Quantile Regression for Dynamic Panel Data with Fixed Effects

Antonio F. Galvao, Jr.*

July, 2009

Abstract

This paper studies estimation and inference in a quantile regression dynamic panel model with fixed effects. Panel data fixed effects estimators are typically biased in the presence of lagged dependent variables as regressors. To reduce the dynamic bias in the quantile regression fixed effects estimator we suggest the use of the instrumental variables quantile regression method of Chernozhukov and Hansen (2006, 2008) along with lagged regressors as instruments. We show that the instrumental variables estimator is consistent and asymptotically normal. We briefly describe how to employ the estimated models for prediction. In addition, Wald and Kolmogorov-Smirnov type tests for general linear restrictions are proposed. Monte Carlo studies are conducted to evaluate the finite sample properties of the estimators and tests. The simulation results show that the instrumental variables approach sharply reduces the dynamic bias, and turns out to be especially advantageous when innovations are non-Gaussian and heavy-tailed. Finally, we illustrate the procedures by testing for the presence of time non-separability in utility using household consumption panel data. The results show evidence of asymmetric persistence in consumption dynamics, and important heterogeneity in the determinants of consumption.

Key Words: Quantile regression, dynamic panel, fixed effects, instrumental variables

JEL Classification: C14, C23, D12, E21

*University of Wisconsin-Milwaukee, Department of Economics, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201, US. E-mail: agalvao@uwm.edu. This paper is based on the first chapter of my doctoral thesis. I am deeply grateful to Roger Koenker for his continuing guidance and encouragement. I am also grateful to Anil Bera, Xuming He, Ted Juhl, Steve Portnoy, Shinichi Sakata, participants in the seminars at Forecasting in Rio, 2008 North American Summer Meeting of the Econometric Society, City University London, 17th Midwest Econometrics Group Meeting, and University of Illinois at Urbana-Champaign for helpful comments and discussions. All the remaining errors are my own.

1 Introduction

Recently, there has been a growing literature on estimation and testing of dynamic panel data models. Consistency of estimators in conventional dynamic panel data models depends critically on the assumptions about the initial conditions of the dynamic process. Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991) have shown that instrumental variables methods are able to produce consistent estimators that are independent of the initial conditions. This paper investigates estimation and inference in a quantile regression formulation of the dynamic panel data model with individual specific intercepts. We find that conventional fixed effects estimation of the quantile regression specification suffers from similar bias problems to those of the least squares estimation. To reduce the dynamic bias in the quantile regression fixed effects estimator, we suggest the use of the instrumental variables quantile regression method of Chernozhukov and Hansen (2005, 2006, 2008) along with lagged (or lagged differences of the) regressors as instruments. Thus, the estimator combines the usual instrumental variables concept for dynamic panel data and the quantile regression instrumental variables framework. We show that under some mild regularity conditions, notably that with $T \to \infty$ as $N \to \infty$ and $N^a/T \to 0$, for some $a > 0$, the estimator is consistent and asymptotically normal. In addition, we briefly describe how to employ the estimated models for prediction. We also propose Wald and Kolmogorov-Smirnov tests for general linear hypotheses, and derive the associated limiting distributions. Monte Carlo experiments show that, even in short panels, the instrumental variables estimator can substantially reduce the dynamic bias. Finally, we illustrate the new approach by testing for the presence of time non-separability in utility using household consumption panel data. The results show evidence of asymmetric persistence in consumption dynamics.

Koenker (2004) introduced a general approach to estimation of quantile regression models for longitudinal data. Individual specific (fixed) effects are treated as pure location shift parameters common to all conditional quantiles and may be subject to shrinkage toward a common value as in the Gaussian random effects paradigm. Controlling for individual specific heterogeneity via fixed effects while exploring heterogeneous covariate effects within the quantile regression framework offers a more flexible approach to the analysis of panel data than that afforded by the classical Gaussian fixed and random effects estimators. Recent work by Lamarche (2006, 2008) and Geraci and Bottai (2007) have elaborated on this form of penalized quantile regression estimator. Abrevaya and Dahl (2008) have introduced an alternative approach to estimating quantile regression models for panel data employing the "correlated random effects" model of Chamberlain (1982).

In econometric applications the modeling of dynamic relationships and the availability of panel data often suggest dynamic model specifications involving lagged dependent variables. It has been recognized at least since Nickell (1981) that classical least squares estimators in dynamic panel models with fixed effects are seriously biased when the temporal dimension of the panel is short. An extensive literature, initiated by Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991), has explored instrumental variables approaches to attenuate the bias.[1]

Conventional quantile regression estimation of dynamic panel data models with fixed effects suffers from similar bias effects to those seen in the least squares case when T is modest. Reliance on the existing least squares strategies for bias reduction is unsatisfactory in the quantile regression setting for at least two reasons. First, differencing is inappropriate, either temporally, or via the usual deviation from individual means (within) transformation. Linear transformations that are completely innocuous in the context of conditional mean models are highly problematic in the conditional quantile models since they alter in a fundamental way what is being estimated. Expectations enjoy the convenient property in that they commute with linear transformations; quantiles do not.[2] Secondly, the implementation of the instrumental variables method needs to be rethought. Fortunately, neither problem is insurmountable. There is no need to transform the quantile regression model to compute the fixed effects estimator. This is a computable convenience in the least squares case, but even when the number of fixed effects is large, interior point optimization methods using modern sparse linear algebra make direct estimation of the quantile regression model quite efficient. The instrumental variables estimator for quantile regression introduced by Chernozhukov

[1]Ahn and Schmidt (1995) study efficient estimation of models for dynamic panel data using GMM estimator. Blundell and Bond (1998) consider estimation of the dynamic error components model proposing two alternative linear estimators that are designed to improve the properties of the standard first differenced GMM estimator. Bun and Carree (2005) propose a bias corrected estimator for fixed effects dynamic panel data. More recently, a number of additional approaches have been proposed to reduce the bias in dynamic and nonlinear panels. These methods use asymptotic approximations derived as both the number of individuals, N, and the number time series, T, go to infinity jointly; see, for example, Arellano and Hahn (2005) for a survey, and Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Hahn and Newey (2004), Bester and Hansen (2007), for specific approaches.

[2]This intrinsic difficulty has been recognized by Abrevaya and Dahl (2008), among others, and is clarified by Koenker and Hallock (2000, p.19): "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects."

and Hansen (2006, 2008) will be adapted to the dynamic panel data setting and serves as an effective bias reduction device.

Monte Carlo simulations show that the quantile regression fixed effects estimator is significantly biased in the presence of lagged dependent variables, while the instrumental variables method sharply reduces the bias even in short panels. In addition, the Monte Carlo experiments suggest that the quantile regression instrumental variables approach for dynamic panel data performs better than ordinary least squares instrumental variables in terms of bias and root mean squared error for non-Gaussian heavy-tailed distributions. Tests based on the fixed effects quantile regression dynamic panel instrumental variables (PQRIV) turn out to be especially advantageous when innovations are heavy-tailed.

There is an emerging literature on forecasting with panel data, see for instance Baltagi (2008) for a survey. However, less emphasis has been devoted to forecasting with dynamic panel data models. The model proposed in this paper is useful for prediction of conditional quantile functions using dynamic panel data with fixed effects. The quantile regression model has a significant advantage over models based on the conditional mean, since it will be less sensitive to the tail behavior of the underlying random variables representing the forecasting variable of interest, and consequently will be less sensitive to observed outliers. Moreover, because of the heterogeneous nature of most of the variables of interest in economics, prediction using quantile regression techniques is an important tool for applied work. The model proposed in this paper can also be used to predict the conditional density function of the variable of interest under very weak assumptions.

We illustrate the methods by testing for the presence of time-nonseparability in utility using household consumption data from the Panel Study on Income Dynamics (PSID) dataset. A growing body of literature has emphasized the importance of allowing for habit formation as a way of modeling time dependence in preferences in order to improve the predictions of time-separable models. The notion of habit persistence has been also used to address other important issues in macroeconomics and finance, such as: the equity premium puzzle; the excess sensitivity of nondurable consumption; the hump-shaped response of consumption to shocks; and the relationship between savings and growth. Previous research, such as Dynan (2000), find no evidence of habit formation using the PSID dataset and ordinary least squares for estimation at annual frequency. However, quantile regression methods reveal important heterogeneity associated with economic agents' behavior in terms of their dynamic consumption growth, which is averaged out by least squares estimators. At the same time, dynamic panel data quantile regression allows to control for individual specific effects in the dynamic panel data context. Our results show evidence of asymmetric persistence in consumption dynamics in the upper quantiles of the conditional distribution. In addition, it is possible to reject the null hypothesis of no effect of past consumption growth on subsequent consumption growth for these quantiles. Thus, for the upper conditional quantile functions of consumption growth, the results suggest that an increase in current consumption growth leads to increases in subsequent consumption growth, and for these corresponding quantiles there is evidence of habit persistence. Moreover, the results show important evidence of heterogeneity in the determinants of consumption such as the number of adult male equivalents in the household, age of the household head, and race.

The rest of the paper is organized as follows. Section 2 presents the quantile regression dynamic panel data instrumental variables with fixed effects estimation and describes how to employ the estimated models for prediction. Inference is described in Section 3. Section 4 describes the Monte Carlo experiment. In Section 5 we illustrate the new approach using the PSID dataset. Finally, Section 6 concludes the paper.

2 The Model and Assumptions

2.1 Estimation

In this section we introduce estimation of the dynamic panel data quantile regression that includes individual specific fixed effects, and study the asymptotic properties of the estimator.

Consider the classical dynamic model for panel data with individual fixed effects[3]

$$y_{it} = \eta_i + \alpha y_{it-1} + x_{it}'\beta + u_{it} \quad i = 1, ..., N; \quad t = 1, ..., T. \tag{1}$$

where y_{it} is the response variable, η_i denotes the individual fixed effects, y_{it-1} is the lag of the response variable, x_{it} is a p-vector of exogenous covariates, and u_{it} is the innovation term. It is possible to write model (1) in a more concise matrix form as,

$$y = Z\eta + \alpha y_{-1} + X\beta + u, \tag{2}$$

[3]To simplify the presentation we focus on the first-order autoregressive processes, since the main insights generalize in a simple way to higher-order cases.

so that $Z = I_N \otimes \iota_T$, and ι_T is a $T \times 1$ vector of ones. Note that Z represents an incidence matrix that identifies the N distinct individuals in the sample.

The analogous version model to (1) for the τth conditional quantile function of the response of the tth observation on the ith individual y_{it} can be represented as

$$Q_{y_{it}}(\tau | z_{it}, y_{it-1}, x_{it}) = z_{it}\eta + \alpha(\tau) y_{it-1} + x_{it}'\beta(\tau) \tag{3}$$

where y_{it} is the outcome of interest, y_{it-1} is the lag of the variable of interest, x_{it} are exogenous variables, z_{it} identifies the fixed effects, and $\eta = (\eta_1, ..., \eta_N)'$ is the $N \times 1$ vector of individual specific effects or intercepts. In model (3) only the effects of the covariates (y_{it-1}, x_{it}) are allowed to depend upon the quantile, τ, of interest. The η's are intended to capture some individual specific source of variability, or "unobserved heterogeneity," that was not adequately controlled for by other covariates. In most applications the time series dimension T is relatively small compared to the number of individuals N. Therefore, it might be difficult to estimate a τ-dependent distributional individual effect, and we restrict the estimates of the individual specific effects to be independent of τ across the quantiles. We restrict the individual effects, η, to be independent of the specific quantile, τ, by estimating the model for several quantiles simultaneously. Koenker (1984) considered an analogous situation in which only the intercept parameter was permitted to depend upon τ and the slope parameters associated with the included covariates were constrained to be identical for several τ's. In this case the slope parameters were estimated as regression L-statistics. In the present paper, as in Koenker (2004), it is the individual effects, η, that are estimated as discretely weighted L-statistics.

Koenker (2004) introduced a general approach to estimation of quantile regression fixed effects models for panel data. Applying this principle to equation (3), one would solve

$$(\hat{\eta}, \hat{\alpha}, \hat{\beta}) = \min_{\eta,\alpha,\beta} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k \rho_\tau (y_{it} - z_{it}\eta - \alpha(\tau_k) y_{it-1} - x_{it}'\beta(\tau_k))$$

where $\rho_\tau(u) := u(\tau - I(u < 0))$ as in Koenker and Bassett (1978), and υ_k are the weights that control the relative influence of the K quantiles $\{\tau_1, ..., \tau_K\}$ on the estimation of the η_i parameter.

However, we will see that the quantile regression fixed effects estimator, as in the ordinary least squares case, is biased in the presence of lagged dependent variables as regressors. In least squares estimation of dynamic panel models it is evident that the unobserved initial values of the dynamic process induce a bias.[4] For long panels, the effect associated with the initial conditions is seen to be $O(T^{-1})$ and therefore negligible. Later we evaluate the dynamic bias in the within-group and quantile regression fixed effects estimators by means of Monte Carlo simulation.[5] We find that the quantile regression fixed effects estimator suffers from similar bias effects to those seen in the least squares case when T is moderate.

Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991), in the linear regression case, show that instrumental variables methods are able to produce consistent estimators for dynamic panel data models that are independent of the initial conditions. These estimators are based on the idea that lagged (or lagged differences of) the regressors are correlated with the included regressor but are uncorrelated with the innovations. Thus, valid instruments, w_{it}, are available from inside the model and can be used to estimate the parameters of interest by instrumental variables methods. In this paper we use an analogous rationality for the construction of instruments.

The problem of bias for the dynamic panel quantile regression can be ameliorated through the use of instrumental variables, w, that affect the determination of lagged y but are independent of innovations. Following Chernozhukov and Hansen (2006, 2008), and assuming the availability of instrumental variables, w_{it}, we consider estimators defined as

$$\hat{\alpha} = \min_{\alpha} \|\hat{\gamma}(\alpha)\|_A,$$

where

$$(\hat{\eta}(\alpha), \hat{\beta}(\alpha), \hat{\gamma}(\alpha)) = \min_{\eta,\beta,\gamma} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k \rho_\tau (y_{it} - z_{it}\eta - \alpha(\tau_k) y_{it-1} - x_{it}'\beta(\tau_k) - w_{it}'\gamma(\tau_k)),$$

with $\|x\|_A = \sqrt{x'Ax}$, and A is a positive definite matrix.[6] Our final parameter estimates of interest are thus

$$\hat{\theta}(\tau) \equiv (\hat{\alpha}(\tau), \hat{\beta}(\tau)) \equiv (\hat{\alpha}(\tau), \hat{\beta}(\hat{\alpha}(\tau), \tau)).$$

The intuition underlying the estimator is that, since w is a valid instrument, it is independent of u and it should have a zero coefficient. Thus, for given α, the quantile regression

[4] See Hsiao (2003), Arellano (2003) and Heckman (1981) for more details.

[5] Nickell (1981) provides analytical calculations for bias in the within-group estimator in a linear dynamic panel model.

[6] As discussed in Chernozhukov and Hansen (2006), the exact form of A is irrelevant when the model is exactly identified, but it is desirable to set A equal to the asymptotic variance-covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$ otherwise.

of $(y_{it} - \alpha y_{it-1})$ on the variables (z_{it}, w_{it}, x_{it}) should generate coefficient zero for the variable w_{it}. Hence, by minimizing the coefficient of the variable w_{it} one can recover the estimator of α. Therefore, the bias generated by inclusion of y_{it-1} in equation (3) is reduced through the presence of instrumental variables, w_{it}, that affect the determination of y_{it} but are independent of u_{it}. Values of y lagged (or differences) two periods or more and/or lags of the exogenous variable x affect the determination of lagged y but are independent of u, so they can be used as instruments to estimate α and β by the quantile regression dynamic panel instrumental variables (PQRIV) method. As suggested by Chernozhukov and Hansen (2008), in practice, a simple procedure is to let the instruments w_{it} either be w_{it} or the predicted value from a least squares projection of lagged y on w_{it} and x_{it}.

