
Volume 140
Issue 11
November 2010
ISSN 0378-3758
ELSEVIER
journal of statistical planning and inference
ScienceDirect

Journal of Statistical Planning and Inference 140 (2010) 3476–3497

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Penalized quantile regression for dynamic panel data

Antonio F. Galvao [a], Gabriel V. Montes-Rojas [b,*]

[a] Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201, USA
[b] Department of Economics, City University London, D306 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK

ARTICLE INFO

Article history:
Received 13 January 2010
Received in revised form
12 May 2010
Accepted 13 May 2010
Available online 28 May 2010

Keywords:
Panel data
Quantile regression
Fixed effects
Dynamic panel
Shrinkage
Penalized regression

ABSTRACT

This paper studies penalized quantile regression for dynamic panel data with fixed effects, where the penalty involves l_1 shrinkage of the fixed effects. Using extensive Monte Carlo simulations, we present evidence that the penalty term reduces the dynamic panel bias and increases the efficiency of the estimators. The underlying intuition is that there is no need to use instrumental variables for the lagged dependent variable in the dynamic panel data model without fixed effects. This provides an additional use for the shrinkage models, other than model selection and efficiency gains. We propose a Bayesian information criterion based estimator for the parameter that controls the degree of shrinkage. We illustrate the usefulness of the novel econometric technique by estimating a "target leverage" model that includes a speed of capital structure adjustment. Using the proposed penalized quantile regression model the estimates of the adjustment speeds lie between 3% and 44% across the quantiles, showing strong evidence that there is substantial heterogeneity in the speed of adjustment among firms.

1. Introduction

Very often researchers wish to use longitudinal data to estimate behavioral relationships that are dynamic in character, namely, models containing lagged dependent variables. However, the fixed effects (FE) regression models for panel data may be biased in the presence of lagged dependent variables as regressors when the time series dimension of the panel data is modest. In standard estimation of dynamic panel models consistency of the estimators depends critically on the assumptions about the initial conditions.[1] Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991) show that consistent estimators for dynamic panel data models can be obtained by using instrumental variables (IV) methods that produces estimators independent of the initial conditions.

Quantile regression (QR) for dynamic panel data offers a systematic strategy for examining how covariates influence the location, scale, and shape of the response distribution, revealing important heterogeneity in dynamic responses. Koenker (2004) proposes a general penalized approach to estimate static QR models for longitudinal data with FE, thus allowing one to control for unobserved individual heterogeneity. Galvao (2009) studies estimation and inference for dynamic QR panel data models with individual specific intercepts, proposing an IV strategy to estimate the model and reduce the bias in the absence of shrinkage.

* Corresponding author. Tel.: +44 20 70 40 89 19.
E-mail addresses: agalvao@uwm.edu (A.F. Galvao), Gabriel.Montes-Rojas.1@city.ac.uk (G.V. Montes-Rojas).

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

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2010.05.008

It is well known that the optimal estimator for the random effects Gaussian model involves shrinking the individual effects toward a common value.[2] When there is an intercept in the model this common value can be taken to be the conditional central tendency of the response at a point determined by the centering of the other covariates. In the QR model this would be some corresponding conditional quantile of the response. As argued in Koenker (2004), particularly when N is large relative to the T, shrinkage may be advantageous in controlling the variability introduced by the large number of estimated individual specific parameters. In addition, by shrinking the FE we are effectively reducing the number of parameters to be estimated and, hence, ameliorating the incidental parameters problem. For the quantile loss function, it is convenient to consider the l_1 penalty, in place of the conventional Gaussian penalty. This choice maintains the linear programming form of the problem and also preserves the sparsity of the resulting design matrix.[3] Several authors, notably Tibshirani (1996) and Donoho et al. (1998), have pointed out that l_1 shrinkage offers some statistical advantages over more conventional Gaussian l_2 penalties in addition to its computational advantages. Moreover, Wang et al. (2007a) note that in the least absolute deviation regression the l_1-penalty is able to provide parameter estimation and variable selection simultaneously.

In the panel data QR set-up, shrinkage models offer some computation advantages because the FE have to be estimated directly. The reliance on the existing bias reduction strategies using least-squares differencing, either temporally or via the usual deviation from individual means (within) transformation, is unsatisfactory in the QR setting. Linear transformations that are 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 that they commute with linear transformations; quantiles do not. This problem is also present in nonlinear panel models (see recent work by Bester and Hansen, 2009).

