to 2005, the change in wages for the 90th percentile earner was 55% higher than for the 10th percentile earner.

[18] See, for example, Leamer (1994), Krugman (2000), Feenstra and Hanson (1999), Katz and Autor (1999), as well as many other papers cited in Feenstra (2010) or in Haskel, Lawrence, Leamer, and Slaughter (2012).

[19] The United States is covered exhaustively by 722 Commuting Zones (Tolbert and Sizer (1996)), each roughly corresponding to a local labor market.

slightly from that in ADH in order to improve clarity for our application—a "group" g in this setting is a given CZ in a given decade. The variable of interest is ΔIPW_g^U, which represents the decadal change in Chinese imports per U.S. worker for the CZ and decade corresponding to group g.[20]

To address endogeneity concerns (i.e., that imports from China may be correlated with unobserved labor demand shocks), the authors instrumented for imports per last-period worker using ΔIPW_g^O, a measure of import exposure that replaces the change in Chinese imports to the United States in a given industry with the change in Chinese imports to other similarly developed nations for the same industry and uses one decade lagged employment shares in calculating the weighted average. Using this 2SLS approach, the authors found that a \$1,000 increase in Chinese imports per worker in a CZ decreases average log weekly wage by -0.76 log points, corresponding to decrease in wages for the average CZ of 0.9% from 1990 to 2000 and 1.4% from 2000 to 2007. When estimated separately by gender, the effect was more negative for males (-0.89 log points) and less so for females (-0.61 log points).[21]

6.3. *Distributional Effects of Increased Import Competition*

We build on the ADH framework to analyze whether low-wage earners were more adversely affected than high-wage earners by Chinese import competition. To apply the grouped IV quantile regression estimator to this setting, we replace $\Delta\overline{\ln w}_g$, the change in the average log weekly wage in equation (9), with $\Delta \ln w_g^u$, the change in the u-quantile of log wages in the CZ and decade corresponding to group g. We calculate these quantiles using micro-level observations from the Census Integrated Public Use Micro Samples for 1990 and 2000 and the American Community Survey for 2006–2008, matching these observations to CZs following the strategy described in ADH.[22] We instrument for ΔIPW_g^U using ΔIPW_g^O as described above. Recall that existing methods for handling endogeneity in quantile models are suited for the case where the

[20]ADH apportion national industry-level import changes to local imports per worker using the weighted average of industry-level changes in the value of Chinese imports to the United States, with weights corresponding to the beginning-of-decade employment share of each industry in each CZ.

[21]As discussed by ADH, the existence of an extensive-margin labor supply response—imports affecting whether individuals are employed—makes these results likely a lower bound for the effect on all workers because we do not observe wages for the unemployed population.

[22]The thought experiment behind the asymptotics in this application is that the estimator is consistent as the number of groups ($G = 722$ CZs $\times$ two decades) and the number of individuals within each group ($N_G = 543$, the size of the smallest group) both grow large. We follow ADH by clustering at the state level and weighting by start-of-decade CZ population in the second stage of our estimator. To cluster, we are relying on Appendix E of the Supplemental Material, which relaxes Assumption 1 to allow for observations to be dependent across groups. We also follow the ADH individual weighting procedure in the first stage given that not all individuals can be mapped to a unique CZ.