The implementation of the quantile regression instrumental variables procedure is straightforward. Define the objective function

$$Q_{NT}(\tau, \eta_i, \alpha, \beta, \gamma) := \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k \rho_\tau (y_{it} - \eta_i - \alpha(\tau_k) y_{it-1} - x'_{it}\beta(\tau_k) - w'_{it}\gamma(\tau_k)) \tag{4}$$

where y_{it-1} is, in general, a $dim(\alpha)$-vector of endogenous variables, η_i are the fixed effects, x_{it} is a $dim(\beta)$-vector of exogenous explanatory variables, w_{it} is a $dim(\gamma)$-vector of instrumental variables such that $dim(\gamma) \geq dim(\alpha)$.

For the special case of $K = 1$, one can use a grid search. The quantile regression instrumental variable estimator for dynamic panel can be implemented as follows:
1) For a given quantile of interest τ, define a grid of values $\{\alpha_j, j = 1, ..., J; |\alpha| < 1\}$, and run the ordinary τ-quantile regression of $(y_{it} - y_{it-1}\alpha_j)$ on (z_{it}, w_{it}, x_{it}) to obtain coefficients $\hat{\eta}(\alpha_j, \tau)$, $\hat{\beta}(\alpha_j, \tau)$ and $\hat{\gamma}(\alpha_j, \tau)$; that is, for a given value of the autoregression structural parameter, say α, one estimates the ordinary panel quantile regression to obtain

$$(\hat{\eta}_i(\alpha_j, \tau), \hat{\beta}(\alpha_j, \tau), \hat{\gamma}(\alpha_j, \tau)) := \min_{\eta_i, \beta, \gamma} Q_{NT}(\tau, \eta_i, \alpha, \beta, \gamma). \tag{5}$$

2) To find an estimate for $\alpha(\tau)$, choose $\hat{\alpha}(\tau)$ as the value among $\{\alpha_j, j = 1, ..., J\}$ that makes $\|\hat{\gamma}(\alpha_j, \tau)\|$ closest to zero. Formally, let

$$\hat{\alpha}(\tau) = \min_{\alpha \in A} [\hat{\gamma}(\alpha, \tau)'] \hat{A}(\tau) [\hat{\gamma}(\alpha, \tau)] \tag{6}$$

where A is a positive definite matrix. The estimate $\hat{\beta}(\tau)$ is then given by $\hat{\beta}(\hat{\alpha}(\tau), \tau)$, which leads to the estimates

$$\hat{\theta}(\tau) = \left(\hat{\alpha}(\tau), \hat{\beta}(\tau)\right) = \left(\hat{\alpha}(\tau), \hat{\beta}(\hat{\alpha}(\tau), \tau)\right). \tag{7}$$

The estimator finds parameter values for α and β through the inverse step (6) such that the value of coefficient $\gamma(\alpha, \tau)$ on w in the ordinary quantile regression step (5) is driven as close to zero as possible. We show that this estimator is consistent and asymptotically normal under some regularity conditions. In addition, for the particular case of $K = 1$, the quantile regression instrumental variables estimator method may be viewed as an appropriate quantile regression analog of the two stage least squares (TSLS). The TSLS estimates can be obtained by using the same two-step procedure as described above for the PQRIV. In Appendix 1 we show the details of the derivation.

For the important case of $K > 1$, where we restrict the η's to be independent of τ, the optimization is very large depending on the number of estimated quantiles. Therefore, instead of using a grid search we use a numerical optimization function in R. As starting values, we use the parameter estimates from the quantile regression fixed effects model without any instruments. The design matrix for the problem of estimating $K > 1$ is as follows

$$[\upsilon \otimes (I_N \otimes \iota_T) \vdots \Upsilon \otimes y_{-1} \vdots \Upsilon \otimes X \vdots \Upsilon \otimes w],$$

where I_N is a $N \times N$ identity matrix, ι_T is a $T \times 1$ vector of ones, Υ is a $K \times K$ diagonal matrix with the weights υ on the diagonal. The corresponding response vector is $\tilde{y} = (\upsilon \otimes y)$. As Koenker (2004) observes, in typical applications the design matrix of the full problem is very sparse, i.e. has mostly zero elements. Standard sparse matrix storage schemes only require space for the non-zero elements and their indexing locations, and this considerably reduces the computational effort and memory requirements in large problems.

Now we discuss the asymptotic properties of the dynamic panel data quantile regression instrumental variables estimator. The existence of the parameter η_i, whose dimension N tends to infinity, raises some new issues for the asymptotic analysis of the proposed estimator. As first noted by Neyman and Scott (1948), leaving the individual heterogeneity unrestricted in a nonlinear or dynamic model generally results in inconsistent estimators of the common parameters due to the incidental parameters problem; that is, noise in the estimation of the fixed effects when the time dimension is short results in inconsistent estimates of the common parameters due to the nonlinearity of the problem. In this respect, quantile regression panel data suffers from this problem, and the presence of the fixed effects parameters, whose dimension N is tending to infinity, raises some new issues for the asymptotic analysis of the estimator. Koenker (2004) overcomes this problem by using a large N and T asymptotics with the restriction that $N^a/T \to 0$, for some $a > 0$. We derive consistency and asymptotic normality of the estimators assuming that $N \to \infty$ and $T \to \infty$ with the restriction that $N^a/T \to 0$, for some $a > 0$. We show that this is a sufficient condition for consistency and asymptotic normality of the IV estimator. We impose the following regularity conditions:

A1. The y_{it} are independent across individuals, covariance stationary, with conditional distribution functions F_{it}, and differentiable conditional densities, $0 < f_{it} < \infty$, with bounded derivatives f'_{it} for $i = 1, ..., N$ and $t = 1, ..., T$;

A2. Let $Z = I_N \otimes \iota_T$, and ι_T a T-vector of ones, $y_{-1} = (y_{it-1})$ be a $NT \times dim(\alpha)$ matrix, $X = (x_{it})$ be a $NT \times dim(\beta)$ matrix, and $W = (w_{it})$ be a $NT \times dim(\gamma)$ matrix. For

$$\begin{aligned}
\Pi(\eta, \alpha, \beta, \tau) &:= E[\upsilon(\tau - 1(Z\eta + y_{-1}\alpha + X\beta))\check{X}(\tau)] \\
\Pi(\eta, \alpha, \beta, \gamma, \tau) &:= E[\upsilon(\tau - 1(Z\eta + y_{-1}\alpha + X\beta + W\gamma))\check{X}(\tau)] \\
\check{X}(\tau) &:= [Z, W, X]',
\end{aligned}$$

Jacobian matrices $\frac{\partial}{\partial(\eta,\alpha,\beta)}\Pi(\eta, \alpha, \beta, \tau)$ and $\frac{\partial}{\partial(\eta,\beta,\gamma)}\Pi(\eta, \alpha, \beta, \gamma, \tau)$ are continuous and have full rank uniformly over $\mathscr{E} \times \mathscr{A} \times \mathscr{B} \times \mathscr{G} \times \mathscr{T}$. The parameter space, $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$, is a connected set. Moreover the image of $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$ under the map $(\eta, \alpha, \beta) \to \Pi(\eta, \alpha, \beta, \tau)$ is simply connected;

A3. Denote $\Phi(\tau_k) = diag(f_{it}(\xi_{it}(\tau_k)))$, where $\xi_{it}(\tau_k) = \eta_i + \alpha(\tau_k)y_{it-1} + x'_{it}\beta(\tau_k) + w'_{it}\gamma(\tau_k)$, $M_{Z_k} = I - P_{Z_k}$ and $P_{Z_k} = Z(Z'\Phi(\tau_k)Z)^{-1}Z'\Phi(\tau_k)$. Let $\tilde{X} = [W', X']'$. Then, the following matrices are positive definite:

$$J_\vartheta = \lim_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \upsilon_1 \tilde{X}' M'_{Z_1} \Phi(\tau_1) M_{Z_1} \tilde{X} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \upsilon_k \tilde{X}' M'_{Z_k} \Phi(\tau_k) M_{Z_k} \tilde{X} \end{pmatrix}$$

$$J_\alpha = \lim_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \upsilon_1 \tilde{X}' M'_{Z_1} \Phi(\tau_1) M_{Z_1} y_{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \upsilon_k \tilde{X}' M'_{Z_k} \Phi(\tau_k) M_{Z_k} y_{-1} \end{pmatrix},$$

and

$$S = \lim_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \Sigma_{11} \tilde{X}' M'_{Z_1} M_{Z_1} \tilde{X} & \cdots & \Sigma_{1k} \tilde{X}' M'_{Z_1} M_{Z_k} \tilde{X} \\ \vdots & \ddots & \vdots \\ \Sigma_{k1} \tilde{X}' M'_{Z_k} M_{Z_1} \tilde{X} & \cdots & \Sigma_{kk} \tilde{X}' M'_{Z_k} M_{Z_k} \tilde{X} \end{pmatrix},$$

where $\Sigma_{ij} = \upsilon_i(\tau_i \wedge \tau_j - \tau_i\tau_j)\upsilon_j$. Now define $[\bar{J}'_\beta, \bar{J}'_\gamma]'$ as a partition of J_ϑ^{-1}, and $H = \bar{J}'_\gamma A[\alpha(\tau_k)]\bar{J}_\gamma$. Then, J_ϑ is invertible, and $J'_\alpha H J_\alpha$ is also invertible;

A4. For all $\tau \in \mathcal{T} = [c, 1-c]$ with $c \in (0, 1/2)$, $(\alpha(\tau), \beta(\tau)) \in \text{int } \mathscr{A} \times \mathscr{B}$, and $\mathscr{A} \times \mathscr{B}$ is compact and convex;

A5. $\max_{it} \|y_{it}\| = O(\sqrt{NT})$; $\max_{it} \|x_{it}\| = O(\sqrt{NT})$; $\max_{it} \|w_{it}\| = O(\sqrt{NT})$;

A6. $T \to \infty$ as $N \to \infty$ and $\frac{N^a}{T} \to 0$ for some $a > 0$.

Condition A1 is a standard assumption in quantile regression literature and imposes a restriction on the density function of y_{it}. We assume covariance stationarity for simplicity.[7] Condition A2 is important for identification of the parameters. The identification is shown through the use of a version of Hadamard's theorem, as discussed in Chernozhukov and Hansen (2006). The continuity and full rank conditions require that the instrument W impacts the conditional distribution of Y at many relevant points. In addition, the condition that the image of the parameter space be simply connected requires that the image can be continuously shrunk to a point. This condition can be interpreted as ruling out "holes" in the image of the set.[8] Assumption A3 states conditions for the matrices that guarantee asymptotic normality. A4 imposes compactness on the parameter space of $\alpha(\tau)$. Such an assumption is needed since the objective function is not convex in α. Assumption A5 imposes bounds on the variables. Finally, condition A6 is the same assumption as in Koenker (2004) and allows T to grow very slowly relative to N. The recent literature analyzing bias in dynamic panel data models develops asymptotic theory where N and T are large. In a linear case, Alvarez and Arellano (2003) establish consistency of the within group estimator (WG), the generalized method of moments (GMM), and the limited information maximum likelihood estimators for a first order autoregressive model with individual effects when both N and T tend to infinity and $\lim(N/T) \equiv c < \infty$. Hahn and Kuersteiner (2002) use the same relative rate for N and T in dynamic linear and nonlinear panels, respectively.

To further comment on the nature of the correlation between Y_{-1} and W required by A2, note that, for a given quantile τ, by A1 we have that

$$\partial E[(\tau - 1(Z\eta + y_{-1}\alpha + X\beta))\check{X}(\tau)]/\partial(\eta, \alpha, \beta) = E[(Z', W', X')'\Phi(Z', Y_{-1}', X')].$$

[7]Koenker and Xiao (2006) present a general discussion of quantile autoregression where, under some mild conditions, the process y_t is globally stationary but can still display local (and asymmetric) persistence in the presence of certain types of shocks. Thus, even in the case where the autoregressive coefficient is greater than unity over some range of quantiles, under some mild conditions, y_t can still be covariance stationary in the long run. Therefore, a quantile autoregressive process may allow for some (transient) forms of explosive behavior while maintaining stationarity in the long run.

[8]We assume that the image of $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$ under the map $(\eta, \alpha, \beta) \to \Pi(\eta, \alpha, \beta, \tau)$ is connected for ease of exposition. However, it is straightforward to show that the image of a connected set by an continuous function is a connected set.

Hence, the Jacobian in A2 takes a form of density-weighted covariance matrix for Z, Y_{-1} and W variables, and A2 requires that this matrix has full rank. In addition, A2 imposes that global identifiability must hold; hence, the impact of W should be rich enough to guarantee that the equations are solved uniquely.

We can now establish consistency and asymptotic normality of the estimator. Proofs appear in Appendix 1. The following theorem states identification and consistency of $\hat{\theta}(\tau)$.

Theorem 1. Given assumptions A1-A6, $(\eta, \alpha(\tau), \beta(\tau))$ uniquely solves the equations $E[\upsilon\psi(Y - Z\eta - y_{-1}\alpha - X\beta)\check{X}(\tau)] = 0$ over $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$, and $\theta(\tau) = (\alpha(\tau), \beta(\tau))$ is consistently estimable.

Under conditions A1-A6 we show the asymptotic properties of the fixed effects PQRIV as N grows at a controlled rate relative to T. Theorem 1 provides a lower bound for the relative rate of growth of T which is sufficient for consistency. The intuition behind this condition is that T must go to infinity fast enough to guarantee consistent estimates for the fixed effects, and then for the other parameters. Under assumption A6 convergence is fast enough to eliminate the inconsistency problem found for very small T and large N asymptotic approximations.

It is important to notice that even though the dimension of the parameter space increases with the number of cross-section, in order to achieve identification, we only need to impose that parameter space, $\mathscr{E} \times \mathscr{B} \times \mathscr{A}$, is connected rather than compact. Therefore, by A2 and applying a Hadamards global univalence theorem for general metric spaces it is possible to show that there is a one-to-one correspondence between the parameter space and $\Pi(\mathscr{E}, \mathscr{B}, \mathscr{A}, \tau)$, the image of $\mathscr{E} \times \mathscr{B} \times \mathscr{A}$ under $\Pi(\cdot, \cdot, \cdot, \tau)$. Then the identification follows from the global convexity of the quantile function and the instrumental variables exclusion restriction.

The intuition behind the proof of consistency relies on the uniform convergence of the objective function over the parameter space. The basic technique used to show uniform convergence is similar to Wei and He (2006) where we divide the growing parameter space into small cubes. The total number of cubes grows at a polynomial rate so that the exponential bound obtained at each cube holds globally and the uniform convergence follows. Using this technique we establish stochastic equicontinuity, and consistency follows from application of an argmax theorem as in van der Vaart and Wellner (1996).[9]

[9] It is important to note that we can use this technique of proof since the parameter space is growing at a

The limiting distribution of parameters of interest using the quantile regression instrumental variables estimator for the dynamic panel model with fixed effects is given by Theorem 2.

Theorem 2 (Asymptotic Normality). Under conditions A1-A6, for a given $\tau \in (0,1)$, $\hat{\theta}$ converges to a Gaussian distribution:

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} N(0, \Omega(\tau)), \quad \Omega(\tau) = (K', L')'S(K', L')$$

where $S = (\min(\tau, \tau') - \tau\tau')E(VV')$, $V = (\upsilon_1 \tilde{X}' M'_{Z_1}, ..., \upsilon_k \tilde{X}' M'_{Z_k})'$, $K = (J'_\alpha H J_\alpha)^{-1} J_\alpha H$, $H = \bar{J}'_\gamma A[\alpha(\tau)]\bar{J}_\gamma$, $L = \bar{J}_\beta M$, $M = I - J_\alpha K$, $[\bar{J}_\beta, \bar{J}_\gamma]$ is a partition of J_ϑ, $\Phi = diag(f_{it}(\xi_{it}(\tau_k)))$, $\tilde{X} = [W', X']'$, and J_ϑ and J_α are as defined in assumption A3.

The proof of asymptotic normality has some elements of the nonlinear panel data literature where we concentrate out the FE. Here we write a asymptotic representation for the FE, then we plug them into the representation for all the coefficients. Therefore, there will be a reminder term coming from this two step procedure. It happens that the large N and T asymptotics with the restriction that $N^a/T \to 0$ is a sufficient condition to ensure that the reminder term is negligible and the estimator is asymptotically normal centered at zero.