In standard IV estimation of dynamic panel data models the instruments become less informative in two important cases. First, when the autoregressive parameter increases toward unity. Even if the autoregressive coefficient is less than one, IV from lagged variables tend to be weak if the autoregressive coefficient is close to one. The second case is when the variance of the FE increases. In these cases, using the lagged dependent variable as IV tend to generate weak instruments (see, for instance Arellano, 2003, Chapter 6). Simulation studies show evidence that the usual estimators based on IV present large finite bias and poor precision. Our own simulations confirm that when the autoregressive coefficient goes toward unity the IV QR with FE estimator becomes severely biased. In this context, in order to improve the efficiency of the GMM estimator, Arellano and Bover (1995) and Ahn and Schmidt (1995) suggest additional orthogonality conditions and Blundell and Bond (1998) and Hahn (1999) suggest additional restrictions based in the steady-state distribution of the dependent variable. In addition, Hahn et al. (2007) propose an IV estimator for dynamic models when identification becomes weak near the unit circle.[4]

In this paper we study penalized QR estimators for dynamic panel data with FE, where the penalty involves l_1 shrinkage of the FE. Using extensive Monte Carlo simulations we present strong evidence that the penalty term reduces the dynamic bias and increases the efficiency of the QR estimators, because the regularization, or shrinkage, of the FE toward a common value can help to mollify the dynamic panel bias caused by the presence of the unobserved individual heterogeneity. The underlying intuition is that, in the limit, when we shrink the FE completely, there is no need to use IV for the lagged dependent variable in the dynamic panel data model. Thus, the results in this paper suggest an additional use for the shrinkage models, other than model selection and efficiency gains. In particular, the shrinkage helps in reducing potential bias in models with endogeneity and weak instruments. In our case, the QR dynamic panel data context, shrinking the FE may help to overcome the dynamic bias associated with the weak instruments problem. These results run parallel to those of Bester and Hansen (2009) and Arellano and Hahn (2006), where the penalized optimization approach substantially reduces the bias in nonlinear FE models.

In order to achieve this aim, we propose a new penalized panel data IV estimator that combines the IV QR method of Chernozhukov and Hansen (2006, 2008) with the l_1 shrinkage of the FE. We derive its consistency and asymptotic normality. Moreover, Monte Carlo simulations are conducted to study the finite sample properties of the proposed estimator. We compare the bias and root mean squared error (RMSE) of this estimator with the simple penalized QR panel data estimator of Koenker (2004) without IV. The results show that under some shrinkage both penalized estimators are able to substantially reduce the bias and RMSE relative to the estimators with zero shrinkage. In fact, surprisingly, a small penalty considerably reduces the bias in the QR model with FE and no IV.

In addition, it is well known that an appropriate value of the regularization parameter is crucial for the performance of the fitted model in any regularized model fitting. We discuss a Bayesian information criterion (BIC) based estimator for the parameter that controls the degree of shrinkage. In practice, one can compute the estimates of the model, then select the value that minimizes the model selection criterion as estimate of the tuning parameter. This avoids a more computationally intensive cross-validation approach. We show in this paper that this approach works in the IV QR

[2] See e.g. Robinson (1991) and Koenker (2004) for more details on shrinkage methods for panel data models.

[3] See Belloni and Chernozhukov (2009) for other properties of l_1-penalty models.

[4] More recently, a number of alternative approaches based on large panels asymptotic theory have been proposed to reduce the bias in dynamic and nonlinear panels. See, for example, Arellano and Hahn (2007) for a survey, and Hahn and Kuersteiner (2002, 2003), Alvarez and Arellano (2003), Hahn and Newey (2004), Arellano and Hahn (2006), Bester and Hansen (2009), for specific approaches.

models with FE. Moreover, the experiments show evidence that the BIC criterion is a suitable estimator for selecting the shrinkage parameter.

