Abstract

We study the conditions under which it is possible to estimate regression quantiles by estimating conditional means. The advantage of this approach is that it allows the use of methods that are only valid in the estimation of conditional means, while still providing information on how the regressors affect the entire conditional distribution. The methods we propose are not meant to replace the well-established quantile regression estimator, but provide an additional tool that can allow the estimation of regression quantiles in settings where otherwise that would be difficult or even impossible. We consider two settings in which our approach can be particularly useful: panel data models with individual effects and models with endogenous explanatory variables. Besides presenting the estimator and establishing the regularity conditions needed for valid inference, we perform a small simulation experiment, present two simple illustrative applications, and discuss possible extensions.

Key words: Endogeneity; Fixed effects; Linear heteroskedasticity; Location-scale model; Quantile regression.

*We thank Geert Dhaene and Paulo Parente for useful comments and advice; the usual disclaimer applies. Address for correspondence:
José Machado, NOVA School of Business and Economics, Portugal. E-mail: jafm@novasbe.pt.
João Santos Silva, School of Economics, University of Surrey, UK. E-mail: jmcss@surrey.ac.uk.
1. INTRODUCTION

We study the conditions under which it is possible to estimate regression quantiles by estimating conditional means. We focus on the conditional location-scale model considered, among others, by Koenker and Bassett (1982), Gutenbrunner and Jurečková (1992), Koenker and Zhao (1984), He (1997), and Zhao (2000), and propose an estimator of the conditional quantiles obtained by combining estimates of the location and scale functions, both of which are defined by conditional expectations of appropriately defined variables.

The advantage of our approach is that it allows the use of methods that are only valid in the estimation of conditional means, such as differencing out individual effects in panel data models, while providing information on how the regressors affect the entire conditional distribution, rather than just some measure of central tendency. These informational gains are perhaps the most attractive feature of quantile regression (see, e.g., the influential papers by Chamberlain, 1994, and Buchinsky, 1994) and were emphasized, for example, in the surveys by Koenker and Hallock (2001) and Cade and Noon (2003). Besides greatly facilitating the estimation of complex models, our approach also leads to estimates of the regression quantiles that do not cross, a crucial requisite often ignored in empirical applications (see also He, 1997, and Chernozhukov, Fernández-Val, and Galichon, 2010).

Because our estimator is based on conditional means, it does not share some of the robustness properties of the seminal quantile regression estimator of Koenker and Bassett (1978), which is based on the check function. For example, our estimator requires stronger assumptions on the existence of moments than those needed for the validity of Koenker and Bassett’s (1978) estimator. However, under the appropriate conditions, our estimator identifies the same conditional quantiles, the optimal predictors under the usual asymmetric loss function, and these are inherently robust.
The setup we consider is restrictive in that we need to assume that the covariates only affect the distribution of interest through known location and scale functions.\textsuperscript{1} However, practitioners are often prepared to make even stronger assumptions,\textsuperscript{2} and we will argue that in spite of its assumptions our approach can be useful in many empirical applications. Importantly, although we do not develop such tests here, it is possible to test the assumption that the covariates only affect the location and scale functions, and therefore it is possible to check whether or not our approach is suitable in a particular application.

The approach we propose is not meant to replace the well established and very attractive estimation methods based on the check-function. Instead, we see our estimator as an additional tool that can complement those techniques and allow the estimation of regression quantiles in settings where otherwise that would be difficult or even impossible. For example, our approach is attractive when panel data are available and the researcher wants to estimate regression quantiles including individual effects.

There is now a substantial literature on quantile regression models with individual effects (see, e.g., Koenker, 2004, Lamarche, 2010, Canay, 2011, Galvão, 2011, Kato, Galvão and Montes-Rojas, 2012, and Powell, 2017). However, none of these methods gained widespread popularity, either because of their computational complexity or because they rely on very restrictive assumptions on how the fixed effects affect the quantiles. Albeit also based on a somewhat restrictive (but testable) assumption, our approach has the advantage of being very easy to implement and it allows the individual effects to affect the entire distribution, rather than being just location shifters as in, e.g., Koenker (2004), Lamarche (2010), and Canay (2011).

\textsuperscript{1}Notice that in a conditional location-scale model the regressors affect all higher-order moments through the scale function. Indeed, in a location-scale model the $m$-th conditional central moment is proportional to the $m$-th power of the scale function.

\textsuperscript{2}For example, the popular Tobit and sample selection models assume that the errors are normally distributed and statistically independent of the regressors.
Our approach can easily be extended to the case where the researcher wants to account for the endogeneity of some of the variables in the model as, for example, in Abadie, Angrist, and Imbens (2002) and in Chernozhukov and Hansen (2005, 2006, and 2008). Strictly speaking, in this context our approach is not based on the estimation of conditional means, but on moment conditions that under exogeneity identify conditional means. The proposed estimator is closely related to that of Chernozhukov and Hansen (2008), in the sense that under suitable regularly conditions it identifies the same structural quantile function, but has the advantage of being applicable to non-linear models and being computationally much simpler to implement, especially in models with multiple endogenous variables.

The remainder of the paper is organized as follows. Section 2 introduces our approach to the estimation of regression quantiles in location-scale models. Section 3 considers the application of our approach in the context of a panel data model with fixed effects. In Section 4 we consider estimation when some of the variables of the model are endogenous. Section 5 presents the results of a small simulation study and Section 6 illustrates the application of the proposed methods with two empirical examples. Section 7 concludes and an Appendix collects the more technical details.

2. THE BASIC IDEA

The rationale of the proposed estimator can be introduced in a simple setup. We are interested in estimating the conditional quantiles of a random variable $Y$ whose distribution conditional on a $k$-vector of covariates $X$ belongs to the location-scale family and can be expressed as

$$Y = \alpha + X'\beta + \sigma(\delta + Z'\gamma)U,$$  

where:

- $(\alpha, \beta', \delta, \gamma')' \in \mathbb{R}^{2(k+1)}$ are unknown parameters;\(^3\)

\(^3\)For simplicity, we assume that $X$ and $Z$ have the same dimension.
• $Z$ is a $k$-vector of known differentiable (with probability 1) transformations of the components of $X$ with element $l$ given by

$$Z_l = Z_l(X), \quad l = 1, \ldots, k;$$

• $\sigma(\cdot)$ is a known $C^2$ function such that

$$P\{\sigma(\delta + Z'\gamma) > 0\} = 1;$$

• $U$ is an unobserved random variable, independent of $X$, with density function $f_U(\cdot)$ bounded away from 0 and normalized to satisfy the moment conditions

$$E(U) = 0 \quad E(|U|) = 1.$$  \hspace{1cm} (2)

A special case of (1) is, of course, the linear heteroskedasticity model in which $\sigma(\cdot)$ is the identity function and $Z = X$. This model has been study by many authors and has a long tradition in the quantile regression literature (see, e.g., Koenker and Basset, 1982, Gutenbrunner and Jurečková, 1992, Koenker and Zhao, 1984, He, 1997, and Zhao, 2000). Our formulation, however, is sufficiently general to also encompass other specifications such as models with multiplicative heteroskedasticity (Harvey, 1976), which have recently been advocated by Romano and Wolf (2017) who specify

$$\sigma(\cdot) = \exp(\delta + \gamma_1 \log(|X_1|) + \cdots + \gamma_k \log(|X_k|)).$$

The specification in (1) differs from the standard quantile regression formulation $Q_Y(\tau|x) = x'\beta(U)$, which can be viewed as representing a linear data generating process where all unobserved heterogeneity comes from random parameter variation. The model in (1) imposes that there is a single source of unobserved heterogeneity, $U$, but our formulation allows for nonlinear quantile effects and, thus, (1) cannot be considered a restricted version of $Q_Y(\tau|x) = x'\beta(U)$, except when $\sigma(\cdot)$ is the identity function.

Model (1) implies that

$$Q_Y(\tau|X) = \alpha + X'\beta + \sigma(\delta + Z'\gamma) q(\tau)$$ \hspace{1cm} (3)
with \( q(\tau) = F_U^{-1}(\tau) \), and therefore \( \Pr(U < q(\tau)) = \tau \). In the case where \( \sigma(\cdot) \) is the identity function and \( Z = X \), the quantiles simplify to

\[
Q_Y(\tau|X) = (\alpha + \delta q(\tau)) + X'(\beta + \gamma q(\tau)).
\]

In general, the marginal effect of the regressor \( X_i \) on the \( \tau \)-th quantile of \( Y \) (the “regression quantile coefficient”) is

\[
\beta_i(\tau, X) = \beta_i + q(\tau) D_{X_i}^\sigma
\]

(4)

with \( D_{X_i}^\sigma = \partial \sigma(\delta + Z'\gamma)/\partial X_i \).

Using (2), and the exogeneity of the regressors, the vector of parameters of interest,

\[
(\alpha, \beta', \delta, \gamma', q(\tau))',
\]

can be identified from the following set of moment conditions (for ease of exposition we assume here i.i.d. data):

\[
E[RX] = 0
\]

\[
E[R] = 0
\]

\[
E[(|R| - \sigma(\delta + Z'\gamma)) D_{\gamma}^\sigma] = 0 \quad \text{(MC1)}
\]

\[
E[(|R| - \sigma(\delta + Z'\gamma)) D_{\delta}^\sigma] = 0
\]

\[
E[I(R \leq q(\tau) \sigma(\delta + Z'\gamma)) - \tau] = 0
\]

where \( R = Y - (\alpha + X'\beta) = \sigma(\delta + Z'\gamma)U, D_{\gamma}^\sigma = \partial \sigma(\delta + Z'\gamma)/\partial \gamma, D_{\delta}^\sigma = \partial \sigma(\delta + Z'\gamma)/\partial \delta \).

Given that the location-scale model specifies the scale function \( \sigma(\cdot) \), we can explore that information and base the identification on the alternative set of moment conditions

\[
E[U] = 0
\]

\[
E[U] = 0
\]

\[
E[(|U| - 1) D_{\gamma}^\sigma] = 0 \quad \text{(MC2)}
\]

\[
E[(|U| - 1) D_{\delta}^\sigma] = 0
\]

\[
E[I(U < q(\tau)) - \tau] = 0
\]
where \( U = (Y - (\alpha + X'\beta)) / \sigma (\delta + Z'\gamma) \).\(^4\)

These conditions form the basis of the estimation procedure (Method of Moments-Quantile Regression, MM-QR) discussed in further detail in the next sections. Conditions (MC1) bear resemblance to those of the Restricted Quantile Regression of He (1997) and Zhao (2000) but we explore different moment conditions. In He (1997) and Zhao (2000) the moment conditions corresponding to (2) are that \( U \) has median at zero and that \(|U|\) has median at 1. Thus, the implied orthogonality condition corresponding to (MC1) are those defining least absolute deviation estimators rather than least squares estimators. Our choice is, admittedly, weaker from a robustness point of view, but we believe that our approach is useful in that it makes it very easy to implement quantile regression in a wider class of models.\(^5\) In particular, we will explore the use of (MC1) in the estimation of panel data models with fixed-effects, and (MC2) in the estimation of structural quantile functions as defined by Chernozhukov and Hansen (2006, 2008).

We conclude this section by establishing that the conditional quantile estimates obtained from (MC1) or (MC2) do not cross; the result follows directly from the unidimensional nature of the quantile estimator implied by the last moment conditions of (MC1) and (MC2).