Remark 1. When $dim(\gamma) = dim(\alpha)$, the choice of $A(\alpha)$ does not affect the asymptotic variance, and the joint asymptotic variance of $\alpha(\tau)$ and $\beta(\tau)$ will generally have the simple form $\Omega(\tau) = (K', L')'S(K', L')$, for S, K and L as defined above. As in Chernozhukov and Hansen (2008), when $dim(\gamma) > dim(\alpha)$, the choice of the weighting matrix $A(\alpha)$ generally matters, and it is important for efficiency. A natural choice for $A(\alpha)$ is given by the inverse of the covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$. Noticing that $A(\alpha)$ is equal to $(\bar{J}_\gamma S \bar{J}_\gamma)^{-1}$ at $\alpha(\tau)$, it follows that the asymptotic variance of $\sqrt{NT}(\hat{\alpha}(\tau) - \alpha(\tau))$ is given by

$$\Omega_\alpha = (J'_\alpha \bar{J}'_\gamma (\bar{J}_\gamma S \bar{J}'_\gamma)^{-1} \bar{J}_\gamma \bar{J}_\alpha)^{-1}.$$

The components of the asymptotic variance matrix that need to be estimated include J_ϑ, J_α and S. The matrix S can be estimated by its sample counterpart

$$\hat{S}(\tau, \tau') = (\min(\tau, \tau') - \tau\tau') \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} V_{it} V'_{it}. \quad (8)$$

known rate, N. There is a large literature showing the asymptotic properties of quantile regression estimator for infinite dimension parameter space when the rate of the parameter increasing is unknown, see e.g. He and Shao (2000) and Portnoy (1985). However, it is possible to achieve a better rate of N relative to T in the first case.

Following Powell (1986), J_ϑ and J_α can be estimated as stated in Theorem 2 above. The typical element of $\hat{J}_\vartheta$ takes the following form

$$\hat{J}_{\vartheta_j} = \upsilon_j \frac{1}{2NTh_n} \sum_{i=1}^{N} \sum_{t=1}^{T} I(|\hat{u}(\tau_j)| \leq h_n) \tilde{X} M_{Z_j} M'_{Z_j} \tilde{X}'. \tag{9}$$

where $\hat{u}(\tau_j) \equiv Y - Z\hat{\eta} - \hat{\alpha}(\tau_j)y_{-1} - X\hat{\beta}(\tau_j)$ and h_n is an appropriately chosen bandwidth, with $h_n \to 0$ and $NTh_n^2 \to \infty$. The estimator of $\hat{J}_{\alpha_j}$ is analogous to $\hat{J}_{\vartheta_j}$. Using the same procedure we can estimate the element $Z\hat{\Phi}(\tau_j)Z$ in P_Z. The consistency of these asymptotic covariance matrix estimators is standard and will not be discussed further in this paper.

2.2 Prediction

Most econometric forecasting has focused on models for the conditional mean under Gaussian conditions. The dynamic panel data quantile regression models described above offer an opportunity to significantly expand the scope of forecasting applications. One-step ahead forecasts of the quantile function of y_{it} for an individual i are immediately available:

$$\hat{Q}_{y_{iT+1}}(\tau | y_{iT}, x_{iT+1}) = \hat{\eta}_i + \hat{\alpha}(\tau) y_{iT} + x'_{iT+1}\hat{\beta}(\tau). \tag{10}$$

Another important application of PQRIV models is out-of-sample prediction. Quantile regression offers a natural approach to the construction of prediction intervals as noted, for example, by Koenker and Zhao (1996), and Zhou and Portnoy (1996). The methods proposed by the later work suggest the construction of a $1-\lambda$ level interval for an s-step-ahead forecast as

$$[Q_{y_{iT+s}}(\lambda/2 - h_n), Q_{y_{iT+s}}(1 - \lambda/2 + h_n)]$$

where $h_n \to 0$ to account for parameter uncertainty.[10]

An important practical aspect of the forecast interval problem involves computing $\hat{Q}_{y_{iT+s}}(\cdot)$. This is straightforward in the one-step ahead case, as shown in equation (10), but more problematic for $s > 1$. Koenker and Zhao (1996) suggest a simply approach based on simulation for implementing out-of-sample quantile regression prediction, and a similar approach seems reasonable for the quantile regression dynamic panel data problem.

[10] Of course, if the parameters of the model were known exactly, the conditional quantile function itself could be used. The interval

$$[Q_{y_{iT+s}}(\lambda/2), Q_{y_{iT+s}}(1 - \lambda/2)]$$

would provide an exact $1 - \lambda$ level interval for an s-step-ahead forecast.

Let $Q_{y_{iT}}(\tau|y_{iT-1}, x_{iT})$ denote the conditional quantile function of y_{iT}, for an individual i, given the information up to time $T-1$. A draw from the one-step-ahead forecast distribution is given by

$$\hat{y}_{iT+1} = \hat{Q}_{y_{iT+1}}(U|y_{iT}, x_{iT+1}) = \hat{\eta}_i + \hat{\alpha}(U)y_{iT} + x'_{iT+1}\hat{\beta}(U), \tag{11}$$

where U is a uniformly distributed random variable on $[0, 1]$ and x_{iT+1} is given. Applying (11) recursively we can compute a sample path of forecasts $(\hat{y}_{iT+1}, \hat{y}_{iT+2}, ..., \hat{y}_{iT+s})'$. Repeatedly applying this procedure, say R times, one can compute the $\lambda/2-h_n$ and $1-\lambda/2+h_n$, quantiles of the empirical distribution of the forecasts and used them to construct the final prediction intervals. It may appear that the use of (11) is computationally prohibitive because it appears to require a quantile regression estimate for each possible realization of $U \in (0, 1)$. However, the entire function $Q_{y_{iT}}(\tau|y_{iT-1}, x_{iT})$ is easily computed by standard parametric linear programming techniques, yielding a piecewise constant function on a known grid that is then readily evaluated by the forecasting simulation.

Conditional density forecasts s-step-ahead, for the individual i, can be constructed based on an ensemble of such forecast paths. The simulation method described previously produces R prediction samples of conditional quantile function of interest for period $T + s$. Thus, one can apply any typical nonparametric density estimator to predict the conditional density function s-step-ahead.

3 Inference

In this section, we turn our attention to inference in the quantile regression dynamic panel instrumental variable (PQRIV) model, and suggest a Wald type test for general linear hypotheses and a Kolmogorov-Smirnov test for linear hypotheses over a range of quantiles $\tau \in \mathcal{T}$.

In the independent and identically distributed setting the conditional quantile functions of the response variable, given the covariates, are all parallel, implying that covariate effects shift the location of the response distribution but do not change the scale or shape. However, slope estimates often vary across quantiles implying that it is important to test for equality of slopes across quantiles. Wald tests designed for this purpose were suggested by Koenker and Bassett (1982a), Koenker and Bassett (1982b), and Koenker and Machado (1999). It is possible to formulate a wide variety of tests using variants of the proposed Wald test, from simple tests on a single quantile regression coefficient to joint tests involving many covariates and distinct quantiles at the same time.

General hypotheses on the vector $\theta(\tau)$ can be accommodated by Wald-type tests. The Wald process and associated limiting theory provide a natural foundation for the hypothesis $R\theta(\tau) = r$ when r is known. We first consider a Wald type test where we test the coefficients for selected quantiles of intest. Later we introduce a test for linear hypothesis over a range of quantiles $\tau \in \mathcal{T}$, instead of focusing only on a selected quantile. The following are examples of hypotheses that may be considered in the former framework. For simplicity of presentation we use the model stated in equation (3) with a single variable in the x_{it} matrix.

Example 1 (No dynamic effect). For a given τ, if there is no dynamic effect in the model, then under $H_0 : \alpha(\tau) = 0$. Thus, $\theta(\tau) = (\alpha(\tau), \beta(\tau))'$, $R = [1, 0]$ and $r = 0$.

Example 2 (Location shifts). The hypotheses of location shifts for $\alpha(\tau)$ and $\beta(\tau)$ can be accommodated in the model. For the first case, $H_0 : \alpha(\tau) = \alpha$, for $|\alpha| < 1$, so $\theta(\tau) = (\alpha(\tau), \beta(\tau))'$, $R = [1, 0]$ and $r = \alpha$. For the latter case, $H_0 : \beta(\tau) = \beta$, so that $R = [0, 1]$ and $r = \beta$.

Portnoy (1984) and Gutenbrunner and Jureckova (1992) show that the quantile regression process is tight and thus the limiting variate viewed as a function of τ is a Brownian Bridge over $\tau \in \mathcal{T}$.[11] Therefore, under the linear hypothesis $H_0 : R\theta(\tau) = r$, conditions A1-A6, and letting $\Gamma = (K', L')'EVV'(K', L')$, we have

$$\mathcal{V}_{NT} = \sqrt{NT}[R\Gamma(\tau)R']^{-1/2}(R\hat{\theta}(\tau) - r) \Rightarrow B_q(\tau), \tag{12}$$

where $B_q(\tau)$ represents a q-dimensional standard Brownian Bridge. For any fixed τ, $B_q(\tau)$ is $N(0, \tau(1-\tau)I_q)$. The normalized Euclidean norm of $B_q(\tau)$

$$Q_q(\tau) = \|B_q(\tau)\| / \sqrt{\tau(1-\tau)}$$

is generally referred to as a Kiefer process of order q. Thus, for given τ, the regression Wald process can be constructed as

$$\mathcal{W}_{NT} = NT(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r), \tag{13}$$

where $\hat{\Omega}$ is a consistent estimator of Ω, and Ω is given by

$$\Omega(\tau) = (K'(\tau), L'(\tau))'S(\tau, \tau)(K'(\tau), L'(\tau)).$$

[11] In a related result, Wei and He (2006) establish tightness of the quantile regression process in the longitudinal data context with increasing parameter dimension, see for instance, Lemma 8.4.

If we are interested in testing $R\theta(\tau) = r$ at a particular quantile $\tau = \tau_0$, a Chi-square test can be conducted based on the statistic $\mathcal{W}_{NT}(\tau_0)$. Under H_0, the statistic $\mathcal{W}_{NT}$ is asymptotically χ^2_q with q-degrees of freedom, where q is the rank of the matrix R. The limiting distribution of the test is summarized in the following theorem

Theorem 3 (Wald Test Inference). Under $H_0 : R\theta(\tau) = r$, and conditions A1-A6, for fixed τ,

$$\mathcal{W}_{NT}(\tau) \overset{a}{\sim} \chi^2_q.$$

Proof. The proof of Theorem 3 is simple; it follows from observing that for any fixed τ, by Theorem 2

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) \Rightarrow N(0, \Omega(\tau))$$

and under the null hypothesis,

$$\sqrt{NT}(R\hat{\theta}(\tau) - r) \Rightarrow N(0, R\Omega(\tau)R')$$

since $\hat{\Omega}(\tau)$ is a consistent estimator of $\Omega(\tau)$, by the Slutsky's theorem,

$$\mathcal{W}_{NT} = NT(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r) \overset{a}{\sim} \chi^2_q.$$

$\blacksquare$

In order to implement the test it is necessary to estimate $\Omega(\tau)$ consistently. It is possible to obtain such an estimator as suggested in Theorem 2 in the previous section, and the main components of $\hat{\Omega}(\tau)$ can be obtained as in equations (8) and (9).

More general hypotheses are also easily accommodated by the Wald approach. Let $\zeta = (\theta(\tau_1)', ..., \theta(\tau_m)')$ and define the null hypothesis as $H_0 : R\nu = r$. The test statistic is the same Wald test as equation (13). In this case Ω is the matrix with (k, l)th block

$$\Omega(\tau_k, \tau_l) = (K'(\tau_k), L'(\tau_k))' S(\tau_k, \tau_l)(K'(\tau_l), L'(\tau_l)),$$

where $S(\tau_k, \tau_l)$ is as defined in Theorem 2, and the other variables are as defined above. The statistic $\mathcal{W}_{NT}$ is still asymptotically χ^2_q under H_0 where q is the rank of the matrix R. This formulation accommodates a wide variety of testing situations, from a simple test on single quantile regression coefficients to joint tests involving several covariates and distinct quantiles. Thus, for instance, we might test for the equality of several slope coefficients across several quantiles.

Example 3 (Same dynamic effect for two distinct quantiles). If there are the same dynamic effects for two given distinct quantiles in the model, then under $H_0 : \alpha(\tau_1) = \alpha(\tau_2)$. Thus, $\zeta = (\theta(\tau_1)', ..., \theta(\tau_m)') = (\alpha(\tau_1), \beta(\tau_1), \alpha(\tau_2), \beta(\tau_2))'$, $R = [1, 0, -1, 0]$ and $r = 0$.

Another important class of tests in the quantile regression literature involves the Kolmogorov-Smirnov (KS) type tests, where the goal is to examine the property of the estimator over a range of quantiles $\tau \in \mathcal{T}$, instead of focusing only on a selected quantile. Thus, if one has interest in testing $R\theta(\tau) = r$ over $\tau \in \mathcal{T}$, one may consider the KS type sup-Wald test. Following Koenker and Xiao (2006), we may construct a KS type test on the dynamic panel data regression quantile process using

$$KS\mathcal{W}_{NT} = \sup_{\tau \in \mathcal{T}} \mathcal{W}_{NT}(\tau). \tag{14}$$

The next example shows a possible application of the KS test in dynamic panel quantile regression.

Example 4 (Asymmetric dynamic effect). It is particularly interesting to analyze data displaying asymmetric dynamics. Thus, one may consider testing the hypothesis that $\alpha(\tau) = \alpha$ over τ, such that, $H_0 : \alpha(\tau) = \alpha$ over $\tau \in \mathcal{T}$.

To perform such a test for $\alpha(\tau)$ one can use the Kolmogorov-Smirnov test given in (14). The test is implemented by computing the test statistic $\mathcal{W}_{NT}(\tau)$, given in (13), for each $\tau \in \mathcal{T}$, and then calculating the maximum over τ. The limiting distribution of the Kolmogorov-Smirnov test is given in the following theorem

Theorem 4 (Kolmogorov-Smirnov Test). Under H_0 and conditions A1-A6,

$$KS\mathcal{W}_{NT} = \sup_{\tau \in \mathcal{T}} \mathcal{W}_{NT}(\tau) \Rightarrow \sup_{\tau \in \mathcal{T}} Q_q^2(\tau).$$

The proof of Theorem 4 follows directly from the continuous mapping theorem and equation (12). Critical values for $\sup Q_q^2(\tau)$ have been tabled by DeLong (1981) and, more extensively, by Andrews (1993) using simulation methods.

4 Monte Carlo Simulation

4.1 Monte Carlo Design

In this section, we describe the design of simulation experiments used to assess the finite sample performance of the quantile regression estimator and inference procedures discussed in the previous sections.[12] Two simple versions of the basic model (1) are considered in the simulation experiments. In the first, reported in Tables 1 and 2, the exogenous covariate, x_{it}, exerts a pure location shift effect. In the second, reported in Tables 3 and 4, x_{it} exerts both location and scale effects. In the former case the response y_{it} is generated by the model,

$$y_{it} = \eta_i + \alpha y_{it-1} + \beta x_{it} + u_{it}$$

while in the latter case,

$$y_{it} = \eta_i + \alpha y_{it-1} + \beta x_{it} + (\gamma x_{it}) u_{it}.$$

We employ two different schemes to generate the disturbances u_{it}. Under Scheme 1, we generate u_{it} as a $N(0, \sigma_u^2)$, and we also used a heavier tailed distribution scheme to generate u_{it}. Under Scheme 2 we generate u_{it} as a t-distribution with 3 degrees of freedom.

The regressor x_{it} is generated according to

$$x_{it} = \mu_i + \zeta_{it}, \tag{15}$$

where ζ_{it} follows an ARMA(1, 1) process,

$$(1 - \phi L)\zeta_{it} = \epsilon_{it} + \theta \epsilon_{it-1}, \tag{16}$$

and ϵ_{it} follows the same distribution as u_{it}, that is, normal distribution and t_3 for Schemes 1 and 2, respectively. In all cases we set $\zeta_{i,-50} = 0$ and generate ζ_{it} for $t = -49, -48, ..., T$ according to

$$\zeta_{it} = \phi \zeta_{it-1} + \epsilon_{it} + \theta \epsilon_{it-1}. \tag{17}$$

This ensures that the results are not unduly influenced by the initial values of the x_{it} process. In generating y_{it} we also set $y_{i,-50} = 0$ and discard the first 50 observations, using the observations $t = 0$ through T for estimation. The fixed effects, μ_i and α_i, are generated as

$$\mu_i = e_{1i} + T^{-1} \sum_{t=1}^{T} \epsilon_{it}, \quad e_{1i} \sim N(0, \sigma_{e_1}^2),$$

$$\eta_i = e_{2i} + T^{-1} \sum_{t=1}^{T} x_{it}, \quad e_{2i} \sim N(0, \sigma_{e_2}^2).$$

The above method of generating μ_i and α_i ensures that the usual random effects estimators are inconsistent because of the correlation that exists between the individual effects and the error term or the explanatory variables.