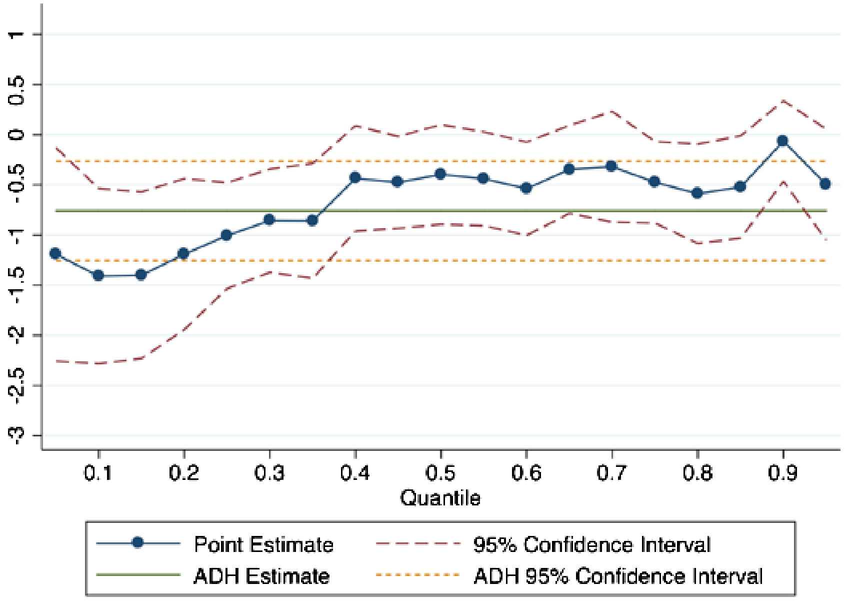

FIGURE 1.—Effect of Chinese import competition on conditional wage distribution: full sample. *Notes*: Figure plots grouped IV quantile regression estimates of the effect of a $1,000 increase in Chinese imports per worker on the conditional wage distribution (β_1 in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta\overline{\ln w_g}$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (9). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

individual-level unobserved conditional quantile itself is correlated with the treatment and would be inconsistent in this setting because the endogeneity consists of a group-level treatment being correlated with the group-level unobservable additive term.

Figures 1, 2, and 3 display the results of the grouped IV quantile regression estimator for the full sample, for females only, and for males only. Each figure displays u-quantile estimates for $u \in \{0.05, 0.1, \dots, 0.95\}$, along with pointwise 95% confidence bands about each estimate. The figures also display the 2SLS effect found in ADH and 95% confidence intervals corresponding to their IV estimate of Chinese import penetration on the change in CZ-level average wages.

Each figure provides evidence that Chinese import competition affected the wages of low-wage earners more than high-wage earners, demonstrating how increases in trade can causally exacerbate local income inequality. For all three samples, the magnitude of the estimated causal effect of Chinese import penetration is much larger for lower quantiles of the conditional wage distribution.

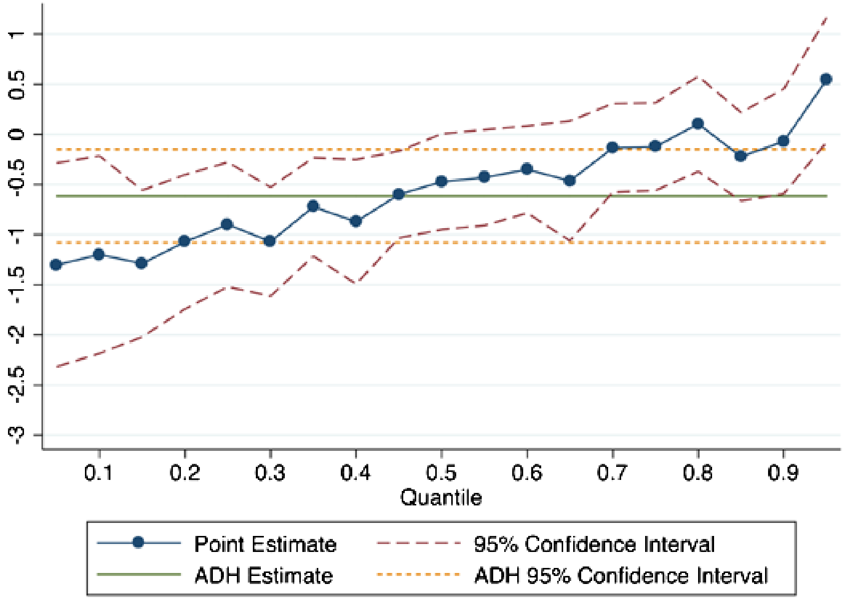

FIGURE 2.—Effect of Chinese import competition on conditional wage distribution: females only. *Notes*: Figure plots grouped IV quantile regression estimates for the female-only sample of the effect of a $1,000 increase in Chinese imports per worker on the female conditional wage distribution (β_1 in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta\overline{\ln w_g}$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (9). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

The point estimates suggest that the average negative effect of Chinese import penetration estimated by ADH is primarily driven by large negative effects for those in the bottom tercile, where the effect is twice as large as the average effect.[23] Wages not in the bottom tercile were less affected than the average—Figure 1 shows that, for most wage-earners (from the 0.35 quantile and above), the effect of Chinese import competition was one-third smaller in magnitude than the effect on the average estimated by ADH. Comparing the pattern of the coefficients across two gender subsamples in Figures 2 and 3, there is more distributional heterogeneity for females than males, a finding that additional testing shows is even more pronounced for non-college educated females. For each sample, we can reject an effect size of zero for almost all quantiles below the median but cannot for all quantiles above the median.