**Proposition 1 (No Quantile-Crossing: He, 1997)** Consider the regression quantile \( Q_Y(\tau | X) \) given by (3) and its estimate \( \hat{Q}_Y(\tau | X) \) obtained from an estimate \( \hat{\sigma} (Z) \) of

\(^4\)Although we do not pursue that here, it is easy to see that the validity of the location-scale model can be tested, for example, by testing the overidentifying restrictions resulting form augmenting (MC1) and (MC2) with conditions imposing the orthogonality between suitable functions of \( U_\alpha \) and functions of the regressors. See, e.g., Hansen (1982) and Newey (1985).

\(^5\)Notice that, due to the normalization in (2), we estimate the scale function rather than the skedastic function. There are two reasons for this. First, in the leading case where the scale is a linear function of the regressors and the quantiles are linear, the scale function can be estimated by ordinary least squares, whereas estimation of the skedastic function would involve non-linear estimation. Additionally, as noted by Koenker and Zhao (1996), the scale function is a more robust measure of dispersion.
\( \sigma (\delta + Z'\gamma) \). Then
\[
\tau \leq \tau' \iff \hat{Q}_Y (\tau | X) \leq \hat{Q}_Y (\tau' | X),
\]
for any design point with \( \hat{\sigma} (Z) > 0 \). \( \square \)

3. PANEL DATA WITH FIXED EFFECTS

3.1. Linear models

The estimation of linear regression quantiles for longitudinal data was seminally considered by Koenker (2004). In Koenker’s model however, the individual effects only cause parallel (location) shifts of the distribution of the response variable (see also Lamarche, 2010, Canay, 2011, and Galvão, 2011). We also start by considering a linear specification, but allow the individual effects to affect the entire distribution.

Given data \( \{(Y_{it}, Z'_{it})\}' \) from a panel of \( n \) individuals \( i = 1, \ldots, n \) over \( T \) time periods, \( t = 1, \ldots, T \), we consider the estimation of the conditional quantiles \( Q_Y (\tau | X) \) for a location-scale model of the form
\[
Y_{it} = \alpha_i + X'_{it}\beta + (\delta_i + Z'_{it}\gamma)U_{it}, \quad (5)
\]
with \( P\{\delta_i + Z'_{it}\gamma > 0\} = 1 \). The parameters \( (\alpha_i, \delta_i), i = 1, \ldots, n, \) capture the individual \( i \) fixed effects and \( Z \) is defined as before. The sequence \( \{X_{it}\} \) is \( i.i.d. \) for any fixed \( i \) and independent across \( t \). \( U_{it} \) are \( i.i.d. \) (across \( i \) and \( t \)), statistically independent of \( X_{it} \), and normalized to satisfy the moment conditions (2).

Model (5) implies that
\[
Q_Y (\tau | X_{it}) = (\alpha_i + \delta_i q (\tau)) + X'_{it}\beta + Z'_{it}\gamma q (\tau).
\]

We will call the scalar coefficient \( \alpha_i (\tau) \equiv \alpha_i + \delta_i q (\tau) \) the quantile-\( \tau \) fixed effect for individual \( i \), or the distributional effect at \( \tau \). The distributional effect differs from the usual fixed effect in that it is not, in general, a location shift. That is, the distributional effect represents the effect of time-invariant individual characteristics which, like other
variables, are allowed to have different impacts on different regions of the conditional distribution of $Y$. The fact that $\int_0^1 q(\tau)\,d\tau = 0$ implies that $\alpha_i$ can be interpreted as the average effect for individual $i$.

Consider now the MM-QR estimator of (5) implied by (MC1). For this model, the moment conditions have a convenient triangular structure with respect to the model parameters that allows the one-step GMM estimator (Hansen, 1982) to be calculated sequentially:

1. Regress $(Y_{it} - \sum_t Y_{it}/T)$ on $(X_{it} - \sum_t X_{it}/T)$ by least squares to obtain $\hat{\beta}$;
2. Estimate $\hat{\alpha}_i = \frac{1}{T} \sum_t (Y_{it} - X'_{it}\hat{\beta})$ and obtain the residuals $\hat{R}_{it} = Y_{it} - \hat{\alpha}_i - X'_{it}\hat{\beta}$;
3. Regress $(|\hat{R}_{it}| - \sum_t |\hat{R}_{it}|/T)$ on $(Z_{it} - \sum_t Z_{it}/T)$ by least squares to obtain $\hat{\gamma}$;
4. Estimate $\hat{\delta}_i = \frac{1}{T} \sum_t (|\hat{R}_{it}| - Z'_{it}\hat{\sigma})$;
5. Estimate $q(\tau)$ by $\hat{q}$, the solution to

$$\min_q \sum_i \sum_t \rho_r \left( \hat{R}_{it} - \left( \hat{\delta}_i + Z'_{it}\hat{\gamma} \right) q \right)$$

where $\rho_r(A) = (\tau - 1)I\{A \leq 0\} + \tau I\{A > 0\}$ is the check-function. (Equivalently, order the standardized residuals $\hat{U} = \hat{R}_{it}/\left( \hat{\delta}_i + Z'_{it}\hat{\gamma} \right)$ and estimate the $\tau$-th sample quantile.)

The following results could be approached using a standard GMM framework for the exactly identified case and the results of, say, Newey and McFadden, (1994, Theorem 7.2). Our approach however, mimics the sequence of steps above and is similar to Zhao’s (2000).

The regression in Step 3 is reminiscent of the one used to compute Glejser’s (1969) test for heteroskedasticity, and the insights in Machado and Santos Silva (2000) and Im (2000) suggest that the MM-QR estimator is greatly simplified if $|R|$ in (MC1) is replaced by

$$2R(I\{R \geq 0\} - P\{R \geq 0\}).$$
Indeed, because $|R| = 2R(I\{R \geq 0\} - 1/2)$, the two transformations differ only in the way the residuals $R$ are weighted: with mean zero in one case and with mean $P\{R \geq 0\} - 1/2$ in the other. Using the assumption that $E[R|Z] = 0$, it is clear that the (population) moment condition

$$E[Z (|R| - \delta_i - Z'\gamma)] = 0$$

identifies $\delta_i$ and $\gamma$ iff

$$E[Z (2R(I\{R \geq 0\} - \eta) - \delta_i - Z'\gamma)] = 0, \quad \eta = P(R \geq 0) = P(U \geq 0).$$

Therefore, in Steps 3 and 4, instead of using $|\hat{R}|$ one may use

$$\hat{R}_{it}[I\{\hat{R}_{it} \geq 0\} - \hat{\eta}]$$

with

$$\hat{\eta} = \frac{1}{nT} \sum_i \sum_t I\{\hat{R}_{it} \geq 0\}.$$

The advantage of using this alternative transformation in Steps 3 and 4 of the algorithm is that it makes asymptotic inference on $\gamma$ independent of the first step estimator. Besides simplifying the treatment of the asymptotic properties of the estimator, this allows the practitioner to make inference about the parameters of the scale function directly from the least squares results in Step 3, without having to take into account the first-step estimation.

Below we present the main results on the asymptotic properties of the MM-QR estimator as a set of two theorems whose proofs are provided in the Appendix. Throughout we use the following notation: for any sequence of random variables $A_{it}, B_{it}$ for which the limits exist,

$$Q_{AB} = \lim_{n \to \infty} \frac{1}{n} \sum_i E[(A_{i1} - \mu_{A_i})(B_{i1} - \mu_{B_i})']$$

with $\mu_{A_i} = E[A_{i1}]$,

$$P_{AB} = \lim_{n \to \infty} \frac{1}{n} \sum_i E[\sigma_{i1}^2(A_{i1} - \mu_{A_i})(B_{i1} - \mu_{B_i})']$$
with $\sigma_{it} = (\delta_i + Z_{it}'\gamma)$, and

$$P_A = \lim_{n \to \infty} \frac{1}{n} \sum_i E[\sigma_{i1}^2(A_{i1} - \mu_{A_i})].$$

The following theorem establishes the asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$.

**Theorem 1 (Slope coefficients)** Consider model (5) satisfying conditions (P) in the Appendix. Assume further that the sequences $\{X_{it}, Z_{it}, U_{it}\}$ satisfy the conditions (U) and (XZ) in the Appendix. Then, as $\{n, T\} \to \infty$

$$\sqrt{nT}(\hat{\beta} - \beta) \xrightarrow{D} Q_{XX}^{-1}N(0, E(U^2)P_{XX})$$

and if $n = o(T)$,

$$\sqrt{nT}(\hat{\gamma} - \gamma) \xrightarrow{D} Q_{ZZ}^{-1}N(0, E(V^2)P_{ZZ})$$

with $V = 2U(I\{U \geq 0\} - P\{U \geq 0\})$. □

Notice that, as is well known, the result for the least squares estimator $\hat{\beta}$ does not require any restriction on the rate of growth of $n$. Notice also that, as mentioned before, the limiting distribution of $\hat{\gamma}$ does not depend on the first-step estimation.

The quantile-$\tau$ fixed effect, $\alpha_{i}(\tau) = \alpha_i + \delta_i q(\tau)$, can be estimated by

$$\hat{\alpha}_{i}(\tau) = \frac{1}{T} \sum_{t=1}^{T} (Y_{it} - X_{it}'\hat{\beta}) + \hat{\gamma} \frac{1}{T} \sum_{t=1}^{T} (|\hat{R}_{it}| - Z_{it}'\hat{\gamma}).$$

Likewise, the $\tau$-th quantile regression coefficient of the regressor $X_{it}$, which is given by (4) and is the main parameter of interest, can be estimated by

$$\hat{\beta}_{i}(\tau, X) = \hat{\beta}_{i} + \hat{\gamma}.$$ 

Theorem 2 below establishes the asymptotic distribution of $\hat{\beta}_{i}(\tau, X)$ for the leading case where $Z = X$ and $\hat{\beta}_{i}(\tau, X) = \hat{\beta}_{i}(\tau)$; the more general case is equally straightforward.\(^6\)

\(^6\)With $Z = X$, $Q_{ZZ} = Q_{XX}$, and $P_{ZZ} = P_{ZX} = P_{XX}$; if $Z \neq X$, $\Xi$ has to be adjusted in a straightforward way.
Theorem 2 (Quantile Regression Coefficients) Consider model (5) satisfying conditions (P) in the Appendix and assume that \( Z = X \). Assume further that the sequences \( \{X_{it}, Z_{it}, U_{it}\} \) satisfy the conditions in (U) and (XZ) in the Appendix. Then, as \( \{n, T\} \to \infty \) with \( n = o(T) \)

\[
\sqrt{nT}(\hat{\beta}(\tau) - \beta(\tau)) \overset{D}{\longrightarrow} N(0, \Omega)
\]

with

\[
\Xi = (Q_{XX}^{-1}|q(\tau)Q_{ZZ}^{-1}|(1/\mu_\sigma)\gamma),
\]

being a \( k \times (2k + 1) \) block matrix, and

\[
\Omega = \begin{pmatrix}
E[U^2]P_{XX} & E[UV]P_{XZ} & E[UW]P_X \\
E[Ṽ^2]P_{ZZ} & E[ṼW]P_X & \mu_\sigma^2 E[Ṽ^2]
\end{pmatrix}
\]

with \( \mu_\sigma = \frac{1}{n} \sum_i (\delta_i + \gamma E[Z_{i1}])^a \), \( a = 1, 2 \) and \( W = \frac{1}{f_U(q(\tau))} \psi_\tau(U - q(\tau)) - U - q(\tau) V \).