[12] The experiment shown in this section builds on Hsiao, Pesaran, and Tahmiscioglu (2002).

In the simulations, we consider $T = 10, 20$ and $N = 50, 100$. We set the number of replications to 2000, and consider the following values for the remaining parameters:

$$(\alpha, \beta) = (0.5, 0.7), (0.8, 0.7);$$
$$\phi = 0.7, \quad \theta = 0.2, \quad \gamma = 0.5, \quad \sigma_u^2 = \sigma_{e_1}^2 = \sigma_{e_2}^2 = 1.$$

In the Monte Carlo study, we compare the coefficient estimates in terms of bias and root mean squared error. We also investigate the small sample properties of the tests based on different estimators, paying particular attention to the size and power of these tests.

4.2 Monte Carlo Results

We study four different estimators in the Monte Carlo experiment, the within group estimator (WG), the OLS instrumental variables estimator (OLS-IV), the fixed effects panel quantile regression estimator (PQR) proposed by Koenker (2004), and the quantile regression dynamic panel instrumental variables estimator (PQRIV) that we propose in this paper. The quantile regression based estimators are analyzed for three quantiles ($\tau = (0.25, 0.5, 0.75)$), estimated simultaneously, with equal weight for each one. For the OLS-IV and PQRIV estimators we considered two different instruments, y_{it-2} and x_{it-1}; the results are essentially the same in both cases, and we simply present results for the x_{it-1} case. We also consider different sample sizes in the experiments. However, due to space limitations we report results for only $T = 10$ and $N = 50$. The results for the other sample size schemes are similar.

4.2.1 Bias and RMSE

We first study the bias and root mean squared error (RMSE) of the estimators. Tables 1 and 2 present bias and RMSE results for estimates of the autoregression coefficient, α, and the exogenous variable coefficient, β, for the location-shift model and the Normal and t_3 distributions of the innovations, respectively. For both configurations of the parameters (α, β) the biases of the estimates are roughly constant. Moreover, the coefficient estimate of the exogenous variable is slightly biased in the WG and PQR cases.

Table 1 shows that when the disturbances are drawn from a Gaussian distribution, then as expected, the autoregression coefficient is biased downward for the the WG case, but the OLS-IV is approximately unbiased. In the same way, in the presence of lagged variables the fixed effects quantile regression estimator proposed by Koenker (2004), PQR is biased

		WG	OLS-IV	PQR			PQRIV		
				$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$\alpha = 0.8$	Bias	-0.0908	0.0016	-0.0962	-0.0920	-0.0895	-0.0187	-0.0071	0.0003
	RMSE	0.094	0.061	0.103	0.099	0.097	0.069	0.067	0.068
$\beta = 0.7$	Bias	0.0193	0.0026	-0.0089	0.0185	0.0473	-0.0111	0.0014	0.0139
	RMSE	0.052	0.065	0.070	0.063	0.080	0.073	0.069	0.073
$\alpha = 0.5$	Bias	-0.0982	-0.0070	-0.1044	-0.0976	-0.0953	-0.0190	-0.0013	0.0097
	RMSE	0.103	0.082	0.113	0.109	0.107	0.093	0.089	0.088
$\beta = 0.7$	Bias	0.0349	-0.0006	-0.0019	0.0325	0.0683	-0.0158	-0.0015	0.0178
	RMSE	0.055	0.065	0.063	0.057	0.091	0.075	0.069	0.069

Table 1: Location-Shift Model: Bias and RMSE of Estimators for Normal Distribution ($T = 10$ and $N = 50$)

downward. However, the instrumental variables estimator is able to largely eliminate the bias. Table 1 reveals that the PQRIV estimator is approximately unbiased for both selections of the parameters α and β. In summary, estimates are biased in both the WG and the PQR cases, and the instrumental variables strategy is able to diminish the bias considerably for both ordinary least squares and quantile regression cases. Regarding the RMSE in the Gaussian case, the OLS-based estimators perform better than the respective quantile regression estimators. Thus, in the Gaussian case, the PQRIV is capable of reducing the bias but it has a larger RMSE when compared with the OLS-IV.

Table 2 presents the results for the t_3-distribution case. The autoregressive estimates of WG and PQR are biased downward, and the WG has a larger bias when compared with the same estimator in the Gaussian case. The PQRIV and OLS-IV are approximately unbiased estimators for both coefficients. In contrast to the case when innovations were Gaussian, with a heavier-tailed distribution of innovations, the RMSE results from the quantile estimators are smaller than their respective OLS based estimators.

Tables 3 and 4 present bias and RMSE results of the estimators for the location-scale-shift model for the Normal and t_3 distributions, respectively. As in the location-shift case, the bias of the estimators are roughly constant for the two different configurations. Again, it is possible to note that the quantile regression instrumental variables estimator presents a much smaller bias when compared with PQR, and a much improved precision when compared with OLS-IV, in the t_3 case.

Table 3 shows that in the Gaussian case the WG and PQR estimators are biased downward and the OLS-IV and PQRIV are approximately unbiased. As in the location-shift case,

		WG	OLS-IV	PQR			PQRIV		
				$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$\alpha = 0.8$	Bias	-0.1540	0.0017	-0.0983	-0.0947	-0.0910	-0.0197	-0.0037	-0.0088
	RMSE	0.161	0.116	0.104	0.100	0.098	0.075	0.070	0.076
$\beta = 0.7$	Bias	0.0391	-0.0038	-0.0182	0.0225	0.0454	-0.0135	-0.0027	0.0090
	RMSE	0.086	0.129	0.077	0.072	0.087	0.080	0.075	0.085
$\alpha = 0.5$	Bias	-0.1432	0.0107	-0.0969	-0.1018	-0.0923	-0.0155	-0.0055	0.0056
	RMSE	0.153	0.151	0.109	0.114	0.106	0.098	0.095	0.094
$\beta = 0.7$	Bias	0.0635	0.0115	-0.0134	0.0364	0.0779	-0.0144	-0.0018	0.0166
	RMSE	0.106	0.117	0.074	0.065	0.111	0.083	0.073	0.092

Table 2: Location-Shift Model: Bias and RMSE of Estimators for t_3 Distribution ($T = 10$ and $N = 50$)

		WG	OLS-IV	PQR			PQRIV		
				$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$\alpha = 0.8$	Bias	-0.0917	-0.0003	-0.0921	-0.0971	-0.1015	-0.0091	-0.0006	0.0047
	RMSE	0.095	0.060	0.099	0.105	0.106	0.061	0.065	0.065
$\beta = 0.7$	Bias	0.0239	0.0085	-0.0465	0.0094	0.0662	-0.0108	0.0003	0.0138
	RMSE	0.068	0.088	0.077	0.072	0.095	0.080	0.092	0.095
$\alpha = 0.5$	Bias	-0.0967	0.0019	-0.0939	-0.0959	-0.0987	-0.0073	-0.0003	0.0069
	RMSE	0.099	0.078	0.104	0.106	0.109	0.083	0.083	0.086
$\beta = 0.7$	Bias	0.0454	0.0003	-0.0337	0.0180	0.0699	-0.0053	-0.0018	0.0172
	RMSE	0.077	0.080	0.081	0.085	0.091	0.082	0.084	0.086

Table 3: Location-Scale Shift Model: Bias and RMSE of Estimators for Normal Distribution ($T = 10$ and $N = 50$)

in the presence of dynamic variables, the fixed effects quantile regression estimator proposed by Koenker (2004) is biased downward, and the instrumental variables proposed in this paper dramatically reduces the bias. The RMSE's present the same features as in the previous location case. Table 4 presents the results for the t_3-distribution case, and the results are qualitatively similar to those in Table 2, where under non-Gaussian heavier-tailed conditions the quantile regression estimators perform better than the least squares based estimators in terms of RMSE.

4.2.2 Size and Power

Now we turn our attention to the size and power of the asymptotic inference procedures given in the previous section. First, we concentrate on tests for selected quantiles, then we consider tests over a range of quantiles. For the former case, we use the same Monte Carlo

		WG	OLS-IV	PQR			PQRIV		
				$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$\alpha = 0.8$	Bias	-0.1710	0.0254	-0.0946	-0.0976	-0.0925	-0.0110	-0.0049	0.0026
	RMSE	0.182	0.233	0.101	0.109	0.102	0.071	0.072	0.069
$\beta = 0.7$	Bias	0.0430	0.0105	-0.0531	0.0123	0.0779	-0.0168	-0.0007	0.0181
	RMSE	0.115	0.159	0.087	0.069	0.109	0.086	0.089	0.097
$\alpha = 0.5$	Bias	-0.1244	0.0173	-0.0881	-0.0894	-0.0905	-0.0152	-0.0044	0.0042
	RMSE	0.139	0.132	0.103	0.107	0.109	0.079	0.085	0.086
$\beta = 0.7$	Bias	0.0727	0.0104	-0.0415	0.0238	0.0905	-0.0186	0.0011	0.0125
	RMSE	0.132	0.153	0.081	0.063	0.119	0.087	0.088	0.096

Table 4: Location-Scale Shift Model: Bias and RMSE of Estimators for t_3 Distribution ($T = 10$ and $N = 50$)

setup as in the calculations of bias and RMSE to calculate the power curves. We present the results for PQRIV as well as for OLS-IV in order to compare the finite sample performance of the estimators. Thus, we consider the PQRIV model in equation (1) and test the hypothesis that $\hat{\alpha}(\tau) = \alpha$ and also that $\hat{\beta}(\tau) = \beta$ for a given τ. We present the results for $\alpha = 0.5$ and $\beta = 0.7$.[13] For models under the alternative, we consider linear deviations $\alpha + d/\sqrt{NT}$ and $\beta + d/\sqrt{NT}$. The construction of the test uses an estimate of the density as given in equation (9). The procedure proposed by Powell (1986) entails a choice of bandwidth. We consider the default bandwidth suggested by Bofinger (1975)

$$h_n = [\Phi^{-1}(\tau + c_n) - \Phi^{-1}(\tau - c_n)] \min(\hat{\sigma}_1, \hat{\sigma}_2)$$

where $c_n = (NT)^{-1/5}((4.5\phi^4(\Phi^{-1}(t)))/(2\Phi^{-1}(t)^2 + 1)^2)^{1/5}$, $\hat{\sigma_1} = \sqrt{Var(\hat{u})}$, and $\hat{\sigma_2} = (\hat{Q}(\hat{u}, .75) - \hat{Q}(\hat{u}, .25))/1.34$. We also use a Gaussian bandwidth, but since the results are essentially the same we only present results for the first choice of bandwidth. The results for $T = 10$ and $N = 50$ are presented in Figures 1 and 2.

Figure 1 shows the finite sample size and power for the estimated α and β coefficients considering Normal distributions and PQRIV and OLS-IV estimators. Part 1 of the figure concerns α and Part 2 shows the results for β. Observe that the size is very close to the established five percent for all estimators. When comparing PQRIV and OLS-IV estimators with respect to the Normal distribution one can see that the OLS based estimators perform better than the quantile regression estimator in terms of power.

[Figure 1 about here]

[13] The results for $\alpha = 0.8$ and $\beta = 0.7$ are similar.

Figure 2 presents the results for finite sample size and power for the estimated α and β coefficients considering t_3 distribution and PQRIV and OLS-IV estimators. As in the previous case, the size is very close to the established five percent for all estimators. When the noise in the model comes from a heavier-tailed distribution, t_3, the PQRIV estimators have a strongly superior performance *vis-a-vis* the OLS-IV estimators, showing that there are large gains in power from using a robust estimator when innovations follow a non-Gaussian heavy tailed distribution.

In summary, the results for the power curves show that the OLS-IV presents more power than PQRIV in the Normal case, but in the t_3 case the opposite occurs, as PQRIV estimators have far more power than OLS-IV. The PQRIV presents higher power *vis-a-vis* the OLS-IV estimator in the non-Gaussian case. In addition, as expected, comparison of the same estimators using different distributions show that the quantile regression estimator performs better in the t_3 distribution case, and the OLS-IV estimator has more power in the Gaussian setting. The results for the other sample cases are qualitatively similar to those of Figures 1 and 2, but also show that, as the sample sizes increase, the tests do have improved power properties, corroborating the asymptotic theory.

[Figure 2 about here]

We also conduct a Monte Carlo experiment to examine the PQRIV-based inference procedures when we are particularly interested in models displaying asymmetric dynamics. Accordingly, we consider the PQRIV model to test the hypothesis that $\alpha(\tau) = \alpha$ for all τ. The data in these experiments were generated from model (1) in the same manner as in Section 4.1, where u_{it} are *i.i.d.* random variables. We consider the Kolmogorov-Smirnov test $KS\mathcal{W}_{NT}$ given by (14) for different sample sizes and innovation distributions, and choose $\mathcal{T}$ = [0.1, 0.9]. Both Normal innovations and student-t innovations (with 3 degrees of freedom) are considered. The number of repetitions is 1000.

Representative results of the empirical size and power of the proposed tests are reported in Table 5. We report results for two choices of $\alpha(\tau)$: $\alpha = 0.25$ and $\alpha = 0.35$. We consider the following model

$$y_{it} = \eta_i + \alpha y_{it-1} + \beta x_{it} + (1 + \gamma y_{it-1})u_{it},$$

with the other variables defined as in the previous section, using $\gamma = 0.3$ to compute power, and $\gamma = 0$ the size.

	Model	Normal	t_3
Size	$\alpha = 0.25$	0.05	0.06
	$\alpha = 0.35$	0.04	0.06
Power	$\alpha = 0.25$	0.88	0.94
	$\alpha = 0.35$	0.85	0.89

Table 5: Size and Power for Normal Distribution

Table 5 reports the empirical size and power for Gaussian innovations and the sample size $T = 10$ and $N = 50$, as well as the results for the student-t innovations and same sample size. Results in Table 5 show that the empirical size of the test is close to the established nominal 5%, and also confirm that using the quantile regression based approach leads to power gains in the presence of heavy-tailed disturbances. (Such gains obviously depend on choosing quantiles at which there is sufficient conditional density.)

5 Application

In this section we illustrate the new approach proposed in this paper by applying the estimator and test procedures to test for the presence of time non-separability in utility using household consumption panel data. A simple model of habit formation implies a condition relating the strength of habits to the evolution of consumption over time. We build on previous work by testing the time separability of preferences with household panel data using the quantile regression dynamic panel instrumental variables (PQRIV) framework.

A growing body of literature has emphasized the importance of allowing for habit formation as a way of modeling time dependence in preferences in order to improve the predictions of time-separable models. For instance, some authors have shown that habit persistence may partially solve the equity premium puzzle, because it smooths consumption growth over and above the smoothing implied by the life cycle-permanent income hypothesis with time-separable preferences (Constantinides (1990)). In addition, if preferences exhibit habit formation, consumption reacts slowly to permanent income shocks, and this can, in principle, explain the excess sensitivity of nondurable consumption observed in the aggregate data. The notion of habit persistence has also been used to address other important issues in macroeconomics and finance, such as the hump-shaped response of consumption to monetary and other shocks (Fuhrer (2000)), the relationship between savings and growth (Carroll, Overland, and Weil (2000)), or the volatility puzzle (Campbell and Cochrane (1999)).

Despite growing interest in the implications of preferences that are time-nonseparable, most empirical papers have used aggregate consumption data to look for empirical evidence of such preferences. Studies of time-nonseparable preferences based on aggregated consumption data obtain mixed conclusions about the strength of habit formation. Dunn and Singleton (1986) and Eichenbaum, Hansen, and Singleton (1988) find very weak evidence of habit formation in US monthly aggregated data. Ferson and Constantinides (1991) find strong evidence of habit formation in monthly, quarterly, and annual US consumption data.