[23]A coefficient of −1.4 log points, for example for the lower quantiles of Figure 1, corresponds to a 2.6% decrease in wages from 2000 to 2007 for the average commuting zone's change in Chinese import exposure.

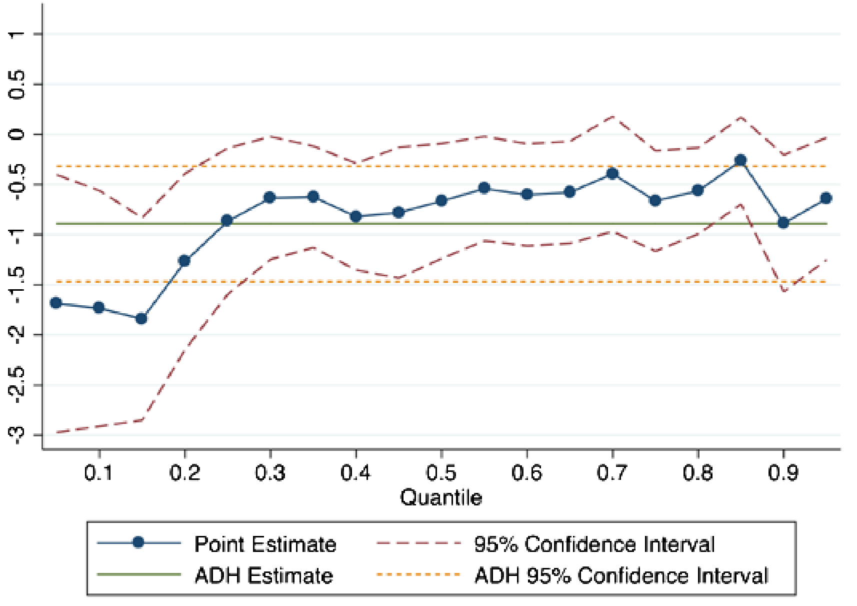

FIGURE 3.—Effect of Chinese import competition on conditional wage distribution: males only. *Notes*: Figure plots grouped IV quantile regression estimates for the male-only sample of the effect of a \$1,000 increase in Chinese imports per worker on the male conditional wage distribution (β_1 in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta\overline{\ln w_g}$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (9). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

7. CONCLUSION

In this paper, we present a quantile extension of Hausman and Taylor (1981), modeling the distributional effects of an endogenous group-level treatment. We develop an estimator, which we refer to as grouped IV quantile regression, and show that the estimator, as well as its standard errors, are easy to compute. We demonstrate that, in contrast to standard quantile regression, this estimator is asymptotically unbiased in the presence of the group-level shocks that are ubiquitous in applied microeconomic models. We illustrate the model and estimator with examples from labor, education, industrial organization, and urban economics. An empirical application to the setting of Autor, Dorn, and Hanson (2013) highlights the usefulness of our approach by estimating the effects of Chinese import competition on the distribution of wages—insights which would be missed by focusing on average effects alone. We believe the estimator has the potential for widespread practical use in applied microeconomics.

REFERENCES

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117. [811, 818]

ABREVAYA, J., AND C. DAHL (2008): "The Effects of Birth Inputs on Birthweight," *Journal of Business and Economic Statistics*, 26, 379–397. [811]

ALTONJI, J., AND R. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models With Endogenous Regressors," *Econometrica*, 73, 1053–1102. [810]

ANGRIST, J., AND K. LANG (2004): "Does School Integration Generate Peer Effects? Evidence From Boston's Metco Program," *American Economic Review*, 94, 1613–1634. [810,818,819]

ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2006): "Quantile Regression Under Misspecification, With an Application to the US Wage Structure," *Econometrica*, 74, 539–563. [817]

ARELLANO, M., AND S. BONHOMME (2013): "Random Effects Quantile Regression," Working Paper. [811]

AUTOR, D., D. DORN, AND G. HANSON (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103 (6), 2121–2168. [811,826,830]

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90, 300–323. [826]

BACKUS, M. (2014): "Why Is Productivity Correlated With Competition?" Working Paper. [810, 821]

BAUM-SNOW, N. (2007): "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122, 775–805. [820]

CANAY, I. (2011): "A Simple Approach to Quantile Regression for Panel Data," *Econometrics Journal*, 14, 368–386. [811]

CHAMBERLAIN, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics, Sixth World Congress*, Vol. 1. Cambridge: Cambridge University Press, 171–209. [817]

CHERNOZHUKOV, V. (2005): "Extremal Quantile Regression," *The Annals of Statistics*, 33, 806–839. [823]