It is not difficult to establish (but of limited practical interest) that the MM-QR estimators of the fixed effects coefficients \( \alpha_i, \delta_i \) converge at rate \( \sqrt{T} \) to a Gaussian distribution. Owing to the faster rates of convergence of the slope estimators, this asymptotic distribution of the fixed effects coefficients is the same as if the slopes were known.

More interesting is to discuss the consistency of this estimators in the sense of the limiting behavior of the random sequences \( \max_{i \leq n} |\hat{\alpha}_i - \alpha_i| \) and \( \max_{i \leq n} |\hat{\delta}_i - \delta_i| \). Lemmata 1 and 4 in the Appendix prove that if \( n/T \to 0 \) as \( \{n, T\} \to \infty \), then

\[
\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_i| = o_P(1)
\]

and

\[
\max_{1 \leq i \leq n} |\hat{\delta}_i - \delta_i| = o_P(1).
\]

The requirement for this consistency is the only source of restrictions on the rate of growth of \( n \) as \( T \to \infty \) and, of course, this impacts on the estimation both of the scale function and of \( q(\tau) \).
Given that most of our results depend on \( \{n, T\} \to \infty \) with \( n = o(T) \), and because in most applications \( n \) is much larger than \( T \), it is important to consider the usefulness of the proposed estimator in applied settings. In Section 5 we present simulation results for a range of values of \( n, T, \) and \( n/T \). Previewing our results, we can say that we find the expected drop in the bias as \( T \) grows, and that the bias also drops with \( n \) for the smaller values of \( n \) considered in the simulations.

As expected, our simulations also show that the precision of the estimators increases with \( nT \). As noted by Hahn and Newey (2004), the fact that the variance of the estimator depends on \( n \) but the bias does not, may lead to confidence intervals with poor coverage in applications where \( n/T \) is large (see also Kato, Galvão, and Montes-Rojas, 2012). In Section 5 we briefly discuss potential solutions for this problem.

### 3.2. Non-linear models

The linear heteroskedasticity model considered so far is particularly attractive for its long history and for its simplicity, but estimation with other specifications of the location and scale functions is also possible. However, in specifications with fixed effects, estimating non-linear models will generally be impractical.

The exception to this are specifications based on the exponential function because in this case, just like in the linear model, there is a transformation that eliminates the fixed effects. Indeed, Wooldridge (1999) shows that the so-called fixed effects Poisson regression with an exponential conditional mean, which conditions-out the individual effects, is valid under very general conditions and is easy to implement (notice that this estimator is valid even if the data are not counts).

Therefore, when either the conditional mean, the conditional variance, or both, are given by exponential functions, all that is needed is to replace the corresponding least squares steps in the algorithm described before with suitable Poisson regressions; naturally, the subsequent computation of the fixed effects needs to be modified accordingly,
but that is trivial. Using the delta-method and our earlier results, it is straightforward to derive the asymptotic distribution of the estimators in these non-linear models.

The possibility of estimating models with \( \sigma(\cdot) = \exp(\cdot) \) is particularly interesting because this specification ensures that \( \sigma(\cdot) > 0 \). Moreover, models with multiplicative heteroskedasticity also have a long history and are popular in many contexts (see, e.g., Harvey, 1976, Wooldridge, 2010, and Romano and Wolf, 2017).7

4. ENDOGENOUS REGRESSORS

We explore now the application of the MM-QR estimator to models with endogenous regressors. Consider a scalar random variable \( Y \) related to a vector of observed random variables \((D', C'_1, C'_2)'\) (with dimensions \( k_D, k_1, k_2 \), respectively, and \( k_2 \geq k_D \)), and to an unobserved scalar random variable \( U \) satisfying (2), according to the structural relationship

\[
Y = D'\beta_D + C'_1\beta_1 + \sigma (D'\gamma_D + C'_1\gamma_1)U
\]

\( D_l = D_l(C_1, C_2, U) \) for \( l = 1, \ldots, k_D \)

\( C_1, C_2 \) statistically independent of \( U \),

where \( D_l(\cdot) : \mathbb{R}^{k_1+k_2+1} \rightarrow \mathbb{R} \) and \( \sigma(\cdot) \) is as defined in Section 2. The parameters \((\beta, \gamma) \in \Omega_2\), satisfy assumption (P1) in the Appendix. Put \( X' = (D', C'_1) \) (the regressors), \( C'' = (C'_1, C'_2) \) (the instruments), \( \beta' = (\beta'_D, \beta'_1) \), and \( \gamma' = (\gamma'_D, \gamma'_1) \).8

The most relevant feature of this model is that the endogenous regressor impacts both the location and scale of \( Y \). Although similar, (6) is neither more nor less restrictive than the structural random coefficients model considered by Chernozhukov and Hansen (2006, 2008). As noted before, we impose that there is only one source of unobserved heterogeneity, \( U \), whereas Chernozhukov and Hansen (2006, 2008) essentially allow one

---

7For example, the latest release of Stata (StataCorp., 2017) includes the command `hetregress` which estimates linear regression models with multiplicative heteroskedasticity.

8Notice that if the location and scale have intercepts, \( C_1 \) will have a column of 1s.
source of unobserved heterogeneity per explanatory variable. However, unlike them, we allow for non-linear quantile effects. Our model is also very easy to estimate even when there are multiple endogenous explanatory variables.

As in Chernozhukov and Hansen (2006, 2008), we are not interested in estimating $Q_Y(\tau|X)$, but the parameters of a function $S_Y(\tau|X)$ such that

$$P\{Y \leq S_Y(\tau|X)\} = P\{Y \leq S_Y(\tau|X)|C\} = \tau.$$  

Therefore, $S_Y(\tau|X)$, the “structural quantile function” in Chernozhukov and Hansen’s (2008) terminology, can be interpreted as $Q_Y(\tau|C)$ and can be written as

$$S_Y(\tau|X) = X'\beta + \sigma(X'\gamma)q(\tau).$$

Given the model, if $(\beta, \gamma)$ were known, the moment condition

$$E\left[\psi_\tau\left(\frac{Y - X'\beta}{\sigma(X'\gamma)} - q\right)\right] = 0$$

with $\psi_\tau(A) = (\tau - I\{A \leq 0\})$, would identify the marginal quantile of $U$, that is $q(\tau)$ such that $P\{U \leq q(\tau)\} = P\{U \leq q(\tau)|C\} = \tau$. This procedure is not feasible since $\beta$ and $\gamma$ are not known but, given the data $\{(Y_i, X'_i, C'_i)\}$, these parameters can be consistently estimated under very general conditions by GMM applied to the sample analogues of the moment conditions in (MC2),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} C_i \left(\frac{Y_i - X'_i\hat{\beta}}{\sigma(X'_i\hat{\gamma})}\right) = 0,$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} C_i \left(\frac{|Y_i - X'_i\hat{\beta}|}{\sigma(X'_i\hat{\gamma})} - 1\right) = 0.$$  

Notice that this MM-QR estimator cannot be solved sequentially, and therefore in this case there is no practical benefit in replacing $|U_i|$ with $2U_i (I\{U_i > 0\} - P\{U > 0\})$.

Given the estimates of $\beta$ and $\gamma$, $q(\tau)$ may be estimated by the condition

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau\left(\frac{Y_i - X'_i\hat{\beta}}{\sigma(X'_i\hat{\gamma})} - q\right) = o_P(1).$$
or, alternatively, by ranking the standardized residuals.

The next theorem formalizes this estimator for the exactly identified case \((k_D = k_2)\); the over-identified case could be handled similarly.

**Theorem 3 (Structural quantile function coefficients)** Consider a sample of \(n\) i.i.d. observations of \((Y, X, C)\) from the structure defined by (6) with \(\text{dim}(X) = \text{dim}(C)\). Then, under assumptions \((P), (U),\) and \((DC)\) in the Appendix, as \(n \to \infty\)

\[
\begin{pmatrix}
\sqrt{n}(\hat{\beta} - \beta) \\
\sqrt{n}(\hat{\gamma} - \gamma) \\
\sqrt{n}(\hat{q} - q(\tau))
\end{pmatrix}
\xrightarrow{\mathcal{D}}
G^{-1} \mathcal{N}(0, \Omega),
\]

where,

\[
\Omega = 
\begin{pmatrix}
E[U^2]E[CC'] & E[UV]E[CC'] & \frac{E[U\psi_q(U-q(\tau))]}{f_U(q(\tau))} E[C] \\
E[UV]E[CC'] & E[V^2]E[CC'] & \frac{E[V\psi_q(U-q(\tau))]}{f_U(q(\tau))} E[C] \\
\frac{1}{f_U(q(\tau))} & \frac{1}{f_U(q(\tau))} & (1 - \tau)
\end{pmatrix},
\]

with \(V = |U| - 1\) and

\[
G = 
\begin{pmatrix}
E[(1/\sigma) CX'] & E[(\sigma'/\sigma) UX'] & 0_{k \times 1} \\
E[(1/\sigma) \text{sign}(U) CX'] & E[(\sigma'/\sigma) |U| CX'] & 0_{k \times 1} \\
E[(1/\sigma) X'] & E[(\sigma'/\sigma) UX'] & 1
\end{pmatrix},
\]

with \(k = k_1 + k_2, \sigma = \sigma(X'\gamma),\) and \(\sigma' = d\sigma(z)/dz\) at \(z = X'\gamma\). □

Inference about \(\beta(\tau, X) = \partial S_Y(\tau|X)/\partial X\), the ultimate parameter of interest, can be performed using the standard delta-method. For example, in the linear case where \(\beta(\tau, X) = \beta(\tau) = \beta + \gamma q(\tau)\) we have that

\[
\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{\mathcal{D}} AG^{-1} \mathcal{N}(0, \Omega),
\]

where

\[
A = (I_{k \times k} \ | \ q(\tau) I_{k \times k} \ | \ \gamma)
\]

is a \(k \times (2k + 1)\) block matrix and \(I\) denotes the identity matrix.

In the next sections we present simulation results and an empirical example illustrating the performance and application of this estimator.
5. SIMULATION EVIDENCE

This section presents the results of two small simulation exercises illustrating the performance of the methods proposed in this paper.

5.1. Panel data models with fixed effects

The first set of experiments is designed to study the performance of the estimator in a panel-data model with fixed effects. For this experiment, 10,000 independent data sets were generated as

\[ Y_{it} = \alpha_i + X_{it} + (1 + \chi_{it} + \kappa \alpha_i) U_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \]  

(7)

where \( \alpha_i \sim \chi^2_{(1)} \) and \( X_{it} = 0.5 (\alpha_i + \chi_{it}) \), with \( \chi_{it} \sim \chi^2_{(1)} \), and three different distributions of \( U_{it} \) are considered: standard normal, \( \chi^2_{(5)} \), and \( t_{(5)} \); in all cases \( U_{it} \) is standardized to have zero mean and unit variance.\(^9\) We performed simulations for \( T \in \{10, 20, 50\} \), \( n \in \{50, 500, 100T\} \), \( \tau \in \{0.25, 0.75\} \), and \( \kappa \in \{0, 1\} \). For \( \kappa = 0 \) the fixed effects are pure location shifts as in Koenker (2004) and Canay (2011); otherwise the fixed effects affect the entire distribution. We estimate (7) using the estimator described in Section 3 and, for comparison, we also estimate the model using the attractive estimator proposed by Canay (2011), which is valid only when \( \kappa = 0 \).\(^10\)

Tables 1 and 2 report the bias and mean squared error (MSE) for all the cases with \( \tau = 0.25 \); for brevity, we do not report the results obtained with \( \tau = 0.75 \) because they lead to similar conclusions.