There is a large literature using household panel data on consumption to examine behavior when preferences are assumed to be time-separable, for instance, Hall and Mishkin (1982), Shapiro (1984), and Zeldes (1989). However, there is a lack of empirical microeconomic evidence. In the recent literature the evidence of habit formation is mixed. While Meghir and Weber (1996) and Dynan (2000) do not find evidence of habit formation in preferences at the household level, Naik and Moore (1996) do find support of habit formation. In the same way, Carrasco, Labeaga, and Lopez-Salido (2005), and Browning and Collado (2007) have found mixed evidence of habit formation using Spanish panel data set.

With habit formation, current utility depends not only on current expenditures, but also on a "habit stock" formed by lagged expenditures. For a given level of current expenditures, a larger habit stock lowers utility. More formally, household i chooses current consumption expenditures, c_{it}, to maximize

$$E\left[\sum_{s=0}^{T}\beta^{s}u\left(\tilde{c}_{it+s};\psi_{it+s}\right)\right],$$

where $\tilde{c}_{it}$ is consumption services in period t, β is a time discount factor, and $\psi_{i,t}$ corresponds to "taste-shifters" – variables that move marginal utility – at time t. Consumption services in period t are positively related to current expenditures and negatively related to lagged expenditures according to

$$\tilde{c}_{it} = c_{it} - \alpha c_{it-1}.$$

The parameter α measures the strength of habit formation; when α is larger, the consumer receives less lifetime utility from a given amount of expenditure. From the first order condition, assuming that the utility function is of the following isoelastic form[14]

$$u(\tilde{c}_{it},\psi_{it}) = \psi_{it}\frac{\tilde{c}_{it}^{1-\rho}}{1-\rho},$$

[14]We provide the details of the derivation in Appendix 2.

and interest rates are constant, and following Dynan (2000), approximating $\Delta \ln(c_{it} - \alpha c_{it-1})$ with $(\Delta \ln c_{it} - \alpha \Delta \ln c_{it-1})$, one can derive the following equation governing consumption dynamics[15]

$$\Delta \ln(c_{it}) = \gamma_0 + \alpha \Delta \ln(c_{it-1}) + \gamma_1 \Delta \ln(\psi_{it}) + \epsilon_{it}. \tag{18}$$

Habit persistence enters the Euler equation through lagged consumption growth and habit persistence coefficient α. The habit-formation model predicts $\alpha > 0$, with its magnitude reflecting the fraction of past expenditures that make up the habit stock and indicating the importance of habit formation in behavior. In other words, the equation shows that habit formation creates a positive link between current and lagged expenditure growth, which stems from consumers' gradual adjustment to permanent income shocks. In contrast to traditional models in which consumption adjusts immediately to innovations to permanent income, habits cause consumers to prefer a number of small consumption changes to one large consumption change.

When studying persistent behavior we have to be careful to distinguish between the different possible sources of persistence in behavior. Consider, for example, smoking. It is clear that the probability of someone smoking in the current period is dependent on smoking behavior in the past, but this could be because some people are 'smokers' (individual heterogeneity) or because something induced them to start at some point and then they continue (state dependence). To have any chance of controlling for individual heterogeneity and state dependence we need dynamic panel data with several periods of observation for each household. In addition, in the study of persistence in behavior, it is extremely important to study the conditional heterogeneity associated with the extension of the addiction for different individuals; for example, it is important to differentiate between the extension of the addiction for heavy and light smokers. Thus, quantile regression for dynamic panel data is a suitable tool for analyzing persistent behavior since it allows to control for individual specific intercepts in the dynamic panel data context, and most importantly, it allows one to explore a range of conditional quantile functions exposing a variety of forms of conditional heterogeneity.

We estimate a quantile regression model using data from the Panel Study on Income Dynamics (PSID), which contains annual information about the income, food consumption,

[15] Although γ_0 is a function of real interest rates, the time discount factor, and the forecast error variance (see Dynan (2000) for details), most Euler equation analyses with household data have assumed these terms constant across households and time periods.

employment, and demographic characteristics of individual households. The PSID has limited consumption information, and we follow a substantial body of the literature in using food expenditures to explore consumption behavior. The baseline sample contains 2132 households, each with 13 observations on food expenditure growth. Although the PSID began in 1968 and continues today, the sample uses spending data only from the period 1974 through 1987 because of interpretation problems in the early years and the suspension of food questions in 1987.

To test the consumption habit formation hypothesis, estimation and testing are based on the following equation

$$Q_{C_{it}}(\tau|\mathcal{F}_{it-1}) = \eta_i + \alpha(\tau)C_{it-1} + X_{it}\beta(\tau), \tag{19}$$

where $C_{it} = \Delta \ln c_{it}$, and X_{it} is a set of covariates. As covariates, following Dynan (2000), we include the following taste shifters in the estimated equations: number of adult male equivalents in the household, age of the head of the household, and age squared. All specifications include time dummies to ensure that aggregate shocks do not lead to inconsistent estimates. In a second round of estimation we also include race as a variable in the model. The consumption variables are measured in logs. We also apply OLS based two stage least squares (TSLS) estimation and testing for comparison. We use two different sets of instruments, C_{it-2} and C_{it-3}, in both PQRIV and TSLS estimation. The results are essentially the same, so we report the estimations for the former instrument.[16] In this model it is very important to take into account time invariant unobserved heterogeneity across households (fixed effects) as stressed by Carrasco, Labeaga, and Lopez-Salido (2005), and Browning and Collado (2007). Individual effects allow us to investigate whether the relationship between current and past consumption reflects habits or heterogeneity. Model (19) is very interesting because it shows how the coefficient of interest, $\alpha(\tau)$, varies along the quantiles, controlling for the individual intercepts.

We are interested in testing $H_0 : \alpha(\tau) = 0$, that is, there is no evidence of habit persistence. Since equation (19) captures the dynamics of consumption, the estimation results will not only serve as a test of this particular model, but will also provide evidence regarding the general importance of habit formation and the determinants of growth in consumption, as well as show evidence of asymmetric persistence in these variables. Point estimates and

[16] We also estimated the model using dummies for income growth as instruments, where income is measured as real labor income of head and wife or real disposable income. The results remain the same if we use this additional instrument.

standard errors for PQRIV and TSLS are presented in Tables 6 and 7. The standard errors appear inside parentheses. Table 6 presents the estimates for the autoregressive coefficient, family size, age, and age squared. In Table 7 we also include race as an additional regressor.

Concerning the TSLS results in Table 6, the point estimate for the autoregressive coefficient is positive, initially indicating evidence of habit formation. However, the results show that $\hat{\alpha}$ is not statistically different from zero even at the 10% level of significance. Thus, based on ordinary least squares estimation there is no evidence of habit formation in consumption; this result is in line with Dynan (2000). As expected, the coefficient of number of adults is positive and statistically significant. In addition, the coefficients of age and age squared are negative and positive, respectively, with the first not statistically different from zero. In Table 7 we add race as an exogenous regressor, and the results are still the same for these four coefficients. In addition, the race coefficient is not statistically different from zero in the ordinary least squares estimation.

Regarding the quantile regression estimation, Table 6 shows that the autoregressive coefficient has an asymmetric impact on consumption growth along the quantiles, ranging from 0.0155 to 0.0681. For the median, as in the least squares case, with coefficient 0.0333 and standard error 0.0225, there is no evidence of habit formation. However, the estimated coefficient is positive and statistically significant for the last two deciles, with point estimates of 0.0526 and 0.0681, and standard errors respectively 0.0236 and 0.0306. Thus, both coefficients are statistically significant at the 5% level of significance. In the high part of the conditional quantile function of consumption growth, for given t, the lag of consumption growth has a significant positive impact on subsequent consumption growth, indicating evidence of habit persistence in this part of the conditional quantile function. In summary, Table 6 shows that based on the estimates of the quantile regression instrumental variables model there is evidence against the null of no effect of previous consumption growth in current consumption growth, $\alpha(\tau) = 0$, for high quantiles, and the results also show a new feature that the persistence of consumption has an asymmetric behavior.

Thus, rejecting the null hypothesis of no habit formation for high quantiles indicates that for upper conditional quantile functions of consumption growth, where for given time period t the percentage difference between current expenditure and lagged expenditure is large, there is evidence of habit formation. Habit formation is then associated with large values of consumption expenditure growth. In periods of high consumption growth agents might experience a permanent positive income shock or lack of a credit constraint, creating habit persistence in preferences. However, for the middle and lower conditional quantile functions of consumption growth, where the consumption growth is close to zero or negative, for a given time period t, there is no evidence of habit formation. Therefore, when current consumption expenditure is similar to the lagged expenditure, agents might be subject to a credit constraint (or face a permanent negative income shock) and adjust consumption immediately from one period to another such that preferences do not exhibit habit persistence.

The other columns of Table 6 show the coefficients and standard errors for the number of adult male equivalents, age of the head of the household, and age-squared. The coefficient on adults is positive and highly significant, and it has a parabolic behavior along the quantiles. For families in both low and high quantiles, an increase in the number of adult males impacts the growth of consumption more than for families around the median.

According to the study of Carroll and Summers (1991) one would expect a negative coefficient on age and a positive coefficient on age-squared. This phenomenon is known as the "hump-shaped" age-consumption profile. The results presented in Table 6 show that the coefficient on age is positive only for the first decile, and it is not statistically different from zero. For the other deciles the coefficient on age is negative and different from zero. The coefficient on age-squared is positive and increases along the quantiles. However, for the first and second deciles it is not statistically significant. This asymmetric behavior indicates the presence of very important heterogeneity features in the determinants of consumption. For high quantiles of conditional consumption growth, the coefficients are consistent with the usual "hump-shaped" age-consumption profile, where the increase in age induces an increase in consumption and there is a valley in consumption growth. However, for very low quantiles this behavior is absent.

In a second round of estimation, following the literature, we include race as an exogenous regressor in the model. The results are presented in Table 7. The signs and significance of the first four coefficients are still the same. An interesting result is that there is an asymmetric dynamic in the race impact on consumption growth. The coefficient is negative for almost all quantiles, turning positive only for high quantiles. Inclusion of this variable has no effect on the findings regarding habit formation. In addition, for low quantiles the estimate is negative and statistically different from zero indicating that in the low part of the conditional quantile function nonwhite households exhibit significantly lower rates of consumption growth. The results for the other coefficients are similar to those in Table 6.

These results, together with the point estimates reported in Tables 6-7, indicate that there is an asymmetric persistence in consumption growth dynamics and in its determinants, and appropriate models are needed to incorporate such behavior.

6 Conclusion

This paper studies estimation and inference in a quantile regression dynamic panel model with fixed effects. Quantile regression for dynamic panel data methods allow one to explore a range of conditional quantiles, thereby exposing a variety of forms of conditional heterogeneity under less restrictive distributional assumptions, permit to control for individual specific intercepts, and provide a framework for robust estimation and inference.

In econometrics, the modeling of dynamic relationships usually requires inclusion of lagged dependent explanatory variables. The standard approaches to estimate dynamic panel models with fixed effects are typically biased in the presence of lagged dependent variables as regressors. To reduce the dynamic bias in the quantile regression fixed effects estimator we suggest the use of the instrumental variables quantile regression method of Chernozhukov and Hansen (2006, 2008) along with lagged regressors as instruments. The instrumental variables strategy sharply reduces the bias in the resulting point estimates. We show that under some mild regularity conditions, notably that with $T \to \infty$ as $N \to \infty$ and $N^a/T \to 0$, for some $a > 0$, the estimator is consistent and asymptotically normal. In addition, we propose Wald and Kolmogorov-Smirnov (KS) type tests for general linear hypotheses and derive their respective limiting distributions.

Monte Carlo studies are conducted to evaluate the finite sample properties of instrumental variables estimator for several types of distributions. It is shown that the quantile regression fixed effects estimator is severely biased in the presence of lagged dependent variables, while the PQRIV essentially eliminates the bias even in short panels. In addition, the PQRIV approach has a better performance *vis-a-vis* IV ordinary least squares-based approach in terms of the bias and root mean square error of the estimators for non-Gaussian heavy-tailed distributions. We also investigate the size and power of the test statistics comparing PQRIV with OLS-IV. The results show that tests based on quantile regression result in large power gains, especially when innovations are non-Gaussian heavy-tailed.

We illustrate the methods by testing for the presence of time-nonseparability in utility using household consumption data from the Panel Study on Income Dynamics (PSID) dataset. Previous work, such as Dynan (2000), find no evidence of habit formation using the PSID dataset and ordinary least squares for estimation at annual frequency. However, quantile regression methods reveal important heterogeneity associated with economic agents' behavior in terms of their dynamic consumption growth, which is averaged out by least squares estimators. At the same time, dynamic panel data quantile regression allows to control for individual specific effects in the dynamic panel data context. Our results show evidence of asymmetric persistence in consumption dynamics in the upper quantiles of the conditional distribution. In addition, it is possible to reject the null hypothesis of no effect of past consumption growth on subsequent consumption growth for these quantiles. Thus, for the upper conditional quantile functions of consumption growth, the results suggest that an increase in current consumption growth leads to increases in subsequent consumption growth, and for these corresponding quantiles there is evidence of habit persistence. Moreover, the results show important evidence of heterogeneity in the determinants of consumption such as the number of adult male equivalents in the household, age of the household head, and race. If economic dynamic panel data displays asymmetric dynamics systematically, then appropriate models are required to incorporate such behavior.

A Appendix 1

A.1 Analogy between IV Quantile regression and TSLS

Regular OLS Case

Consider the model

$$y = X\beta + Z\alpha + u$$

where Z is an endogenous variable and X are exogenous covariates. Let W be a valid instrument for Z. In the two stage least squares procedure

$$\hat{\alpha} = (Z' P_{M_X W} Z)^{-1} Z' P_{M_X W} y, \tag{20}$$

where $P_{M_X W} = M_X W (W' M_X W)^{-1} W' M_X$. Note that if the number of columns in Z is the same as in W, then $W' M_X Z$ is an invertible matrix and we can simplify (20) to

$$\hat{\alpha} = (W' M_X Z)^{-1} W' M_X y. \tag{21}$$

Grid Case

Now we consider the Grid case, in which one estimates the parameter of interest using a grid search. The model is the same

$$y = X\beta + Z\alpha + u,$$

where Z is endogenous, and X are the exogenous covariates. Let $\hat{W}$ be an instrument for Z, defined as the projection of Z on X and W,

$$\hat{W} := \hat{Z} = X\hat{\theta} + W\hat{\delta}.$$

Define a grid for α, $\{\alpha_j, j = 1, ..., J; |\alpha| < 1\}$, and for given α_j, consider the following regression

$$y - Z\alpha_j = X\beta + \hat{W}\gamma + v.$$

The estimator $\hat{\gamma}(\alpha_j)$ is

$$\hat{\gamma}(\alpha_j) = (\hat{W}' M_X \hat{W})^{-1} \hat{W}' M_X (y - Z\alpha_j).$$

Now define the estimator of α as

$$\hat{\alpha} = \arg\min_{\alpha} \|\hat{\gamma}(\alpha)\|_A,$$

where $\|\hat{\gamma}(\alpha)\|_A = \hat{\gamma}(\alpha)' A \hat{\gamma}(\alpha)$, $\hat{\gamma}(\alpha)$ is the vector containing the estimates $\hat{\gamma}(\alpha_j)$, and A is the identity matrix. Finally, noticing that $M_X \hat{W} = M_X W (W' M_X W)^{-1} W' M_X Z$, and from the first order condition yields

$$\hat{\alpha} = (Z' M_X W (W' M_X W)^{-1} W' M_X Z)^{-1} Z' M_X W (W' M_X W)^{-1} W' M_X y.$$

Using the definition of $P_{M_X W}$ we have

$$\hat{\alpha} = (Z' P_{M_X W} Z)^{-1} Z' P_{M_X W} y, \tag{22}$$

which is the same estimator as in equation (20). When the model is exactly identified

$$\hat{\alpha} = (W' M_X Z)^{-1} W' M_X y, \tag{23}$$

which is the same estimator as in equation (21).