CHERNOZHUKOV, V., AND I. FERNÁNDEZ-VAL (2011): "Inference for Extremal Conditional Quantile Models, With an Application to Market and Birthweight Risks," *Review of Economic Studies*, 78, 559–589. [823]

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. [811,814,818]

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525. [811,814,816]

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398. [811,814]

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41, 2786–2819. [823]

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. [811]

CHETVERIKOV, D., B. LARSEN, AND C. PALMER (2016): "Supplement to 'IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA12121. [810]

FEENSTRA, R. C. (2010): *Offshoring in the Global Economy: Microeconomic Structure and Macroeconomic Implications*. Cambridge, MA: MIT Press. [826]

FEENSTRA, R. C., AND G. H. HANSON (1999): "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the U.S., 1979–1990," *Quarterly Journal of Economics*, 114 (3), 907–940. [826]

GALVAO, A. (2011): "Quantile Regression for Dynamic Panel Data With Fixed Effects," *Journal of Econometrics*, 164, 142–157. [811,820]

GALVAO, A., AND L. WANG (2013): "Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data," Working Paper. [811,820]

GRAHAM, B., AND J. POWELL (2012): "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Panel Data Models," *Econometrica*, 80, 2105–2152. [815]

HAHN, J., AND J. MEINECKE (2005): "Time-Invariant Regressor in Nonlinear Panel Model With Fixed Effects," *Econometric Theory*, 21, 455–469. [815,816]

HASKEL, J., R. LAWRENCE, E. E. LEAMER, AND M. J. SLAUGHTER (2012): "Globalization and U.S. Wages: Modifying Classic Theory to Explain Recent Facts," *Journal of Economic Perspectives*, 26 (2), 119–140. [826]

HAUSMAN, J. (2001): "Mismeasured Variables in Econometric Analysis: Problems From the Right and Problems From the Left," *Journal of Economic Perspectives*, 15, 57–67. [810]

HAUSMAN, J., AND W. TAYLOR (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398. [809-811,814,815,825,830]

HAUSMAN, J., Y. LUO, AND C. PALMER (2014): "Errors in the Dependent Variable of Quantile Regression Models," Working Paper. [810]

IMBENS, G., AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512. [811]

KATO, K., AND A. GALVAO (2011): "Smoothed Quantile Regression for Panel Data," Working Paper. [811,815]

KATO, K., A. GALVAO, AND G. MONTES-ROJAS (2012): "Asymptotics for Panel Quantile Regression Models With Individual Effects," *Journal of Econometrics*, 170, 76–91. [811,815, 817]

KATZ, L. F., AND D. AUTOR (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, Vol. 3A, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier Science, 1463–1555. [826]

KOENKER, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89. [811,815,820]

KOENKER, R., AND J. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50. [810, 815-817]

KOMAROVA, T., T. SEVERINI, AND E. TAMER (2012): "Quantile Uncorrelation and Instrumental Regressions," *Journal of Econometric Methods*, 1, 2–14. [816]

KRUGMAN, P. (2000): "Technology, Trade and Factor Prices," *Journal of International Economics*, 50 (1), 51–71. [826]

LAMARCHE, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408. [811,820]

LARSEN, B. (2014): "Occupational Licensing and Quality: Distributional and Heterogeneous Effects in the Teaching Profession," Working Paper. [810,819,820]

LEAMER, E. E. (1994): "Trade, Wages and Revolving Door Ideas," Working Paper 4716, NBER. [826]

LEE, S. (2007): "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, 141, 1131–1158. [811]

PALMER, C. (2011): "Suburbanization and Urban Decline," Working Paper. [810,820]

PONOMAREVA, M. (2011): "Identification in Quantile Regression Panel Data Models With Fixed Effects and Small T," Working Paper. [811]

ROSEN, A. (2012): "Set Identification via Quantile Restrictions in Short Panels," *Journal of Econometrics*, 166, 127–137. [811]

TOLBERT, C. M., AND M. SIZER (1996): "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Economic Research Service Staff Paper 9614. [826]

VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer. [824]

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838. [825]

Dept. of Economics, UCLA, 315 Portola Plaza, Los Angeles, CA 90095, U.S.A.; chetverikov@econ.ucla.edu,

Dept. of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, U.S.A. and NBER; bjlarsen@stanford.edu,

and

Haas School of Business, University of California at Berkeley, Berkeley, CA 94720, U.S.A.; cjpalmer@ berkeley.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received December, 2013; final revision received August, 2015.