When \( \kappa = 0 \) both estimators are valid and there is little to choose between them. However, Canay’s (2011) estimator, which imposes the valid restriction that the fixed effects are pure location shifters, systematically has lower MSE than the MM-QR estimator, and often has somewhat smaller bias.

---

\(^9\)Using this normalization rather than \( E |U_{it}| = 1 \) is immaterial and facilitates the data generation.

\(^10\)Notice that the first step in Canay’s estimator coincides with the first step of our estimator.
The bias of the MM-QR estimator changes little when $\kappa$ is set to 1, but we generally observe an increase in the MSE resulting from the additional variability in the data. In contrast, the performance of Canay’s (2011) estimator deteriorates sharply due to its large bias when the fixed effects are more than location shifters. That is, Canay’s (2011) estimator is naturally sensitive to departures from its key assumption and, in cases where this is a concern, the MM-QR estimator offers a practical alternative to the methods that assume that the fixed effects only shift the location of the distribution.

It is reassuring to find that, for both values of $\kappa$, the performance of the MM-QR estimator is reasonable even when the errors are non-normal. Indeed, with $\kappa = 0$, it is

<table>
<thead>
<tr>
<th>$n = 50$</th>
<th>$n = 500$</th>
<th>$n = 100 \times T$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1: $N(0,1)$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 10$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.071</td>
<td>0.096</td>
</tr>
<tr>
<td>MSE</td>
<td>0.080</td>
<td>0.110</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.035</td>
<td>0.046</td>
</tr>
<tr>
<td>MSE</td>
<td>0.038</td>
<td>0.054</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.014</td>
<td>0.021</td>
</tr>
<tr>
<td>MSE</td>
<td>0.015</td>
<td>0.022</td>
</tr>
<tr>
<td><strong>Model 2: $\chi^2_{(5)}$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 10$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.088</td>
<td>0.153</td>
</tr>
<tr>
<td>MSE</td>
<td>0.061</td>
<td>0.075</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.044</td>
<td>0.079</td>
</tr>
<tr>
<td>MSE</td>
<td>0.027</td>
<td>0.030</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.017</td>
<td>0.031</td>
</tr>
<tr>
<td>MSE</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Model 3: $t_{(5)}$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 10$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.048</td>
<td>0.053</td>
</tr>
<tr>
<td>MSE</td>
<td>0.070</td>
<td>0.108</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>MSE</td>
<td>0.034</td>
<td>0.055</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>Canay</td>
<td>MM-QR</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.013</td>
<td>0.011</td>
</tr>
<tr>
<td>MSE</td>
<td>0.013</td>
<td>0.022</td>
</tr>
</tbody>
</table>
only with non-normal errors that the bias of the MM-QR estimator is smaller than that of Canay’s; this is especially true in the unreported results for $\tau = 0.75$.

Focusing now on the performance of the MM-QR estimator, we find the expected drop in the bias as $T$ grows; in particular, our results suggest that the bias is essentially proportional to $1/T$. Moreover, we find that initially the bias also drops with $n$, but increasing $n$ has no effect on the bias beyond a certain point. In these simulations, the value of $n/T$ ranges from 1 to 100, and the results in Table 1 suggest that the value of this ratio is not particularly relevant for the bias of the estimator. In particular, we see that the bias does not increase by increasing $n$ for a given $T$, and it may even decrease for small values of $n$.

Table 2: Bias and MSE results for $\tau = 0.25$ and $\kappa = 1$

<table>
<thead>
<tr>
<th></th>
<th>$n = 50$</th>
<th>$n = 500$</th>
<th>$n = 100 \times T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canay</td>
<td>MM-QR</td>
<td>Canay</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>$T = 10$</td>
<td>-0.436</td>
<td>0.327</td>
<td>-0.452</td>
</tr>
<tr>
<td></td>
<td>-0.453</td>
<td>0.078</td>
<td>-0.453</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>-0.492</td>
<td>0.314</td>
<td>-0.506</td>
</tr>
<tr>
<td></td>
<td>-0.506</td>
<td>0.037</td>
<td>-0.506</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>-0.526</td>
<td>0.310</td>
<td>-0.538</td>
</tr>
<tr>
<td></td>
<td>-0.539</td>
<td>0.014</td>
<td>-0.539</td>
</tr>
</tbody>
</table>

Model 2: $\chi^2_{(5)}$

<table>
<thead>
<tr>
<th></th>
<th>$n = 50$</th>
<th>$n = 500$</th>
<th>$n = 100 \times T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canay</td>
<td>MM-QR</td>
<td>Canay</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>$T = 10$</td>
<td>-0.432</td>
<td>0.281</td>
<td>-0.449</td>
</tr>
<tr>
<td></td>
<td>-0.449</td>
<td>0.129</td>
<td>-0.449</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>-0.505</td>
<td>0.303</td>
<td>-0.520</td>
</tr>
<tr>
<td></td>
<td>-0.520</td>
<td>0.065</td>
<td>-0.520</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>-0.549</td>
<td>0.327</td>
<td>-0.562</td>
</tr>
<tr>
<td></td>
<td>-0.563</td>
<td>0.026</td>
<td>-0.563</td>
</tr>
</tbody>
</table>

Model 3: $t_{(5)}$

<table>
<thead>
<tr>
<th></th>
<th>$n = 50$</th>
<th>$n = 500$</th>
<th>$n = 100 \times T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canay</td>
<td>MM-QR</td>
<td>Canay</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>$T = 10$</td>
<td>-0.396</td>
<td>0.282</td>
<td>-0.402</td>
</tr>
<tr>
<td></td>
<td>-0.401</td>
<td>0.048</td>
<td>-0.401</td>
</tr>
<tr>
<td>$T = 20$</td>
<td>-0.427</td>
<td>0.247</td>
<td>-0.437</td>
</tr>
<tr>
<td></td>
<td>-0.437</td>
<td>0.021</td>
<td>-0.437</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>-0.447</td>
<td>0.229</td>
<td>-0.459</td>
</tr>
<tr>
<td></td>
<td>-0.460</td>
<td>0.008</td>
<td>-0.460</td>
</tr>
</tbody>
</table>
As expected, the precision of the estimators increases with $nT$ and that is reflected in the values of the MSE. As noted before, the fact that the variance of the estimators decreases with $n$ but the bias does not, may lead the asymptotic distribution of the estimator to be biased when $n/T$ is large (see Hahn and Newey, 2004). This is a problem that affects other quantile regression estimators for fixed effects models (see, e.g., Kato, Galvão, and Montes-Rojas, 2012), but it may be possible to address it using bias correction methods of the type developed by Hahn and Newey (2004), Dhaene and Jochmans (2015), and Fernández-Val and Weidner (2016). In the particular case of the MM-QR estimator, these methods can be implemented in different ways because we can either correct $\hat{\beta}(\tau)$ directly, or construct a bias-corrected estimator from bias-corrected estimators of $\gamma$ and $q(\tau)$; recall that $\hat{\beta}$ is not biased. Very preliminary simulation results suggest that the bias correction based on the split-panel jackknife of Dhaene and Jochmans (2015) can dramatically reduce the bias without a significant loss of precision. We leave it for further research to investigate the extent of the potential bias in the asymptotic distribution of $\hat{\beta}(\tau)$ when $n/T$ is large, and the effectiveness of the different bias corrections.

5.2. Cross-sectional model with endogeneity

The second set of experiments was designed to study the behavior of the MM-QR estimator in presence of an endogenous explanatory variable. In this case, 10,000 independent cross-sectional data sets were simulated from

$$Y_i = 1 + D_i + (1 + D_i) U_i, \quad i = 1, \ldots, N,$$

with $D_i = \sqrt{((1 - \lambda) C_i + \lambda |U_i|)/2}$, where $0 < \lambda < 1$ is a parameter, $C_i = |\xi_i|$, $\xi_i$ has the same distribution as $U_i$, and again we consider three different distributions for the error: standard normal, $\chi^2(5)$, and $t(5)$; in all cases $U_i$ is standardized to have zero mean and unit variance. In this design $D_i$ is endogenous and $C_i$ is a valid instrument for it. Because of the endogeneity, the distribution of $D_i$ necessarily varies with the distribution of $U_i$;
we also let the distribution of $C_i$ vary with the distribution of $U_i$ so that the strength of the instrument depends only on the parameter $\lambda$. We performed simulations for $N \in \{200, 1000, 5000\}$, $\tau \in \{0.25, 0.75\}$, and $\lambda \in \{0.50, 0.25\}$. We estimate (8) using the MM-QR estimator described in Section 4 and, for comparison, we also estimate the model using the IVQR estimator of Chernozhukov and Hansen (2008).\footnote{This estimator was implemented using a grid search with 20 equally-spaced points between $\pm \left(60/\sqrt{N}\right) \times 100\%$ of the true parameter.}

Table 3 reports the bias and MSE for all the cases in this set of experiments for which $\tau = 0.25$. As before, for brevity, we do not report the results with $\tau = 0.75$ which lead essentially to the same conclusions.

Because both estimators are valid in all cases, there is little to choose between them: the MM-QR generally has smaller MSE, but in general the performance of the estimators is very evenly matched. Again, it is reassuring to verify that the MM-QR estimator performs well even with non-normal errors.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$N = 200$</th>
<th>$N = 1000$</th>
<th>$N = 5000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IVQR</td>
<td>MM-QR</td>
<td>IVQR</td>
</tr>
<tr>
<td>Model 1: $\mathcal{N}(0, 1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 BIAS</td>
<td>0.093</td>
<td>0.118</td>
<td>0.017</td>
</tr>
<tr>
<td>MSE</td>
<td>0.384</td>
<td>0.409</td>
<td>0.074</td>
</tr>
<tr>
<td>0.25 BIAS</td>
<td>0.040</td>
<td>0.044</td>
<td>0.007</td>
</tr>
<tr>
<td>MSE</td>
<td>0.209</td>
<td>0.162</td>
<td>0.040</td>
</tr>
<tr>
<td>Model 2: $\chi^2_{(5)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 BIAS</td>
<td>0.071</td>
<td>0.081</td>
<td>0.014</td>
</tr>
<tr>
<td>MSE</td>
<td>0.199</td>
<td>0.155</td>
<td>0.038</td>
</tr>
<tr>
<td>0.25 BIAS</td>
<td>0.033</td>
<td>0.040</td>
<td>0.007</td>
</tr>
<tr>
<td>MSE</td>
<td>0.105</td>
<td>0.073</td>
<td>0.020</td>
</tr>
<tr>
<td>Model 3: $t_{(5)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50 BIAS</td>
<td>0.056</td>
<td>0.070</td>
<td>0.015</td>
</tr>
<tr>
<td>MSE</td>
<td>0.322</td>
<td>0.356</td>
<td>0.064</td>
</tr>
<tr>
<td>0.25 BIAS</td>
<td>0.021</td>
<td>0.033</td>
<td>0.005</td>
</tr>
<tr>
<td>MSE</td>
<td>0.172</td>
<td>0.167</td>
<td>0.034</td>
</tr>
</tbody>
</table>
6. ILLUSTRATIVE APPLICATIONS

In this section we present two simple examples to illustrate the application of the methods developed in this paper. To facilitate the comparison of our results with those in the extant literature, we only consider linear specifications of the conditional quantiles. That is, in all cases we assume that $\sigma(\cdot)$ is the identity function.