A.2 Proof of the Theorems

The next three lemmas help in the derivation of the results. Lemma 1 shows the identification of the parameters. Lemma 2 guarantees a law of large numbers. Finally, Lemma 3 states an *Argmax Process* argument that is helpful in the derivation of consistency. Later we show consistency of $\hat{\theta}(\tau)$.

Lemma 1. Given assumptions A1-A6, $(\eta, \beta(\tau), \alpha(\tau))$ uniquely solves the equations $E[\upsilon\psi(Y - Z\eta - Y_{-1}\alpha - X\beta)\check{X}(\tau)] = 0$ over $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$.

Proof. We want to show that $(\eta, \alpha(\tau), \beta(\tau))$ uniquely solves the limit problem for each τ, that is, $\eta(\alpha^*(\tau)) = \eta$, $\alpha^*(\tau) = \alpha(\tau)$, and $\beta(\alpha^*(\tau), \tau) = \beta(\tau)$. Let $\psi(u) := (\tau - I(u < 0))$. Define:

$$\Pi(\eta, \alpha, \beta, \tau) := E[\upsilon\psi(Y - Z\eta - Y_{-1}\alpha(\tau) - X\beta(\tau))\check{X}(\tau)]$$
$$H(\eta, \alpha, \beta, \tau) := \frac{\partial}{\partial(\eta, \alpha, \beta)} E[\upsilon\psi(Y - Z\eta - Y_{-1}\alpha(\tau) - X\beta(\tau))\check{X}(\tau)]$$

By assumption A2, $H(\eta, \alpha, \beta, \tau)$ is continuous in (η, α, β) and full rank, uniformly over $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$. Moreover, by A2, the image of the set $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$ under the mapping $(\eta, \alpha, \beta) \to \Pi(\eta, \alpha, \beta, \tau)$ is assumed to be simply connected. As in Chernozhukov and Hansen (2005), the application of Hadamard's global univalence theorem for general metric spaces, e.g., Theorem 1.8 in Ambrosetti and Prodi (1995), yields that the mapping $\Pi(\cdot, \cdot, \cdot, \tau)$ is a homeomorphism (one-to-one) between $(\mathscr{E} \times \mathscr{A} \times \mathscr{B})$ and $\Pi(\mathscr{E}, \mathscr{A}, \mathscr{B}, \tau)$, the image of $\mathscr{E} \times \mathscr{A} \times \mathscr{B}$ under $\Pi(\cdot, \cdot, \cdot, \tau)$. Since $(\eta, \alpha, \beta) = (\eta, \alpha(\tau), \beta(\tau))$ solves the equation $\Pi(\eta, \alpha, \beta, \tau) = 0$; and thus it is the only solution in $(\mathscr{E} \times \mathscr{A} \times \mathscr{B})$. This argument is valid for $\tau \in \mathscr{T}$. So, we have that the true parameters $(\eta, \alpha, \beta) = (\eta, \alpha(\tau), \beta(\tau))$ uniquely solve the equation

$$E[\upsilon\psi(Y - Z\eta - Y_{-1}\alpha - X\beta - W0)\check{X}(\tau)] = 0. \tag{24}$$

Define $\vartheta \equiv (\eta, \beta, \gamma)$. By assumption and in view of the global convexity of $Q(\alpha, \vartheta, \tau)$ in ϑ for all τ and α, $\vartheta(\alpha, \tau)$ is defined by the subgradient condition

$$E[\upsilon\psi(Y - Z\eta(\alpha,\tau) - Y_{-1}\alpha - X\beta(\alpha,\tau) - W\gamma(\alpha,\tau))\check{X}(\tau)]v \geq 0 \tag{25}$$

$$\text{for all} \quad v : \vartheta(\alpha, \tau) + v \in \mathscr{E} \times \mathscr{B} \times \mathscr{G}.$$

In fact, if $\vartheta(\alpha, \tau)$ is in the interior of $\mathscr{E} \times \mathscr{B} \times \mathscr{G}$, it uniquely solves the first order condition version of (25)

$$E[\upsilon\psi(Y - Z\eta(\alpha,\tau) - Y_{-1}\alpha - X\beta(\alpha,\tau) - W\gamma(\alpha,\tau))\check{X}(\tau)] = 0. \tag{26}$$

We need to find $\alpha^*(\tau)$ by minimizing $\|\gamma(\alpha, \tau)\|$ over α subject to (25) holding. By (24) $\alpha^*(\tau) = \alpha(\tau)$ makes $\|\gamma(\alpha^*, \tau)\| = 0$ and satisfies (26) and hence (25) at the same time. According to the preceding argument, it is the only such solution. Thus, also by (26) $(\eta(\alpha^*(\tau)) = \eta, \beta(\alpha^*(\tau), \tau) = \beta(\tau))$. ∎

Lemma 2. Let $\xi_{it}(\tau_k) = \eta z_{it} + \alpha(\tau_k) y_{it-1} + \beta(\tau_k) x_{it} + \gamma(\tau_k) w_{it}$, and $u_{it}(\tau_k) = y_{it} - \xi_{it}(\tau_k)$. Let $\vartheta := (\eta, \alpha, \beta, \gamma)$ be a parameter vector in $\mathscr{V} := \mathscr{E} \times \mathscr{A} \times \mathscr{B} \times \mathscr{G}$. Let

$$\delta = \begin{pmatrix} \delta_\eta \\ \delta_\alpha \\ \delta_\beta \\ \delta_\gamma \end{pmatrix} = \begin{pmatrix} \sqrt{T}(\hat{\eta} - \eta) \\ \sqrt{NT}(\hat{\alpha}(\tau_k) - \alpha(\tau_k)) \\ \sqrt{NT}(\hat{\beta}(\tau_k) - \beta(\tau_k)) \\ \sqrt{NT}(\hat{\gamma}(\tau_k) - \gamma(\tau_k)), \end{pmatrix}.$$

Under conditions A1-A6,

$$\sup_{\vartheta \in \mathscr{V}} (NT)^{-1} \Big| \sum_k \sum_i \sum_t \Big[\rho_\tau \Big(u_{it}(\tau_k) - \frac{z_{it}\delta_\eta}{\sqrt{T}} - \frac{y_{it-1}\delta_\alpha}{\sqrt{NT}} - \frac{x_{it}\delta_\beta}{\sqrt{NT}} - \frac{w_{it}\delta_\gamma}{\sqrt{NT}} \Big) - \rho_\tau \Big(u_{it}(\tau_k) \Big) \tag{27}$$
$$- E\Big[\rho_\tau \Big(u_{it}(\tau_k) - \frac{z_{it}\delta_\eta}{\sqrt{T}} - \frac{y_{it-1}\delta_\alpha}{\sqrt{NT}} - \frac{x_{it}\delta_\beta}{\sqrt{NT}} - \frac{w_{it}\delta_\gamma}{\sqrt{NT}} \Big) - \rho_\tau \Big(u_{it}(\tau_k) \Big) \Big] \Big] \Big| = o_p(1).$$

Proof. With some abuse of notation let $\tilde{x}_{it} = (y_{it-1}, x_{it}, w_{it})$, $\tilde{\beta} = (\alpha, \beta, \gamma)$, and $\vartheta = (\tilde{\beta}, \eta)$. Let $\|\cdot\|$ denote the Euclidean norm. It is sufficient to show that for any $\epsilon > 0$

$$
\begin{aligned}
P\Big(\sup_{\vartheta \in \mathscr{V}} (NT)^{-1} \Big| \sum_k \sum_i \sum_t \Big[&\rho(u_{it} - \tilde{x}_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it}) \\
&- E[\rho(u_{it} - \tilde{x}_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it})] \Big] \Big| > \epsilon \Big) \to 0.
\end{aligned} \tag{28}
$$

For each k, consider a partition the parameter space $\Gamma = \mathscr{V}$ into K_N disjoint parts $\Gamma_1, \Gamma_2, \ldots, \Gamma_{K_N}$, such that the diameter of each part is less than $q_{NT} = \frac{\epsilon N^a}{12 K C_1 T}$. Let p_N be the dimension of ϑ, $p_N = O(N)$, and then $K_N \leq (2\sqrt{p_N}/q_{NT} + 1)^{p_N}$ (c.f. Wei and He (2006)).

Let $\zeta_i \in \Gamma_i, i = 1, \ldots, K_N$ be fixed points. Then the left-hand side of (28) can be bounded by $P_1 + P_2$, where

$$
\begin{aligned}
P_1 = P\Big(\max_{1 \leq q \leq K_N} \sup_{\vartheta \in \Gamma_q} (NT)^{-1} \Big| \sum_k \sum_i \sum_t \Big[&\rho(u_{it} - \tilde{x}_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it}) \\
&- \rho(u_{it} - \tilde{x}_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) + \rho(u_{it}) \\
&- E[\rho(u_{it} - \tilde{x}_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it})] \\
&+ E[\rho(u_{it} - \tilde{x}_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho(u_{it})] \Big] \Big| > \epsilon/2 \Big)
\end{aligned}
$$

and

$$
\begin{aligned}
P_2 = P\Big(\max_{1 \leq q \leq K_N} (NT)^{-1} \Big| \sum_k \sum_i \sum_t \Big[&\rho(u_{it} - \tilde{x}_{it} \frac{\zeta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\zeta_\eta}{\sqrt{T}}) + \rho(u_{it}) \\
&- E[\rho(u_{it} - \tilde{x}_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho(u_{it})] \Big] \Big| > \epsilon/2 \Big).
\end{aligned}
$$

Therefore, it suffices to show that both P_1 and P_2 are $o_p(1)$. Noting that $|\rho(x+y) - \rho(x)| \leq$

$2|y|$, we have

$$\max_{1\le q\le K_N}\sup_{\vartheta\in\Gamma_q}(NT)^{-1}\Big|\sum_k\sum_i\sum_t\Big[\rho(u_{it}-\tilde{x}_{it}\frac{\delta_{\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\delta_\eta}{\sqrt{T}})-\rho(u_{it})$$

$$-\rho(u_{it}-\tilde{x}_{it}\frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\zeta_{q\eta}}{\sqrt{T}})+\rho(u_{it})$$

$$-E[\rho(u_{it}-\tilde{x}_{it}\frac{\delta_{\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\delta_\eta}{\sqrt{T}})-\rho(u_{it})]$$

$$+E[\rho(u_{it}-\tilde{x}_{it}\frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\zeta_{q\eta}}{\sqrt{T}})-\rho(u_{it})]\Big]\Big|$$

$$\le \max_{1\le q\le K_N}\sup_{\vartheta\in\Gamma_q}(NT)^{-1}\Big[4\Big|\sum_k\sum_i\sum_t[\frac{\tilde{x}_{it}}{\sqrt{NT}}(\delta_{\tilde{\beta}}-\zeta_{q\tilde{\beta}})-\frac{z_{it}}{\sqrt{T}}(\delta_\eta-\zeta_{q\eta})]\Big|$$

$$+2\Big|\sum_k\sum_i\sum_t E[\frac{\tilde{x}_{it}}{\sqrt{NT}}(\delta_{\tilde{\beta}}-\zeta_{q\tilde{\beta}})-\frac{z_{it}}{\sqrt{T}}(\delta_\eta-\zeta_{q\eta})]\Big|\Big]$$

$$\le \max_{1\le q\le K_N}\sup_{\vartheta\in\Gamma_q}(NT)^{-1}\Big[4\sum_k\sum_i\sum_t[\frac{1}{\sqrt{NT}}\|\tilde{x}_{it}\|\|(\delta_{\tilde{\beta}}-\zeta_{q\tilde{\beta}})\|-\frac{1}{\sqrt{T}}\|z_{it}\|\|(\delta_\eta-\zeta_{q\eta})\|]$$

$$+2\sum_k\sum_i\sum_t E[\frac{1}{\sqrt{NT}}\|\tilde{x}_{it}\|\|(\delta_{\tilde{\beta}}-\zeta_{q\tilde{\beta}})\|-\frac{1}{\sqrt{T}}\|z_{it}\|\|(\delta_\eta-\zeta_{q\eta})\|]$$

$$\le\Big[4q_{NT}+2E[q_{NT}]\Big]\le\frac{\epsilon}{2},$$

where the last equality follows from assumptions A5-A6, and implies $P_1=o_p(1)$. It remains to show that $P_2=o_p(1)$.

If we write $m_{it}=\sum_k[\rho(u_{it}-\tilde{x}_{it}\frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\zeta_{q\eta}}{\sqrt{T}})-\rho(u_{it})]$ then

$$P_2=P\Big(\max_{1\le q\le K_N}(NT)^{-1}\Big|\sum_i\sum_t[m_{it}-Em_{it}]\Big|>\epsilon/2\Big)$$

$$\le\sum_{q=1}^{K_N}P\Big((NT)^{-1}\Big|\sum_i\Big[\sum_t m_{it}-E\sum_t m_{it}\Big]\Big|>\epsilon/2\Big).$$

For fixed i,

$$m_{it}\le\sup_{\zeta\in\Gamma}\Big|\sum_k\Big[\rho(u_{it}-\tilde{x}_{it}\frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}}-z_{it}\frac{\zeta_{q\eta}}{\sqrt{T}})-\rho(u_{it})\Big]\Big|$$

$$\le\sup_{\zeta\in\Gamma}2\Big|\sum_k\Big[\tilde{x}_{it}\frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}}+z_{it}\frac{\zeta_{q\eta}}{\sqrt{T}}\Big]\Big|$$

$$\le\sup_{\zeta\in\Gamma}2\sum_k\Big[\frac{1}{\sqrt{NT}}\|\tilde{x}_{it}\|\|\zeta_{q\tilde{\beta}}\|+\frac{1}{\sqrt{T}}\|z_{it}\|\|\zeta_{q\eta}\|\Big]\le 2C_2,$$

and,

$$\begin{aligned}
Var(\sum_{t=1}^{T} \mathrm{m}_{it}) &\le \sum_{t=1}^{T} Var(\mathrm{m}_{it}) + 2\sum_{t_1<t_2} Cov(\mathrm{m}_{it_1}, \mathrm{m}_{it_2}) \\
&\le \sum_{t=1}^{T} Var(\mathrm{m}_{it}) + 2\sum_{t_1<t_2} \sqrt{Var(\mathrm{m}_{it_1})Var(\mathrm{m}_{it_2})} \\
&\le \sum_{t=1}^{T} Var(\mathrm{m}_{it}) + \sum_{t_1<t_2} [Var(\mathrm{m}_{it_1}) + Var(\mathrm{m}_{it_2})] \\
&\le T\sum_{t=1}^{T} Var(\mathrm{m}_{it}).
\end{aligned}$$

Note that,

$$|\frac{\tilde{x}_{it}}{\sqrt{NT}}\zeta_{q\tilde{\beta}} + \frac{z_{it}}{\sqrt{T}}\zeta_{q\eta}|^2 \le C_3\|\zeta_{q\tilde{\beta}}\| + C_4\|\zeta_{q\eta}\|$$

Using the fact that $\|\zeta_q\| < C$ for any q, and the independence across individuals in A1, we have

$$Var\sum_{i=1}^{N}(\sum_{t=1}^{T}\mathrm{m}_{it}) \le T\sum_{i=1}^{N}\sum_{t=1}^{T} Ey_{it}^2 \le NT^2C_5 + o(1)$$

By the assumption of independence, we can bound P_2 using Bernstein's inequality:

$$\begin{aligned}
P_2 &\le 2K_N exp\Big(-\frac{\left(\frac{\epsilon NT}{2}\right)^2}{\sum_i Var\sum_t m_{it} + \frac{TC_2}{3}\left(\frac{\epsilon NT}{2}\right)}\Big) \\
&\le 2(2\sqrt{p_N}/q_{NT}+1)^{p_N} exp\Big(-\frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2C_5 + o(1) + \frac{TC_2}{3}\left(\frac{\epsilon NT}{2}\right)}\Big) \\
&= 2exp\Big(p_N\ ln(2\sqrt{p_N}/q_{NT}+1) - \frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2C_5 + o(1) + \frac{TC_2}{3}\left(\frac{\epsilon NT}{2}\right)}\Big) \\
&= 2exp\Big(O(N)\ ln(24O(N^{1/2-a}T)C_1/\epsilon + 1) - \frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2C_5 + o(1) + \frac{TC_2}{3}\left(\frac{\epsilon NT}{2}\right)}\Big)
\end{aligned}$$

Hence, $P_2 \to 0$ as $N, T \to \infty$ and $N^a/T \to 0$ and the lemma follows. ■

Lemma 3 (Corollary 3.2.3 - van der Vaart and Wellner (1996)). Let M_n be stochastic processes indexed by a metric space Θ, and let $M : \Theta \to \Re$ be a deterministic function. Suppose that $\|M_n - M\|_\Theta \xrightarrow{p} 0$ and that there exists a point θ_0 such that

$$M(\theta_0) > \sup_{\theta \notin D} M(\theta)$$

for every open set D that contains θ_0. Then any sequence $\hat{\theta}_n$, such that $Q_n(\hat{\theta}_n) > \sup_\theta Q_n(\theta) - o_p(1)$, satisfies $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Let $\theta(\tau) = (\theta(\tau_1)', ..., \theta(\tau_K)')'$. Now we prove Theorem 1 (Identification and Consistency).