6.1. The determinants of government surpluses

Persson and Tabellini (2003) study the economic effects of constitutional reforms by looking at the relation between measures of economic performance and countries’ economic, social, cultural, and political characteristics. For this illustration we focus on the determinants of the budget surplus (see Persson and Tabellini, 2003, Ch. 3).

Persson and Tabellini (2003) use data from 1960 to 1998 for 58 countries to estimate the relation between the surplus of the central government in percent of GDP (denoted SPL) and the following set of country characteristics: POLITY, the measure of the quality of democracy developed by Eckstein and Gurr (1975);\textsuperscript{12} LYP, the log of real per capita income; TRADE, the sum of exports and imports of goods and services in percent of GDP; P1564, the percentage of the population between 15 and 64 years of age; P65, the percentage of the population over the age of 65; LSPL, one-year lag of SPL; OILIM, oil prices in US dollars times a dummy variable equal to 1 if the country is a net importer of oil; OILEX, oil prices in US dollars times a dummy variable equal to 1 if the country is a net exporter of oil; YGAP, the output gap. See Persson and Tabellini (2003) for full details on the sources and definition of variables used.

The first row in Table 4 displays the least squares estimates reported by Persson and Tabellini (2003) in column 4 of their Table 3.4, with standard errors clustered by country in parenthesis.\textsuperscript{13} These results suggest that worse democracies (countries with

\textsuperscript{12}Higher values of the index indicate worse democracies.

\textsuperscript{13}Notice that the original data used in the book contained some mistakes; the correct results and the data are available at Guido Tabellini’s web-page: http://faculty.unibocconi.eu/guidotabellini/.
higher values of POLITY) tend to have larger surpluses. The second row in Table 4
displays the estimates of the parameters in $\sigma(\cdot)$, again with clustered standard errors in
parenthesis; as noted above, we assumed that the scale function is linear so as to preserve
the linearity of the quantiles and facilitate the comparison with the estimates obtained
with other methods. These results show that POLITY has effects with opposite signs on
the location and dispersion,\textsuperscript{14} suggesting that the coefficient of POLITY will be larger
for lower conditional quantiles of SPL. That is, increasing the quality of the democracies
reduces the average surplus, but also increases the dispersion of observed surpluses.

Rows 3 to 5 of Table 4 report quantile regression estimates of the same model using
the method proposed by Canay (2011), which treats the fixed effects as location shifts.
Because the model contains a lagged dependent variable, we also estimated the model
using the method proposed by Galvão (2011).\textsuperscript{15} To allow the fixed effects to differ across
quantiles, Galvão’s (2011) estimator was applied to each quantile at the time; these
results are presented in rows 6 to 8. Finally, rows 9 to 11 display the quantile regression
estimates obtained with the MM-QR estimator presented in Section 3.

For most variables, the three quantile regression estimators lead to similar conclu-
sions in terms of the magnitude and significance of the estimates. For example, the
three methods lead to very similar estimates of the coefficient on the LSPL, the lagged
dependent variable.

However, there are also some very important differences between the results obtained
with the different methods, especially between the results obtained imposing that the
fixed effects only affect the location and the results of less restrictive estimators. Indeed,
the results obtained with the Galvão and MM-QR estimators suggest that the effect
of the quality of the democracy is very heterogeneous, being large for countries whose
budget surplus is low relatively to that of countries with similar characteristics, and
negligible for countries with high budget surpluses relatively to that of countries with

\textsuperscript{14}Similar effects are observed for LSPL and OILIM
\textsuperscript{15}We implemented the estimator using a grid search between 0.30 and 0.95 in steps of 0.01, and using
the lag of LSPL as an instrument for it.
similar characteristics. This pattern is in line with what could be expected from the estimates of the scale function, and it is particularly clear in the results of the MM-QR estimator, for which the difference between the estimates for $\tau = 0.25$ and $\tau = 0.75$ is statistically significant at the 5% level. This finding contrasts sharply with the results obtained with Canay’s estimator, which suggest that the effect of the quality of the democracy is essentially the same across the three quartiles, which seems implausible given the estimates of the parameters in the scale function.

Table 4: The determinants of government surpluses

<table>
<thead>
<tr>
<th></th>
<th>POLITY</th>
<th>LYP</th>
<th>TRADE</th>
<th>P1564</th>
<th>PP65</th>
<th>LSPL</th>
<th>OILIM</th>
<th>OILEX</th>
<th>YGAP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.12</td>
<td>-0.72</td>
<td>0.03</td>
<td>0.12</td>
<td>0.03</td>
<td>0.69</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.47)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.07)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Scale</td>
<td>-0.10</td>
<td>-1.02</td>
<td>0.01</td>
<td>0.06</td>
<td>0.14</td>
<td>-0.14</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(1.25)</td>
<td>(0.01)</td>
<td>(0.05)</td>
<td>(0.12)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Canay</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau = 0.25$</td>
<td>0.13</td>
<td>-0.84</td>
<td>0.03</td>
<td>0.15</td>
<td>0.05</td>
<td>0.74</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.51)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\tau = 0.50$</td>
<td>0.10</td>
<td>-0.67</td>
<td>0.03</td>
<td>0.11</td>
<td>0.04</td>
<td>0.70</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.49)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\tau = 0.75$</td>
<td>0.11</td>
<td>-0.76</td>
<td>0.03</td>
<td>0.10</td>
<td>0.04</td>
<td>0.65</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.51)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Galvão</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau = 0.25$</td>
<td>0.15</td>
<td>-0.50</td>
<td>0.03</td>
<td>0.12</td>
<td>0.02</td>
<td>0.76</td>
<td>-0.06</td>
<td>-0.01</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.61)</td>
<td>(0.01)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\tau = 0.50$</td>
<td>0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.08</td>
<td>-0.01</td>
<td>0.71</td>
<td>-0.05</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.32)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\tau = 0.75$</td>
<td>0.06</td>
<td>-0.30</td>
<td>0.02</td>
<td>0.10</td>
<td>0.05</td>
<td>0.65</td>
<td>-0.03</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.59)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>MM-QR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau = 0.25$</td>
<td>0.19</td>
<td>-0.24</td>
<td>0.03</td>
<td>0.09</td>
<td>-0.04</td>
<td>0.76</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.69)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.11)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\tau = 0.50$</td>
<td>0.11</td>
<td>-0.76</td>
<td>0.03</td>
<td>0.12</td>
<td>0.03</td>
<td>0.68</td>
<td>-0.05</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.51)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\tau = 0.75$</td>
<td>0.03</td>
<td>-1.26</td>
<td>0.03</td>
<td>0.15</td>
<td>0.10</td>
<td>0.62</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.89)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

The dependent variable is SPL; all regressions include country fixed effects. Unbalanced panel with 58 countries and 1659 observations. QR standard errors estimated by bootstrap resampling countries.
The time-series in this panel vary in length from 2 to 38 observations and therefore it is proper to be concerned with the validity of estimators that require large $T$. To check the robustness of the results the estimations were repeated using only data for the 55 countries for which there are at least 10 observations; this reduces the total sample size to 1640. The results obtained with all estimators were remarkably insensitive to dropping the shorter series, and essentially the same estimates were obtained with the two samples.

This data set is reasonably small and therefore all estimators are somewhat imprecise. An example of the challenges posed by these data is that the three quartiles estimated using Galvão’s method cross in 14 occasions, with all quartiles crossing each of the other two. In these cases, if valid, the additional structure imposed by the MM-QR estimator can be helpful.

6.2. Returns to training

Chernozhukov and Hansen (2008) use the data studied by Abadie, Angrist, and Imbens (2002) to illustrate the application of their instrumental variable quantile regression (IVQR) estimator. Here we use the same data to illustrate the application of the MM-QR estimator in a situation where one of the explanatory variables of the model is endogenous.

Briefly, these data were obtained from a randomized training experiment performed under the Job Training Partnership Act in which individuals were randomly assigned the offer of training, but had the option to reject it. Because only 60% of those offered training accepted the offer, the actual training is self-selected but the randomly assigned offer provides a credible instrument for it.

The data used by Chernozhukov and Hansen (2008) contains information on 5102 adult males. Besides details on training assignment and actual training status, the data contains information on earnings and on a number of individual characteristics such as
age, education, and ethnic background. Further details on the data are provided in Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2008).

Table 5 reports different estimates of the returns to training at a range of quantiles. As in Chernozhukov and Hansen (2008), for brevity we do not report the estimates of the parameters associated with the controls.

The first row of Table 5 reports the estimates of the returns to training obtained with Koenker and Bassett’s (1978) estimator that ignores the possible endogeneity of the treatment status; these estimates are all positive and statistically significant, suggesting that the training program had a strong positive impact across the distribution, especially in the upper tail.

<table>
<thead>
<tr>
<th></th>
<th>( \tau = .15 )</th>
<th>( \tau = .25 )</th>
<th>( \tau = .50 )</th>
<th>( \tau = .75 )</th>
<th>( \tau = .85 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>QR</td>
<td>1187 (221)</td>
<td>2510 (360)</td>
<td>4420 (639)</td>
<td>4678 (940)</td>
<td>4807 (1059)</td>
</tr>
<tr>
<td>IVQR</td>
<td>-200 (287)</td>
<td>500 (575)</td>
<td>300 (1101)</td>
<td>2700 (1603)</td>
<td>3200 (1535)</td>
</tr>
<tr>
<td>MM-QR</td>
<td>211 (709)</td>
<td>389 (685)</td>
<td>1008 (763)</td>
<td>1972 (1187)</td>
<td>2575 (1525)</td>
</tr>
</tbody>
</table>

5102 observations. Robust standard errors in parenthesis; IVQR and MM-QR standard errors estimated by bootstrap.

This contrasts with the results obtained with the MM-QR estimator, in which the actual training status is instrumented with the assignment indicator. These results show that the effect of the treatment status variable is not statistically significant at the 10% level neither in the location nor in the scale functions; the corresponding estimates are 1331 (p-value: 0.13) and 956 (p-value: 0.13), respectively. These results suggest that the training is unlikely to have had any impact on the lower tail of the distribution and, at best, may have had some impact on the upper tail. This is exactly what we find.

---

16 Standard errors were obtained by bootstrap both for the IVQR and the MM-QR.
17 Indeed, if the location-scale model is adequate, the conditional mean will be a conditional quantile and the slope parameters will be smaller than 1331 in the quantiles below the mean, and larger for the quantiles above.
using Chernozhukov and Hansen’s (2008) estimator, where again actual training status is instrumented with the assignment indicator.\footnote{The estimator was implemented as in Chernozhukov and Hansen (2008).} Indeed, the results in the second row of Table 5 suggest that the program only had an economically and statistically significant impact on the upper tail of the distribution. The results obtained with the MM-QR, displayed in the third row, paint a similar picture, with the effect of training being economically meaningful and statistically significant at the 10\% only for \( \tau = 0.75 \) and \( \tau = 0.85 \).

Considering the precision of the estimates, the MM-QR and IVQR results are reasonably close and effectively lead to the same conclusion: allowing for the possible endogeneity of the treatment status we find that, if anything, the training only had a significant impact on the upper tail of the distribution.

In these linear models, the validity of the MM-QR depends on assumptions that are stronger than those required by the IVQR but, when these assumptions are valid, the MM-QR has some potential advantages. For example, in this sample, the five structural quantile functions estimated by IVQR cross about 200 times, whereas the MM-QR estimator leads to estimates of these functions that necessarily do not cross.\footnote{It is possible to combine the IVQR with the method proposed by Chernozhukov, Fernández-Val, and Galichon (2010) to obtain structural quantile functions that do not cross.} Imposing this restriction, which is necessarily true, may result in efficiency gains and improved small-sample behavior, as documented by Zhao (2000). An additional advantage of the MM-QR estimator is that it is computationally much easier to implement than the IVQR, specially if there are multiple endogenous explanatory variables. Moreover, in contrast with the IVQR estimator, the MM-QR does not require the choice of a bandwidth and of a grid over which to perform a search, and can easily be used to estimate non-linear models.
7. CONCLUSIONS

In a conditional location-scale model, the information provided by the conditional mean and the conditional scale function is equivalent to the information provided by regression quantiles in the sense that these functions completely characterize how the regressors affect the conditional distribution. This is the result we use to estimate quantiles from estimates of the conditional mean and of the conditional scale function. Naturally, our approach is more restrictive than the traditional quantile regression, but we believe that the additional structure we impose can be useful in many applied settings. In particular, our approach provides an easy way to estimate regression quantiles in situations where using the traditional approach that is difficult or impossible.

The two very different applications we present illustrate that our method leads essentially to the same conclusions that are obtained with methods that are computationally much more demanding. This suggests that the proposed estimator can, at least, be useful in an exploratory phase, for example to provide starting values for other methods and to guide in the choice of the limits of the grid searches used in the Chernozhukov and Hansen (2008) and Galvão (2011) estimators.

Even when the effects of the regressors on the distribution of interest are not limited to their effects on the location and scale functions, i.e., when the location-scale model is inadequate, making a serious effort to model the heteroskedasticity can still be useful in applied work. Heteroskedasticity is often viewed as a nuisance, or interesting only inasmuch as knowledge of it can be used to improve the estimation of the conditional mean (see, e.g., Leamer, 2010, and Romano and Wolf, 2017).\textsuperscript{20} However, the specification and estimation of the scale function is a simple and convenient way of gaining information on how the regressors affect features of the conditional distribution of interest other than its location. When the location-scale model not appropriate, the information that can be obtained from the location and scale functions is not as rich as that provided by

\textsuperscript{20}Of course, heteroskedasticity can also be of interest in itself; the literature on ARCH/GARCH models is a leading example of that (see, e.g., Engle, 2001).
conditional quantiles, but may be interesting in itself, especially when estimation of conditional quantiles is not practical.

There are a number of aspects of the proposed approach that would be interesting to investigate. In the present paper we do not study the performance of the covariance matrix estimator and the quality of the inference based on it. As mentioned before, in a panel data context, standard confidence intervals may not be correctly centered when $n/T$ is large and therefore it would be interesting to study the severity of this problem and the performance of bias-corrected estimators. It would also be useful to develop and study simple tests for the assumption that the location-scale model is adequate in the sense that the effects of the regressors on the distribution of interest are limited to their effects on the location and scale functions. In Section 2 we suggested that such tests can be constructed as tests for overidentifying restrictions, but it may be possible to develop simpler regression-based procedures. Finally, it would naturally be interesting to see if in other applications the results obtained with the proposed method are also similar to those obtained with computationally more demanding estimators, as was the case in the applications we considered.

APPENDIX

A1. Assumptions

The results in the paper were derived under the following assumptions.

**(P): On the parameter space**

1. $(\alpha_i, \delta_i)_{i=1}^n \in \Theta_1$, $(\beta, \gamma) \in \Theta_2$, where $\Theta_1$ and $\Theta_2$ are compact subsets of $\mathbb{R}^{2n}$ and $\mathbb{R}^{2k}$, respectively.

2. The true parameter values are interior points of $\Theta_1$ and $\Theta_2$.

3. Let $F_U$ be the c.d.f. of $U$ satisfying (U1) below and $F_U^{-1}$ its inverse. $\tau \in T = (\epsilon, 1 - \epsilon)$, for some $\epsilon > 0$. The interval $(\lim_{\tau \searrow \epsilon} q(\tau); \lim_{\tau \nearrow (1-\epsilon)} q(\tau))$ is bounded.
(U): On the error term

(1) The random variables $U_{it}$ are i.i.d. (across $i$ and $t$) and independent of $X_{it}$ and $Z_{it}$.

(2) The random variables $U_{it}$ have a continuous density function $f_U$ and $f_U(u) > \zeta > 0$, $\forall u \in \text{supp}(U)$.

(3) $E|U|^{2+\nu} < \infty$ for some $\nu > 0$.

(XZ): On the regressors

(1) The sequence of random $k$-vectors $\{X_{it}\}$ is i.i.d. for any fixed $i$ and independent across $t$.

(2) $Z_{it}$ is a random $k$-vector defined by $Z_{itl} = Z_l(X_{it})$, for $l = 1, \ldots, k$, where $Z_l: \mathbb{R}^k \to \mathbb{R}$ is a known function of class $C^1$ for a.e.-$X$. ($Z_{itl}$ denotes the $l$-th coordinate of the vector $Z_{it}$.)

(3) $\max_{i \leq n} E|X_{i1l}|^{2+\nu} < K < \infty$ for some $K$ and $\nu > 0$, for $l = 1, \ldots, k$. ($X_{i1l}$ denotes the $l$-th coordinate of the vector $X_{i1}$.)

(4) $\max_{i \leq n} E|Z_{i1l}|^{4+\nu} < K < \infty$ for some $K$ and $\nu > 0$, for $l = 1, \ldots, k$.

(5) $(1/n) \sum_i E[(X_{i1} - \bar{X}_i)(X_{i1} - \bar{X}_i)']$ is uniformly p.d. and as a constant limit $Q_{XX}$.

(6) $(1/n) \sum_i E[(Z_{i1} - \bar{Z}_i)(Z_{i1} - \bar{Z}_i)']$ is uniformly p.d. and has a constant limit $Q_{ZZ}$.

(7) $(1/nT) \sum_{it} \pi^2_{it} \sigma^2_{it} = O_P(1)$, with $\pi_{iT} = (1/T) \sum_i (1/\sigma_{it})$.

(8) $(1/n) \sum_i E[(1/\sigma_{i1})X_{i1}] = O(1)$.

---

Assumption (U2) implies that the c.d.f. $F_U$ is strictly monotone and therefore that the quantiles $q(\tau), \tau \in \mathcal{T}$ are unique.

Assumption (U3) implies that $E|V|^{2+\nu} = 2U[I\{U \geq 0\} - P\{U \geq 0\}] - 1$, is also finite.
(9) $\max_{i\leq n} E|Z_{i1a}Z_{i1b}X_{i1c}|^{2+\nu} < K < \infty$ and $\max_{i\leq n} E|Z_{i1a}X_{i1c}X_{i1d}|^{2+\nu} < K < \infty$ for some $K$ and $\nu > 0$, for $a, b, c, d = 1, \ldots, k$.\textsuperscript{23}

(10) The matrices $(1/n) \sum_{i} E[\sigma_{i1}^2 (X_{i1} - \bar{X}_i)(X_{i1} - \bar{X}_i)']$ and $(1/n) \sum_{i} E[\sigma_{i1}^2 (Z_{i1} - \bar{Z}_i)(X_{i1} - \bar{X}_i)']$ have constant limits denoted by $P_{XX}$ and $P_{XZ}$, respectively.

**(DC): On the regressors and instruments**

(1) $E[|D_l|^{2+\nu}] < K < \infty$ for some $K$ and $\nu > 0$, for $l = 1, \ldots, k_D$. ($D_l$ denotes the $l$-th coordinate of the vector $D$.)

(2) $E[|C_1l|^{4+\nu}] < K < \infty$ ($l = 1, \ldots, k_1$) and $E[|C_2l|^{2+\nu}] < K < \infty$ ($l = 1, \ldots, k_2$) for some $K$ and $\nu > 0$.

(3) $E[|\sigma'(X'\gamma)|^{2+\nu}] < K < \infty$ and $E[1/(|\sigma(X'\gamma)|^{2+\nu})] < K < \infty$.

(4) $E[CC']$ is non-singular.

(5) $E[(\sigma'/\sigma)U|CX'] - (E[(1/\sigma)\text{sign}(U)CX'])(E[(1/\sigma)CX'])^{-1}(E[(\sigma'/\sigma)UCX'])$ and $E[(1/\sigma)CX']$ are non-singular.

**A2. Proofs**

For simplicity, the proofs of the theorems will be decomposed into a series of partial results (lemmata). Some are merely instrumental others may be of interest on their own. For economy of space we generally will not refer to any of the assumption above in the

\textsuperscript{23}Applying Minkovski’s inequality it easy to see that this assumption implies that the $(2 + \nu)$-th absolute moments of $\sigma_{it}Z_{it}X'_{it}$ and $\sigma_{it}X_{it}X'_{it}$ exist and are uniformly bounded.
statement of these results. In this appendix we will use the following notation

\[ \Delta_{1i} = \Delta_{1n,T} = \sqrt{T} (\hat{\alpha}_i - \alpha_i) \]
\[ \Delta_2 = \Delta_{2n,T} = \sqrt{nT} (\hat{\beta} - \beta) \]
\[ \Delta_{3i} = \Delta_{3n,T} = \sqrt{T} (\hat{\delta}_i - \delta_i) \]
\[ \Delta_4 = \Delta_{4n,T} = \sqrt{nT} (\hat{\gamma} - \gamma) \]
\[ \Delta_5 = \Delta_{5n,T} = \sqrt{nT} (\hat{q} - q(\tau)) \]

\[ \text{Lemma 1} \quad \text{If } n/T \to 0 \text{ as } \{n, T\} \to \infty, \]
\[ \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_i| = o_P(1). \]

\[ \text{Proof:} \quad \text{Standard least squares results show that} \]
\[ \Delta_{2n,T} = Q_{XX}^{-1} \frac{1}{\sqrt{nT}} \sum_{i,t} \sigma_{it} (X_{it} - \bar{X}_i) U_{it} + o_P(1) \]

with \( \bar{X}_i = (1/T) \sum_t X_{it} \), and

\[ \Delta_{1n,T} = -\frac{1}{\sqrt{n}} \bar{X}'_i \Delta_{2n,T} + \frac{1}{\sqrt{T}} \sum_t \sigma_{it} U_{it} = \frac{1}{\sqrt{T}} \sum_t \sigma_{it} U_{it} + o_P(1). \]

For any \( n \) and \( T \), \( E[\hat{\alpha}_i - \alpha_i] = 0 \) and \( V(\hat{\beta} - \beta) = O_P(1/nT) \).

\[ V(\hat{\alpha}_i - \alpha_i) = \bar{X}'_i V(\hat{\beta} - \beta) \bar{X}_i + \frac{E[U^2]}{T^2} \sum_t \sigma_{it}^2 = O_P(1/nT) + O_P(1/T) = O_P(1/T). \]

Consider now \( \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_i| \).