Proof. The first part of the theorem follows from Lemma 1. For the second part, we need to show that under conditions A1-A6

$$\hat{\theta}(\tau) = \theta(\tau) + o_p(1).$$

Define

$$\mathscr{P} = (\eta, \alpha, \beta, \gamma) \to \rho_\tau(Y - Z\eta - Y_{-1}\alpha - X\beta - W\gamma)$$

and note that $\mathscr{P}$ is continuous. Therefore, uniform convergence is given by Lemma 2, such that

$$\sup_{\theta \in \Theta} |M_n - M| = o_p(1)$$

where $M_n \equiv \frac{1}{nT} \sum_{k=1}^K \sum_{i=1}^n \sum_{t=1}^T \rho(u_{it} - \tilde{x}_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - z_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it})$, and $M = EM_n$.

Therefore, denoting $\vartheta = (\eta, \beta, \gamma)$, by Lemma 3 we have that $\|\hat{\vartheta}(\alpha, \tau) - \vartheta(\alpha, \tau)\| \xrightarrow{p} 0$ (*) which implies that $\|\|\hat{\gamma}(\alpha, \tau)\| - \|\gamma(\alpha, \tau)\|\| \xrightarrow{p} 0$, which by a simple Argmax process over a compact set argument (assumption A4) implies that $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \xrightarrow{p} 0$, which by (*) implies that $\|\hat{\beta}(\tau) - \beta(\tau)\| \xrightarrow{p} 0$ and $\|\hat{\gamma}(\hat{\alpha}, \tau) - 0\| \xrightarrow{p} 0$. Therefore, $\|\hat{\theta}(\tau) - \theta(\tau)\| \xrightarrow{p} 0$ and the theorem follows. ■

Now we prove Theorem 2, i.e., under conditions A1-A6 the proposed estimator is asymptotically Normal.

Proof. Consider the following model

$$y_{it} = \eta_i + \alpha y_{it-1} + \beta x_{it} + u_{it}. \tag{29}$$

The objective function is

$$\min_{\eta_i, \alpha, \beta, \gamma} \sum_{k=1}^K \sum_{i=1}^N \sum_{t=1}^T \upsilon_k \rho_\tau(y_{it} - \eta_i - \alpha(\tau_k) y_{it-1} - \beta(\tau_k) x_{it} - \gamma(\tau_k) w_{it}). \tag{30}$$

Consider a collection of closed balls $B_n(\alpha(\tau_k))$, centered at $\alpha(\tau_k)$, with radius π_n, and $\pi_n \to 0$ slowly enough. Note that, for any $\alpha_n(\tau_k) \xrightarrow{p} \alpha(\tau_k)(\delta_\alpha \xrightarrow{p} 0)$ we can write the objective function as

$$V_{NT}(\delta) = \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T} \upsilon_k \rho_\tau(y_{it} - \xi_{it}(\tau_k) - z_{it}\delta_\eta/\sqrt{T} - y_{it-1}\delta_\alpha/\sqrt{NT} - x_{it}\delta_\beta/\sqrt{NT} - w_{it}\delta_\gamma/\sqrt{NT}) - \upsilon_k\rho_\tau(y_{it} - \xi_{it}(\tau_k)) \tag{31}$$

where $\xi_{it}(\tau_k) = \eta_i + \alpha(\tau_k)y_{it-1} + \beta(\tau_k)x_{it} + \gamma(\tau_k)w_{it}$, and

$$\hat{\delta}_n = \begin{pmatrix} \hat{\delta}_\eta \\ \hat{\delta}_\alpha \\ \hat{\delta}_\beta \\ \hat{\delta}_\gamma \end{pmatrix} = \begin{pmatrix} \sqrt{T}(\hat{\eta}(\alpha_n) - \eta) \\ \sqrt{NT}(\alpha_n(\tau_k) - \alpha(\tau_k)) \\ \sqrt{NT}(\hat{\beta}(\alpha_n, \tau_k) - \beta(\tau_k)) \\ \sqrt{NT}(\hat{\gamma}(\alpha_n, \tau_k) - 0) \end{pmatrix}.$$

Note that, for fixed $(\delta_\alpha, \delta_\beta, \delta_\gamma)$, we can consider the behavior of δ_η. Let $\psi(u) \equiv (\tau - I(u < 0))$ and for each i

$$g_i(\delta_\eta, \delta_\alpha, \delta_\beta, \delta_\gamma) = -\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k \psi_\tau(y_{it} - \xi_{it} - \frac{\delta_\eta}{\sqrt{T}} - \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} - \frac{\delta_\beta}{\sqrt{NT}}x_{it} - \frac{\delta_\gamma}{\sqrt{NT}}w_{it}).$$

For, fixed $\delta_\beta, \delta_\gamma$, $\sup_{\tau\in t} \|\alpha_n(\tau) - \alpha(\tau)\| \xrightarrow{p} 0$, and $K > 0$

$$\sup_{\|\delta\|<K} \|g_i(\delta_{\eta_i}, \delta_\alpha, \delta_\beta, \delta_\gamma) - g_i(0,0,0,0) - E[g_i(\delta_\eta, \delta_\alpha, \delta_\beta, \delta_\gamma) - g_i(0,0,0,0)]\| = o_p(1).$$

Expanding we have

$E[g_i(\delta_\eta, \delta_\alpha, \delta_\beta, \delta_\gamma) - g_i(0,0,0,0)]$

$$\begin{aligned}
&= E\Big(-\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k \psi_\tau(y_{it} - \xi_{it} - \frac{\delta_\eta}{\sqrt{T}} - \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} - \frac{\delta_\beta}{\sqrt{NT}}x_{it} - \frac{\delta_\gamma}{\sqrt{NT}}w_{it}) \\
&\qquad + \frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k \psi_\tau(y_{it} - \xi_{it})\Big) \\
&= -\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k \left[\tau_k - F(\xi_{it} + \frac{\delta_\eta}{\sqrt{T}} + \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it})\right] \\
&= \frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k \left[F(\xi_{it} + \frac{\delta_\eta}{\sqrt{T}} + \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it}) - \tau_k\right] \\
&= \frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T} \upsilon_k f_{it}(\xi_{it}(\tau_k)) \left[\frac{\delta_\eta}{\sqrt{T}} + \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right] + R_{it}.
\end{aligned}$$

Optimality of $\hat{\delta}_{\eta_i}$ implies that $g_i(\delta_\eta, \delta_\alpha, \delta_\beta, \delta_\gamma) = o(T^{-1})$, and thus

$$\hat{\delta}_\eta = -\overline{f}_i^{-1}\Big[\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k f_{it}(\xi_{it}(\tau_k))\left[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right]$$
$$-\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k\psi_\tau(y_{it} - \xi_{it}(\tau_k))\Big] + R_{it}$$

where $\overline{f}_{ik} = T^{-1}\sum_k\sum_t \upsilon_k f_{it}$ and R_{it} is the remainder term for each i.

Substituting $\hat{\delta}_\eta$'s, we denote

$$G(\delta_\alpha, \delta_\beta, \delta_\gamma) = -\frac{1}{\sqrt{NT}}\sum_{k}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\upsilon_k X_{it}\psi_\tau(y_{it} - \xi_{it} - \frac{\hat{\delta}_\eta}{\sqrt{T}} - \frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} - \frac{\delta_\beta}{\sqrt{NT}}x_{it} - \frac{\delta_\gamma}{\sqrt{NT}}w_{it})$$

where $X_{it} = (x_{it}', w_{it}')'$, and $\delta_\alpha(\alpha_n) = \sqrt{NT}(\alpha_n(\tau_k) - \alpha(\tau_k))$, $\delta_\beta(\alpha_n) = \sqrt{NT}(\hat{\beta}(\alpha_n, \tau_k) - \beta)$, and $\delta_\gamma(\alpha_n) = \sqrt{NT}(\hat{\gamma}(\alpha_n, \tau_k) - 0)$. According to Lemma B2 in Chernozhukov and Hansen (2006)

$$\sup_{\|\delta\|<K}\|g_i(\delta_\alpha, \delta_\beta, \delta_\gamma) - g_i(0,0,0) - E[g_i(\delta_\alpha, \delta_\beta, \delta_\gamma) - g_i(0,0,0)]\| = o_p(1),$$

and at the minimizer, $G(\hat{\delta}_\alpha, \hat{\delta}_\beta, \hat{\delta}_\gamma) = o((NT)^{-1})$. Expanding, as above,
$E[G(\delta_\alpha, \delta_\beta, \delta_\gamma) - G(0,0,0)]$

$$= \frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\left(\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it} + \frac{\hat{\delta}_\eta}{\sqrt{T}}\right) + o_p(1)$$
$$= \frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\Big[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it} - \frac{1}{\sqrt{T}}\overline{f}_i^{-1}\big[T^{-1/2}\sum_k\sum_t \upsilon_k f_{it}(\xi_{it}(\tau_k))$$
$$\left[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1} + \frac{\delta_\beta}{\sqrt{NT}}x_{it} + \frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right] - T^{-1/2}\sum_k\sum_{t=1}^{T}\upsilon_k\psi_\tau(y_{it} - \xi_{it}(\tau_k))\big] + R_{it}\Big] + o_p(1)$$
$$= \frac{1}{NT}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\left(y_{it-1} - \overline{f}_i^{-1}T^{-1}\sum_k\sum_t \upsilon_k f_{it} y_{it-1}\right)\delta_\alpha$$
$$+ \frac{1}{NT}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\left(x_{it} - \overline{f}_i^{-1}T^{-1}\sum_k\sum_t \upsilon_k f_{it} x_{it}\right)\delta_\beta$$
$$+ \frac{1}{NT}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\left(w_{it} - \overline{f}_i^{-1}T^{-1}\sum_k\sum_t \upsilon_k f_{it} w_{it}\right)\delta_\gamma$$
$$- \frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it}\overline{f}_i^{-1}T^{-1/2}\sum_k\sum_t \upsilon_k\psi(y_{it} - \xi_{it})$$
$$- \frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t \upsilon_k X_{it} f_{it} R_{it}/\sqrt{T} + o_p(1),$$

where the order of the final term is controlled by the bound on the derivative of the conditional density. It is important to note that $G(\hat{\delta}_\alpha, \hat{\delta}_\beta, \hat{\delta}_\gamma) = 0$ and then $E[G(\delta_\alpha, \delta_\beta, \delta_\gamma) - G(0,0,0)] = G(0,0,0)$, $\Phi = \mathrm{diag}(f_{it}(\xi_{it}(\tau_k)))$, and let Ψ_τ be a NT-vector $(\psi_\tau(y_{it} - \xi_{it}(\tau)))$. Define $\delta_\vartheta = (\delta_\beta, \delta_\gamma)$, then omitting the subscript k we have

$$\upsilon(X' M_Z \Phi M_Z Y_{-1}/NT)\delta_\alpha + \upsilon(X' M_Z \Phi M_Z X/NT)\delta_\vartheta - \upsilon X' P_Z \Psi/\sqrt{NT} = -\upsilon X'\Psi/\sqrt{NT} + R_{nT}$$
$$J_\vartheta \delta_\vartheta = [-(\upsilon X' M_Z \Psi/\sqrt{NT} - R_{nT}) - J_\alpha \delta_\alpha]$$
$$\hat{\delta}_\vartheta = J_\vartheta^{-1}[-(\upsilon X' M_Z \Psi/\sqrt{NT} - R_{nT}) - J_\alpha \delta_\alpha]$$

where

$$M_Z = I - P_Z$$
$$P_Z = Z(Z'\Phi Z)^{-1} Z'\Phi$$

$$J_\vartheta = \lim_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \upsilon_1 \tilde{X}' M'_{Z_1} \Phi(\tau_1) M_{Z_1} \tilde{X} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \upsilon_k \tilde{X}' M'_{Z_k} \Phi(\tau_k) M_{Z_k} \tilde{X} \end{pmatrix}$$
$$J_\alpha = \lim_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \upsilon_1 \tilde{X}' M'_{Z_1} \Phi(\tau_1) M_{Z_1} y_{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \upsilon_k \tilde{X}' M'_{Z_k} \Phi(\tau_k) M_{Z_k} y_{-1} \end{pmatrix},$$

and

$$R_{nT} = \frac{1}{\sqrt{NT}} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t=1}^{T} \upsilon_k X_{it} f_{it} R_{it}/\sqrt{T} + o_p(1).$$

Let $[\bar{J}'_\beta, \bar{J}'_\gamma]'$ be the conformable partition of J_ϑ^{-1}, then

$$\hat{\delta}_\gamma = \bar{J}_\gamma[-(\upsilon X' M_Z \Psi/\sqrt{NT} - R_{nT}) - J_\alpha \delta_\alpha] \tag{32}$$
$$\hat{\delta}_\beta = \bar{J}_\beta[-(\upsilon X' M_Z \Psi/\sqrt{NT} - R_{nT}) - J_\alpha \delta_\alpha]. \tag{33}$$

The remainder term R_{nT} has a dominant component that comes from the Bahadur representation of the η's. By A1 and A5, we have for a generic constant C_1

$$R_{nT} = T^{-1/4} \frac{C_1 K}{\sqrt{N}} \sum_{i=1}^{N} R_{i0} + o_p(1).$$

The analysis of Knight (2001) shows that the summands converge in distribution, that is as $T \to \infty$ the remainder term $T^{1/4} R_{it} \xrightarrow{d} R_{i0}$, where R_{i0} are functionals of Brownian motions with finite second moment. Therefore, independence of y_{it}, and condition A6 ensure that the contribution of the remainder is negligible. Thus (32) and (34) simplify to

$$\hat{\delta}_\gamma = \bar{J}_\gamma[-(\upsilon X' M_Z \Psi/\sqrt{NT}) - J_\alpha \delta_\alpha]$$
$$\hat{\delta}_\beta = \bar{J}_\beta[-(\upsilon X' M_Z \Psi/\sqrt{NT}) - J_\alpha \delta_\alpha].$$

By consistency, $wp \to 1$

$$\hat{\delta}_\alpha = \min_{\delta_\alpha \in B_n(\alpha(\tau))} \hat{\delta}_\gamma(\delta_\alpha)' A \hat{\delta}_\gamma(\delta_\alpha),$$

assuming that $\hat{\delta}_\gamma' A \hat{\delta}_\gamma$ is continuous in δ_α, and from the first order condition

$$\hat{\delta}_\alpha = -[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1}[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma(\upsilon X' M_Z \Psi/\sqrt{NT})].$$

Substituting $\hat{\delta}_\alpha$ back in δ_β we obtain

$$\begin{aligned}\hat{\delta}_\beta &= -\bar{J}_\beta[(X' M_Z \Psi/\sqrt{NT}) - J_\alpha[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma(\upsilon X' M_Z \Psi/\sqrt{NT})] \\ &= -\bar{J}_\beta[(I - J_\alpha[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma)(\upsilon X' M_Z \Psi/\sqrt{NT})].\end{aligned}$$

It is also important to analyze $\hat{\delta}_\gamma$. Thus, replacing $\hat{\delta}_\alpha$ in δ_γ

$$\begin{aligned}\hat{\delta}_\gamma &= -\bar{J}_\gamma[(X' M_Z \Psi/\sqrt{NT}) - J_\alpha[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma(\upsilon X' M_Z \Psi/\sqrt{NT})] \\ &= -\bar{J}_\gamma[(I - J_\alpha[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma)(\upsilon X' M_Z \Psi/\sqrt{NT})].\end{aligned}$$

By condition A3, using the fact that $\bar{J}_\gamma J_\alpha$ is invertible

$$\hat{\delta}_\gamma = 0 + O_p(1) + o_p(1).$$

Let $\Psi_k = \mathrm{diag}(\psi_{\tau_k}(y_{it} - \xi_{it}(\tau_k)))$. Notice that $\Psi_k 1_{NT} 1_{NT}' \Psi_l = (\tau_k \wedge \tau_l - \tau_k \tau_l) I_{NT}$, and that conditions A1-A6 imply a central limit theorem. Thus, neglecting the remainder term, and using the definition of δ_α and δ_β we have

$$\sqrt{NT}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega), \quad \Omega = (K', L')' S (K', L')$$

where $S = (\tau \wedge \tau' - \tau\tau') E(VV')$, $V = X' M_Z$, $K = (J_\alpha' H J_\alpha)^{-1} J_\alpha H$, $H = \bar{J}_\gamma' A[\alpha(\tau)] \bar{J}_\gamma$, $L = \bar{J}_\beta M$, $M = I - J_\alpha K$. ■

B Appendix 2

Household i chooses current consumption expenditure, c_{it}, to solve

$$\max_{\{c_{it}\}_{t=0}^{\infty}} E\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{it} - \alpha c_{it-1}; \psi_{it}\right)\right]$$

$$s.t = \begin{cases} a_{it+1} = r_{it+1}(a_{it} + y_{it} - c_{it}) \\ \text{given} \quad a_{i0} \end{cases},$$

where a_{it}, y_{it} and c_{it} are wealth, income, and consumption respectively. Since ψ_{it} is only a "taste-shifter" we will omit it for simplification reasons. The Bellman equation can be written as

$$V(a_{it}, c_{it-1}) = \max_{c_{it}, a_{it+1}} \{u(c_{it} - \alpha c_{it-1}) + \beta E_t[V(a_{it+1}, c_{it})]\}.$$

From the first-order condition

$$u'(c_{it} - \alpha c_{it-1}) + \beta E_t[V_2(a_{it+1}, c_{it})] = \beta E_t[V_1(a_{it+1}, c_{it}) r_{it+1}]. \tag{34}$$

By the envelope theorem

$$V_1(a_{it}, c_{it-1})] = \beta E_t[V_1(a_{it+1}, c_{it}) r_{it+1}] \tag{35}$$

$$V_2(a_{it}, c_{it-1})] = u'(c_{it} - \alpha c_{it-1})(-\alpha). \tag{36}$$

Combining (34), (35), and (36)

$$u'(c_{it} - \alpha c_{it-1}) - \alpha\beta E_t[u'(c_{it+1} - \alpha c_{it})] = \beta E_t[u'(c_{it+1} - \alpha c_{it}) r_{it+1}] - \alpha\beta^2 E_t[u'(c_{it+2} - \alpha c_{it+1}) r_{it+1}].$$

Therefore

$$E_t[MU_{it} - \alpha\beta MU_{it+1}] = E_t[r_{it+1}\beta MU_{it+1} - r_{it+1}\alpha\beta^2 MU_{it+2}], \tag{37}$$

where $MU_{it} = u'(c_{it} - \alpha c_{it-1})$. Assuming that r_{it} is constant and T is large, Dynan (2000) shows that (37) simplifies to

$$E_t\Big[r\beta \frac{MU_{it+1}}{MU_{it}}\Big] = 1.$$

Using the following utility function

$$u(c_{it} - \alpha c_{it-1}) = \psi_{it} \frac{(c_{it} - \alpha c_{it-1})^{1-\rho}}{1-\rho},$$

we have

$$r\beta \frac{\psi_{it}}{\psi_{it-1}} \Big(\frac{c_{it} - \alpha c_{it-1}}{c_{it-1} - \alpha c_{it-2}}\Big)^{-\rho} = 1 + \epsilon_{it}. \tag{38}$$

Taking natural logarithm of (38) and writing $\gamma_0 = \frac{1}{\rho}(\ln(r) + \ln(\beta))$

$$\Delta \ln(c_{it} - \alpha c_{it-1}) = \gamma_0 + \gamma_1 \Delta \ln(\psi_{it}) + \epsilon_{it}.$$

Finally, using the approximation $\Delta \ln(c_{it} - \alpha c_{it-1}) \approx (\Delta \ln c_{it} - \alpha \Delta \ln c_{it-1})$ we have equation (18).

References

Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 26, 379–397.

Ahn, S. C., and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

Alvarez, J., and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159.

Ambrosetti, A., and G. Prodi (1995): *A Primer of Nonlinear Analysis, Cambridge Studies in Advanced Mathematics, vol. 34*. Cambridge University Press, Cambridge, Massachusetts.

Anderson, T. W., and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Arellano, M. (2003): *Panel Data Econometrics*. Oxford University Press, New York.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–279.

Arellano, M., and J. Hahn (2005): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Mimeo, CEMFI.

Baltagi, B. H. (2008): "Forecasting with Panel Data," forthcoming in the Journal of Forecasting.

Bester, C. A., and C. Hansen (2007): "A Penalty Function Approach to Bias Reduction in Non-Linear Panel Models with Fixed Effects," *Journal of Business and Economic Statistics*, forthcoming.

Blundell, R., and S. R. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Bofinger, E. (1975): "Estimation of a Density Function Using Order Statistics," *Australian Journal of Statistics*, 17, 1–7.

Browning, M., and M. D. Collado (2007): "Habits and Heterogeneity in Demands: A Panel Data Analysis," *Journal of Applied Econometrics*, 22, 625–640.

Bun, M. J., and M. A. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23, 200–210.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit. A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Carrasco, R., J. M. Labeaga, and J. D. Lopez-Salido (2005): "Consumption and Habits: Evidence From Panel Data," *The Economic Journal*, 115, 144–165.

Carroll, C. D., J. Overland, and D. N. Weil (2000): "Saving and Growth with Habit Formation," *American Economic Review*, 90, 341–355.

Carroll, C. D., and L. H. Summers (1991): "Consumption Growth Parallels Income Growth: Some New Evidence," in *National Saving and Economic Performance*, ed. by B. Bernheim, and J. Shoven, pp. 305–343. University of Chicago Press.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effects Models," *Journal of Econometrics*, 132, 491–525.

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

Constantinides, G. M. (1990): "Habit Formation. A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98, 519–543.

DeLong, D. M. (1981): "Crossing Probabilities for a Square Root Boundary By a Bessel Process," *Communications in Statistics, Part A - Theory and Methods*, 10, 2197–2213.

Dunn, K. B., and K. J. Singleton (1986): "Modeling the Term Structure of Interest Rates under Non-separable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

Dynan, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data," *American Economic Review*, 90, 391–406.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Ferson, W. E., and G. M. Constantinides (1991): "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests," *Journal of Financial Economics*, 29, 199–240.

Fuhrer, J. C. (2000): "Habit Formation in Consumption and its Implications for Monetary Policy," *American Economic Review*, 90, 367–390.

Geraci, M., and M. Bottai (2007): "Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution," *Biostatistics*, 8, 140–154.

Gutenbrunner, C., and J. Jureckova (1992): "Regression Rank Scores and Regression Quantiles," *The Annals of Statistics*, 20, 305–330.

Hahn, J., and G. M. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large," *Econometrica*, 70, 1639–1657.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319.

Hall, R. E., and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461–481.

He, X., and Q.-M. Shao (2000): "On Parameters of Increasing Dimensions," *Journal of Multivariate Analysis*, 73, 120–135.

Heckman, J. J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by C. F. Manski, and D. McFadden, pp. 179–195. MIT Press.

Hsiao, C. (2003): *Analysis of Panel data*. Cambridge University Press, Cambridge, Massachusetts, 2d edn.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

Knight, K. (2001): "Comparing Conditional Quantile Estimators: First and Second order Considerations," Mimeo.

Koenker, R. (1984): "A Note on L-Estimates for Linear Models," *Statistics and Probability Letters*, 2, 323–325.

——— (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

Koenker, R., and G. W. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

——— (1982a): "Robust Tests for Heteroscedasticity Based on Regression Quantile," *Econometrica*, 50, 43–61.

——— (1982b): "Tests of Linear Hypothesis and l_1 Estimation," *Econometrica*, 50, 1577–1584.

Koenker, R., and K. Hallock (2000): "Quantile Regression an Introduction," Manuscript, University of Illinois at Urbana-Champaign.

Koenker, R., and J. A. F. Machado (1999): "Godness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1310.

Koenker, R., and Z. Xiao (2006): "Quantile Autoregression," *Journal of the American Statistical Association*, 101, 980–990.

KOENKER, R., AND Q. ZHAO (1996): "Conditional Quantile Estimation and Inference for Arch Models," *Econometric Theory*, 12, 793–813.

LAMARCHE, C. (2006): "Robust Penalized Quantile Regression Estimation for Panel Data," Mimeo.

——— (2008): "Penalized Quantile Regression Estimation for a Model with Endogenous Individual Effects," Mimeo.

MEGHIR, C., AND G. WEBER (1996): "Intertemporal Non-Separability or Borrowing Restrictions? A Disaggregate Analysis Using a US Consumption Panel," *Econometrica*, 64, 1151–1182.

NAIK, N. Y., AND M. J. MOORE (1996): "Habit Formation and Intertemporal Substitution in Individual Food Consumption," *Review of Economics and Statistics*, 78, 321–328.

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

PORTNOY, S. L. (1984): "Tightness of the Sequence of Empiric cdf Processes Defined from Regression Fractiles," in *Robust and Nonlinear Time Series Analysis*, ed. by J. Frank, W. Hardle, and D. Martin. Spring-Verlag: New York.

——— (1985): "Asymptotic Behavior of M Estimators of p Regression Parameters when p^2/n is Large; II. Normal Approximation," *The Annals of Statistics*, 13, 1403–1417.

POWELL, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

SHAPIRO, M. D. (1984): "The Permanent Income Hypothesis and the Real Interest Rate: Some Evidence from Panel Data," *Economics Letters*, 14, 93–100.

VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag Press, New York, New York.

WEI, Y., AND X. HE (2006): "Conditional Growth Charts," *The Annals of Statistics*, 34, 2069–2097.

Zeldes, S. P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97, 305–346.

Zhou, K. Q., and S. L. Portnoy (1996): "Direct Use of Regression Quantiles to Construct Confidence Sets in Linear Models," *The Annals of Statistics*, 24, 287–306.

Table 6: Results for PQRIV and TSLS

PQRIV				
Quantiles	Auto	Adults	Age	Age2
0.1	0.0155	0.0443	0.0009	-0.0724
	(0.031)	(0.005)	(0.003)	(0.291)
0.2	0.0246	0.0316	-0.0013	0.0942
	(0.026)	(0.004)	(0.002)	(0.249)
0.3	0.0287	0.0231	-0.0042	0.3598
	(0.027)	(0.004)	(0.002)	(0.255)
0.4	0.0373	0.0144	-0.0059	0.5109
	(0.027)	(0.004)	(0.002)	(0.260)
0.5	0.0333	0.0073	-0.0088	0.7750
	(0.023)	(0.003)	(0.002)	(0.197)
0.6	0.0304	0.0049	-0.0101	0.8782
	(0.023)	(0.003)	(0.002)	(0.208)
0.7	0.0383	0.0074	-0.0116	1.0177
	(0.025)	(0.004)	(0.002)	(0.241)
0.8	0.0526	0.0152	-0.0158	1.3970
	(0.024)	(0.005)	(0.003)	(0.308)
0.9	0.0681	0.0196	-0.0228	2.0872
	(0.031)	(0.005)	(0.003)	(0.300)
TSLS				
-	0.0286	0.0076	-0.0021	0.5013
	(0.026)	(0.003)	(0.002)	(0.217)

Notes: Standard errors in parenthesis. PQRIV: dynamic panel instrumental variables quantile regression fixed effects; TSLS: fixed effects least squares. Auto: autoregression coefficient, $\hat{\alpha}$, associated with C_{it-1}; Adults: number of adult male equivalents in the household; Age: age of the head of the household; Age2: age squared of the head of the household.

Table 7: Results for PQRIV and TSLS including race as a regressor

PQRIV					
Quantiles	Auto	Adults	Age	Age2	Race
0.1	0.0103	0.0450	0.0009	-0.0630	-0.0384
	(0.027)	(0.004)	(0.002)	(0.256)	(0.012)
0.2	0.0241	0.0330	-0.0013	0.0889	-0.0238
	(0.024)	(0.004)	(0.002)	(0.221)	(0.008)
0.3	0.0318	0.0229	-0.0041	0.3451	-0.0168
	(0.022)	(0.004)	(0.002)	(0.201)	(0.007)
0.4	0.0367	0.0144	-0.0059	0.5084	-0.0129
	(0.022)	(0.004)	(0.002)	(0.194)	(0.007)
0.5	0.0332	0.0074	-0.0086	0.7615	-0.0090
	(0.022)	(0.003)	(0.002)	(0.199)	(0.007)
0.6	0.0299	0.0008	-0.0102	0.8849	-0.0010
	(0.023)	(0.003)	(0.002)	(0.208)	(0.006)
0.7	0.0379	0.0074	-0.0115	1.0110	0.0027
	(0.024)	(0.004)	(0.002)	(0.238)	(0.007)
0.8	0.0520	0.0151	-0.0158	1.4039	0.0105
	(0.024)	(0.004)	(0.003)	(0.286)	(0.009)
0.9	0.0697	0.0201	-0.0235	2.1599	0.0180
	(0.030)	(0.004)	(0.003)	(0.288)	(0.010)
TSLS					
-	0.0328	0.0076	-0.0024	0.5301	-0.0010
	(0.025)	(0.003)	(0.002)	(0.228)	(0.007)

Notes: Standard errors in parenthesis. PQRIV: dynamic panel instrumental variables quantile regression fixed effects; TSLS: fixed effects least squares. Auto: autoregression coefficient, $\hat{\alpha}$, associated with C_{it-1}; Adults: number of adult male equivalents in the household; Age: age of the head of the household; Age2: age squared of the head of the household.

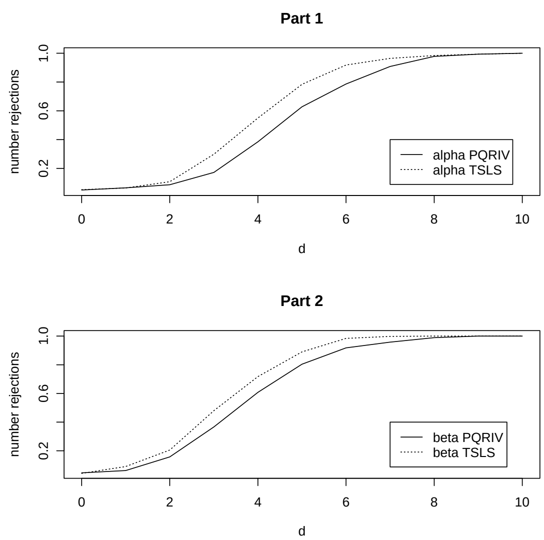

Figure 1: Asymptotic Power for Normal distribution

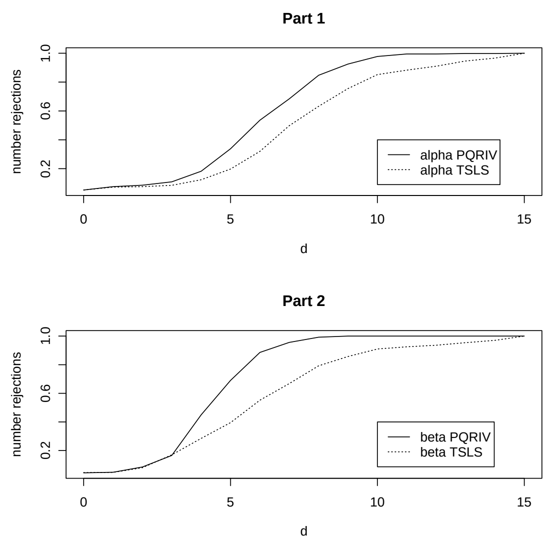

Figure 2: Asymptotic Power for t_3 distribution