\[ P\{ \max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_i| > \epsilon \} \leq \sum_{i=1}^n P\{|\hat{\alpha}_i - \alpha_i| > \epsilon \} \leq \frac{1}{\epsilon^2} \sum_i V(\hat{\alpha}_i - \alpha_i) \leq \frac{1}{T \epsilon^2} \left( \frac{1}{n} \sum_i \bar{X}'_i V(\Delta_{2n,T}) \bar{X}_i \right) + \frac{E[U^2]}{\epsilon^2 T} \left( \frac{1}{nT} \sum_t \sum_i \sigma_{it}^2 \right) \leq O(1/T) + \frac{n}{T} O(1) = o(1) \text{ if } n/T \to 0. \]
Lemma 2  Let $\hat{R}_{it} = Y_{it} - \hat{\alpha}_i - X_{it}\hat{\beta}$ and $\eta = E(I\{U > 0\})$. Then, as $\{n,T\} \to \infty$ with $n = o(T)$,

$$\frac{1}{\sqrt{T}} \sum_t [\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}] - \frac{1}{\sqrt{T}} \sum_t \sigma_{it}[U_{it}(I\{U_{it} > 0\} - \eta) - 1] = o_P(1) \ (i = 1, \ldots, n),$$

and

$$\frac{1}{\sqrt{nT}} \sum_{it} Z_{it}[\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}] - \frac{1}{\sqrt{nT}} \sum_{it} Z_{it}\sigma_{it}[U_{it}(I\{U_{it} > 0\} - \eta) - 1] = o_P(1).$$

Proof: Put,

$$L_{n,T}(X_{it}, \Delta) = (1/\sqrt{T})\Delta_{1i} + (1/\sqrt{nT})X_{it}'\Delta_2,$$

and

$$M_{2n,T}(\Delta) = \frac{1}{\sqrt{nT}} \sum_{it} Z_{it}[\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}]$$

$$= \frac{1}{\sqrt{nT}} \sum_{it} Z_{it}((\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta)) \times$$

$$\times (I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \eta) - \sigma_{it}]$$

($\Delta = ((\Delta_{1i})^n, \Delta_2)$) and

$$\tilde{M}_{2n,T}(\Delta) = M_{2n,T}(\Delta) - E[M_{2n,T}(\Delta)].$$

We will first prove the stochastic equicontinuity of the empirical process $\tilde{M}_{2n,T}(\cdot)$. The proof will follow Andrews (1994). The function

$$m(U, Z, X, \delta_i, \gamma, \Delta) = [\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta)][I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \eta]$$

is of CV-type 1 with envelope

$$\sup_{\delta_i, \gamma, \Delta} m(U, Z, X, \delta_i, \gamma, \Delta) = c_1 + c_2|U| + c_3||Z|||U| + c_4||X||$$

for some constants $c_j$. Pollard’s entropy condition (Andrews, 1994, section 4.2) is satisfied if

$$\lim_{n \to \infty} (1/n) \sum_i (E[||Z_{i1}||^{2+\nu} + 1) \sup_{\delta_i, \gamma, \Delta} \|m(U, Z, X, \delta_i, \gamma, \Delta)||^{2+\nu} < \infty.$$
It suffices that,

$$
\lim_{n \to \infty} \left\{ E|U|^{2+\nu} + \frac{1}{n} \sum_{i} [E\|Z_{ii}\|^{4+\nu} + E|U|^{2+\nu}E\|X_{ii}\|^{2+\nu} + E|U|^{2+\nu}E\|Z_{ii}\|^{2+\nu} + E|U|^{2+\nu}E\|X_{ii}\|^{2+\nu} + E|U|^{2+\nu}E\|X_{ii}\|^{2+\nu}] \right\} < \infty.
$$

Assumption (U3), (XZ3) and (XZ8) yield the desired result and prove the stochastic equicontinuity of $\tilde{M}_{2n,T}(\cdot)$.

Stochastic equicontinuity and the fact that $\max_{i} |(1/\sqrt{T})\Delta_{1i}| = o_{P}(1)$ and $(1/\sqrt{nT})\Delta_{2} = o_{P}(1)$ imply (Andrews, 1994, p. 2265) that

$$
\tilde{M}_{2n,T}(\Delta) - \tilde{M}_{n,T}(0) = o_{P}(1).
$$

Consequently,

$$
M_{2n,T}(\Delta) = E[M_{2n,T}(\Delta)] + \tilde{M}_{2n,T}(0) + [M_{2n,T}(\Delta) - \tilde{M}_{2n,T}(0)]
$$

$$
= E[M_{2n,T}(\Delta)] + \tilde{M}_{2n,T}(0) + o_{P}(1)
$$

$$
= E[M_{2n,T}(\Delta)] + M_{2n,T}(0)
$$

since $E[M_{2n,T}(0)] = 0$.

The lemma is proved as a first-order Taylor series expansion of $E[M_{n,T}(\Delta)]$ around $\Delta = 0$ yields

$$
E[M_{2n,T}(\Delta)] = -E[I\{U_{it} > 0\} - \eta] \frac{1}{\sqrt{nT}} \sum_{it} E[L_{n,T}(X_{it}, \Delta)] = 0.
$$

Now put

$$
M_{1n,T}(\Delta) = \frac{1}{\sqrt{T}} \sum_{t} [\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}].
$$

The same arguments yield

$$
M_{1n,T}(\Delta) = M_{1n,T}(0) + o_{P}(1).
$$

**Lemma 3** Let,

$$
\hat{\eta} = \frac{1}{nT} \sum_{it} I\{\hat{R}_{it} > 0\}.
$$
Then,
\[ \sqrt{nT}(\hat{\eta} - \eta) = O_P(1) \quad \text{as} \quad \{n, T\} \to \infty \quad \text{with} \quad n = o(T). \]

**Proof:** Using the notation of lemma 2, let,
\[ \tilde{R}_{n,T}(U, X, \Delta) = \frac{1}{nT} \sum_{it} I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \frac{1}{nT} \sum_{it} E[I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\}] \]

The process \( \tilde{R}_{n,T}(\cdot) \) satisfies trivially Pollard’s entropy condition and so it is equicontinuous (see Andrews, 1994, p. 2273). Since \( \max_i |(1/\sqrt{T})\Delta_{1i}| = o_P(1) \) and \((1/\sqrt{nT})\Delta_{2} = o_P(1)\)
\[ \tilde{R}_{n,T}(\cdot, \Delta) = \tilde{R}_{n,T}(\cdot, 0) = O_P(1) \]
since \( \tilde{R}_{n,T}(\cdot, 0) = o_P(1) \) by the law of large numbers. Now, a Taylor series expansion yields,
\[ \sqrt{nT}(\hat{\eta} - \eta) = -\frac{f_U(0)}{\sqrt{nT}} \sum_{it} \frac{1}{\sigma_{it}} \left[ \frac{1}{\sqrt{T}} \Delta_{1i} + \frac{1}{\sqrt{nT}} X'_{it} \Delta_{2} \right] + o_P(1), \]
which establishes the result for \((1/nT)\sum_{it}(1/\sigma_{it})X_{it} = O_P(1)\) and (see assumption (XZ7))
\[ \frac{1}{\sqrt{nT}} \sum_{it} \frac{1}{\sigma_{it}} \frac{1}{\sqrt{T}} \Delta_{1i} = \frac{1}{\sqrt{nT}} \sum_{it} \frac{1}{\sigma_{it}} \frac{1}{\sqrt{T}} \left[ -\frac{1}{\sqrt{n}} \bar{X}'_{it} \Delta_{2_{n,T}} + \frac{1}{\sqrt{T}} \sum_{t} \sigma_{it} U_{it} \right] \]
\[ = O_P(1) + \frac{1}{\sqrt{nT}} \sum_{it} \pi_{i,T} \sigma_{it} U_{it} \]
\[ = O_P(1), \]
where the last equality follows from applying the central limit theorem. \( \blacksquare \)

**Proof (Theorem 1):** The the moment conditions defining the estimators of \( \delta_i \) \( (i = 1, \ldots, n) \) and \( \gamma \) are,
\[ \frac{1}{\sqrt{T}} \sum_{t} \left\{ \left[ \hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \hat{\eta}) - \sigma_{it} \right] - \frac{1}{\sqrt{T}} \Delta_{3i} - \frac{1}{\sqrt{nT}} Z'_{it} \Delta_{4} \right\} = 0 \]
\[ \frac{1}{\sqrt{nT}} \sum_{it} Z_{it} \left\{ \left[ \hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \hat{\eta}) - \sigma_{it} \right] - \frac{1}{\sqrt{T}} \Delta_{3i} - \frac{1}{\sqrt{nT}} Z'_{it} \Delta_{4} \right\} = 0, \]
which can be written as

\[
G_n \left( \frac{\Delta_3}{\Delta_4} \right) = \left( \begin{array}{c}
M_{1,n,T}(0) \\
M_{2,n,T}(0)
\end{array} \right) + (\tilde{\eta} - \eta) \left( \frac{1}{\sqrt{nT}} \sum_{it} Z_{it} \sigma_{it} U_{it} - L_{n,T}(X_{it}, \Delta_1, \Delta_2) \right)
\]

with

\[
G_n = \left( \begin{array}{c}
\frac{1}{(1/\sqrt{n})} \sum_i Z_i \\
(1/\sqrt{n}) \tilde{Z}_i
\end{array} \right)
\]

where \(\tilde{Z}_i = (1/T) \sum_i Z_{it}\). Lemma 3 implies that the second term on the right-hand side is \(o_P(1)\). Solving the system for \(\Delta_4\) gives,

\[
Q_{zz} \Delta_4 = \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it} (Z_{it} - \tilde{Z}_i) [U_{it} (I \{U_{it} > 0\} - \eta) - 1].
\]

The central limit theorem establishes the desired result.

**Lemma 4** If \(n/T \to 0\) as \(\{n, T\} \to \infty\),

\[
\max_{1 \leq i \leq n} |\hat{\delta}_i - \delta_i| = o_P(1).
\]

**Proof:** The first equation of the system in the proof of Theorem 1 implies that (adopting the same notation)

\[
\frac{1}{\sqrt{1}} \Delta_3 = \frac{1}{\sqrt{T}} M_{1,n,T}(0) - \frac{1}{\sqrt{nT}} \tilde{Z}_i \Delta_4.
\]

For any \(\epsilon > 0\),

\[
P \left\{ \max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{1}} \Delta_3 \right| > \epsilon \right\} \leq \sum_i P \left\{ \left| \frac{1}{\sqrt{T}} M_{1,n,T}(0) \right| > \frac{\epsilon}{2} \right\} + \\
+ \sum_i P \left\{ \left| \frac{1}{\sqrt{nT}} Z_i \Delta_4 \right| > \frac{\epsilon}{2} \right\} \\
\leq \frac{2E[V^2]}{\epsilon^2} \frac{1}{T} \left( \frac{1}{nT} \sum_{it} E[\sigma_{it}^2] \right) + \\
+ \frac{1}{T} \left( \frac{1}{T} \sum_{it} E[\tilde{Z}_i E[\Delta_4 \Delta_4] \tilde{Z}_i] \right) \\
= \frac{n}{T} O(1) + O(T^{-1}).
\]

36
Lemma 5 Let $\tilde{q}$ solve $\min q S_{n,T}(q, \Delta) = (1/nT) \sum_{it} \rho_r(\hat{R}_{it} - q \hat{\sigma}_{it})$. (With $\hat{\sigma}_{it} = \hat{\delta}_i + Z_{it} \hat{\gamma}$ and $\Delta = ((\Delta_1)_1^n, \Delta_2, (\Delta_3)_1^n, \Delta_4$). Then, as $\{n, T\} \to \infty$ with $n = o(T)$,

$$\tilde{q} \xrightarrow{P} q(\tau).$$

Proof: By well-known arguments is suffices to show that

$$S_{n,T}(q, \Delta) \xrightarrow{P} E[S_{n,T}(q, 0)].$$

The compactness of the parameter space (or the convexity of $\rho_r$) implies that the convergence is uniform in $q$. The law of large numbers implies that

$$S_{n,T}(q, \Delta) \xrightarrow{P} E[S_{n,T}(q, \Delta)].$$

But, with $E[\psi_r(w)] = dE[\rho_r(w)]/dw$,

$$E[S_{n,T}(q, \Delta)] = E[S_{n,T}(q, 0)]$$

$$= E[S_{n,T}(q, 0)]$$

$$- E[\psi_r(U - q)] \frac{1}{nT} \sum_{it} \{L_{nT}(X_{it}, (\Delta_1)_1^n, \Delta_2) + L_{nT}(Z_{it}, (\Delta_3)_1^n, \Delta_4)\}$$

where $L_{nT}(\cdot, \cdot)$ was defined in Lemma 2. The result is established noting that both $(1/nT) \sum_{it} \{L_{nT}(X_{it}, (\Delta_1)_1^n, \Delta_2)$ and $(1/nT) \sum_{it} \{L_{nT}(Z_{it}, (\Delta_3)_1^n, \Delta_4)$ are $O_P(1/nT)$.

Proof (Theorem 2): Let $\psi_r(A) = -(I\{A \leq 0\} - \tau), \Delta = ((\Delta_1)_1^n, \Delta_2, (\Delta_3)_1^n, \Delta_4, \Delta_5)$,

$$\Psi_{n,T}(U, X, Z, \Delta) = \frac{1}{\sqrt{nT}} \sum_{it} \hat{\sigma}_{it} \psi_r \left[ \hat{R}_{it} - q \hat{\sigma}_{it} \right]$$

$$= \frac{1}{\sqrt{nT}} \sum_{it} \left\{ \sigma_{it} + L_{nT}(Z_{it}, (\Delta_3)_1^n, \Delta_4) \right\} \psi_r(\sigma_{it} U_{it} - L_{nT}(X_{it}, (\Delta_1)_1^n, \Delta_2))$$

$$- \left(q(\tau) - \frac{1}{\sqrt{nT}} \Delta_5 \right) (\sigma_{it} + L_{nT}(Z_{it}, (\Delta_3)_1^n, \Delta_4))$$

$$= o_P(1)$$

and

$$\tilde{\Psi}_{n,T}(U, X, Z, \Delta) = \Psi_{n,T}(U, X, Z, \Delta) - E[\Psi_{n,T}(U, X, Z, \Delta)].$$
The boundedness of $\psi_t(\cdot)$ and the moment conditions suffice to yield the stochastic equicontinuity of $\tilde{\Psi}_{n,T}(U, X, Z, \Delta)$. As, $\max_i |(1/\sqrt{T})\Delta_{1i}|, (1/\sqrt{nT})\Delta_2, \max_i |(1/\sqrt{T})\Delta_{3i}|, (1/\sqrt{nT})\Delta_4,$ and $(1/\sqrt{nT})\Delta_5$ are all $o_P(1)$ as \( \{n,T\} \to \infty \) with $n/T \to 0$,

$$
\tilde{\Psi}_{n,T}(U, X, Z, \Delta) - \tilde{\Psi}_{n,T}(U, X, Z, 0) = o_P(1).
$$

Consequently (note that $E[\Psi_{n,T}(U, X, Z, 0)] = 0$),

$$
\Psi_{n,T}(U, X, Z, \Delta) = E[\Psi_{n,T}(U, X, Z, \Delta)] + \Psi_{n,T}(U, X, Z, 0) + o_P(1).
$$

The first term on the right-hand side can be approximated to the first order around $\Delta = 0$ by

$$
E[\Psi_{n,T}(U, X, Z, \Delta)] = -f_U(q(\tau)) \left\{ \frac{1}{\sqrt{nT}} \sum_{it} L_{nT}(X_{it}, (\Delta_{1i})_1^n, \Delta_2) + q(\tau) L_{nT}(Z_{it}, (\Delta_{3i})_1^n, \Delta_4) + \frac{1}{nT} \sum_{it} \sigma_{it}\Delta_5 \right\}
$$

The second term

$$
\Psi_{n,T}(U, X, Z, 0) = \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}\psi_t(U_{it} - q(\tau))
$$

is an asymptotically normal sequence. Putting the two terms together,

$$
\sqrt{n}\Delta_1 + \tilde{X}'\Delta_2 + q(\tau)[\sqrt{n}\Delta_3 + \tilde{Z}'\Delta_4] = \frac{1}{f_U(q(\tau))} \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}\psi_t(U_{it} - q(\tau)) + o_P(1)
$$

with $\Delta_1 = (1/n) \sum_i \Delta_{1i}$ and $\tilde{X} = (1/nT) \sum_{it} X_{it}$ (and likewise for $\Delta_3$ and $\tilde{Z}$). Note that,

$$
\sqrt{n}\tilde{X}'\Delta_2 = \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}U_{it}
$$

and

$$
\sqrt{n}\tilde{Z}'\Delta_4 = \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}V_{it}.
$$

Consequently,

$$
\mu_5\Delta_5 = \frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it} \left[ \frac{1}{f_U(q(\tau))} \psi_t(U_{it} - q(\tau)) + U_{it} + q(\tau)V_{it} \right].
$$
Combining this result with the representation of $\Delta_4$ in the proof of Theorem 1 and with the usual representation of the least squares estimator $\Delta_2$ gives,

$$
\begin{pmatrix}
Q_{XX} & O & O \\
O & Q_{ZZ} & O \\
0' & 0' & \mu_o
\end{pmatrix}
\begin{pmatrix}
\Delta_2 \\
\Delta_4 \\
\Delta_5
\end{pmatrix} = 
\begin{pmatrix}
\frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}(X_{it} - \bar{X}_i)U_{it} \\
\frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}(Z_{it} - \bar{Z}_i)V_{it} \\
\frac{1}{\sqrt{nT}} \sum_{it} \sigma_{it}W_{it}
\end{pmatrix}.
$$

where $O$ and 0 denote a $k \times k$ matrix and $k$-vector of 0s, respectively. The result then follows from the central limit theorem and the delta-method.

**Proof (Theorem 3):** Put $\Delta_1 = \sqrt{n}(\hat{\beta} - \beta)$, $\Delta_2 = \sqrt{n}(\hat{\gamma} - \gamma)$, $\Delta_3 = \sqrt{n}(\hat{q} - q(\tau))$.

Standard GMM arguments (Newey and McFadden, 1994) or, for $\hat{q}$, arguments as in Lemma 5, prove the consistency of $(\hat{\beta}, \hat{\gamma}, \hat{q})$.

Let us start with the linear representation of $\Delta_3$ conditional on root-$n$ consistent estimators of $\beta$ and $\gamma$,

$$
\frac{Y - X'\hat{\beta}}{\sigma(X'\hat{\gamma})} - \hat{q} = (U - q(\tau)) - L_n(U, X, \Delta_1, \Delta_2) - \frac{1}{\sqrt{n}} \Delta_3 - K_n(U, X, \Delta_1, \Delta_2),
$$

where

$$
L_n(U, X, \Delta_1, \Delta_2) = \frac{1}{\sigma} \frac{1}{\sqrt{n}} X' \Delta_1 + \frac{1}{\sigma} \frac{1}{\sqrt{n}} U X' \Delta_2,
$$

and

$$
K_n(U, X, \Delta_1, \Delta_2) = \frac{\sigma'}{\sigma} \frac{1}{n} (X' \Delta_1)(X' \Delta_2).
$$

The moments conditions (DC) ensure that

$$
\frac{1}{\sqrt{n}} \sum_i K_n(U_i, X_i, \Delta_1, \Delta_2) = \frac{1}{\sqrt{n}} \Delta_1 \left( \frac{1}{n} \sum_i \frac{\sigma_i'}{\sigma_i} X_i X_i' \right) \Delta_2 = o_P(1).
$$

Using the stochastic equicontinuity arguments in the proof of Theorem 2, $\psi_\tau[(U - q(\tau)) - L_n(U, X, \Delta_1, \Delta_2) - 1/\sqrt{n}\Delta_3]$ can by expanded around $\Delta = 0$ to yield,

$$
\Delta_3 + \left( \frac{1}{n} \sum_i \frac{1}{\sigma_i} X_i' \right) \Delta_1 + \left( \frac{1}{n} \sum_i \frac{\sigma_i'}{\sigma_i} U_i X_i' \right) \Delta_2 = -\frac{1}{f_U(q(\tau))} \frac{1}{\sqrt{n}} \sum_i \psi_\tau(U_i - q(\tau)) + o_P(1).
$$

39
Consider now $\Delta_1$ and $\Delta_2$. The moment conditions can be written as,

\[
M_{1,n}(U, X, \Delta) = \frac{1}{\sqrt{n}} \sum_i C_i(\hat{R}_i/\hat{\sigma}) \\
= \frac{1}{\sqrt{n}} \sum_i C_i(U_i - L_n(U_i, X_i, \Delta_1, \Delta_2)) \\
= o_P(1),
\]

and

\[
M_{2,n}(U, X, \Delta) = \frac{1}{\sqrt{n}} \sum_i C_i(|\hat{R}_i|/\hat{\sigma}) \\
= \frac{1}{\sqrt{n}} \sum_i 2C_i[U_i - L_n(U_i, X_i, \Delta_1, \Delta_2)] \times \\
\left[ 1/2 - I \left\{ U_i \leq \frac{1}{\sqrt{n}} \frac{1}{\sigma_i} X'_i \Delta_1 \right\} \right] \\
= o_P(1),
\]

As in the proof of lemma 2, the moment conditions ensure the stochastic equicontinuity of \( \{M_{2,n}(U, X, \Delta) - E[M_{2,n}(U, X, \Delta)]\} \). Together with the consistency of $\Delta$ and the fact that $E[M_{2,n}(U, X, 0)] = 0$, this allows to write,

\[
M_{2,n}(U, X, \Delta) = E[M_{2,n}(U, X, \Delta)] + M_{2,n}(U, X, 0).
\]

The linear representation is completed by noting the first-order Taylor series expansion of $E[M_{2,n}(U, X, \Delta)]$ around $\Delta = 0$,

\[
E[M_{2,n}(U, X, \Delta)] = E[(1/\sigma) \text{sign}(U) CX']\Delta_1 + E[(\sigma'/\sigma) |U| CX']\Delta_2 + o(1).
\]

A final remark about the non-singularity of $G$. It suffices to show that

\[
\begin{pmatrix}
E[(1/\sigma) CX'] & E[(\sigma'/\sigma) UCX'] \\
E[(1/\sigma) \text{sign}(U) CX'] & E[(\sigma'/\sigma) |U| CX']
\end{pmatrix}
\]

is non-singular which is ensured by (DC5).
REFERENCES


StataCorp. (2017). *Stata Statistical Software: Release 15*, College Station: TX: StataCorp LLC.