We illustrate the usefulness of the econometric technique by estimating a ''target leverage'' model that includes a speed of capital structure adjustment. There is no consensus about the speed of adjustment in the literature.[5] Moreover, it has become clear that firm-specific factors affect capital structure (Lemmon et al., 2008), but the econometric uncertainties associated with dynamic panel data have made it difficult to achieve consensus on the importance of these factors. The results for the proposed penalized QR model show that the estimates of the adjustment speeds lie between 3% and 44% across the quantiles, showing strong evidence that there is substantial heterogeneity in the speed of adjustment among firms. The results also suggest that unlike some recent studies, which estimate a relatively quickly return to target leverage, the return of the firms to their target leverage ratios is strongly asymmetric and varies along the conditional quantile function of the debt ratio.

The rest of the paper is organized as follows. Section 2 presents the dynamic panel data QR FE models and reviews its asymptotic and computational properties. Section 3 discusses selection of the tuning parameter. Section 4 describes the Monte Carlo experiment. In Section 5 we illustrate the new approach. Conclusions appear in Section 6.

2. The model and estimators

2.1. Panel data models and a heuristic introduction to penalties

Consider the classical linear panel data model

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

which we will write in matrix form as

$$y = Z\eta + X\beta + u.$$

The matrix Z represents an incidence matrix that identifies the N distinct individuals in the sample.

For the fixed effects (FE) case, η_i is assumed to be fixed parameters to be estimated. However, there are a large number of parameters in the FE model and the loss of degrees of freedom can be avoided if the η_i can be assumed random. Suppose that u and η are independent Gaussian vectors with $u \sim \mathrm{N}(0,R)$ and $\eta \sim \mathrm{N}(0,Q)$. Observing that $v = Z\eta + u$ has covariance matrix

$$Evv' = R + ZQZ'$$

one can show that the minimum variance unbiased estimator of β is

$$\hat{\beta} = (X'R + ZQZ')^{-1}X)^{-1}X'(R + ZQZ')^{-1}y.$$

The optimization problem that generates $\hat{\beta}$ suggests a large class of possible candidates estimators under non-Gaussian conditions. It is possible to show (Proposition 1 in Koenker, 2004) that

$$\hat{\beta} \text{ solves } \min_{(\eta,\beta)} \|y - Z\eta - X\beta\|^2_{R^{-1}} + \|\eta\|^2_{Q^{-1}}.$$

The approach proposed in this paper is closely related to the random effects framework (see for instance, Robinson, 1991; Jiang, 1998). Robinson (1991) showed that the Gaussian random effects estimator can be viewed as the penalized least squares. The implicit estimation of the random effects may appear strange, but viewing the random effects estimator as a penalized least squares estimator opens the door to the consideration of alternative penalties. In the Bayesian paradigm the penalty formulation is natural, as emphasized by Lindley and Smith (1972), and many subsequent authors.[6]

2.2. Penalized quantile regression for dynamic panel data

Consider the following model for the τ th conditional quantile functions of the response of the tth observation on the ith individual y_{it},[7]

$$Q_{y_{it}}(\tau|\eta_i, y_{it-1}, x_{it}) = \eta_i + \alpha(\tau)y_{it-1} + x_{it}'\beta(\tau), \quad i = 1,\ldots,N, \; t = 1,\ldots,T, \tag{2}$$

where y_{it} is the outcome of interest, y_{it-1} is the lag of the latter, x_{it} are the exogenous variables, and $\eta = (\eta_1,\ldots,\eta_N)'$ is the $N \times 1$ vector of individual specific effects or intercepts. The effects of the covariates (y_{it-1}, x_{it}) are allowed to depend upon the quantile, τ, of interest, while the FE, η_i, do not.

[5] Fama and French (2002) estimate that firms adjust between 7% and 18% each year, while Roberts (2002) documents adjustment speeds approaching 100% for some industries. Welch (2004), Leary and Roberts (2005), and Flannery and Rangan (2006) also estimate quite different adjustment speeds.

[6] Another related literature concerning penalization in panel data models poses a different question, whether to estimate the panel data model separately for each cross-sectional unit or by pooling the data. See Maddala and Hu (1995), Maddala et al. (1997, 2001) for related discussion.

[7] 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.

It is important to note that in QR framework there is no transformation which can eliminate the fixed effects, thus we are required to deal with the full problem (i.e. estimation of the FE). 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."

The panel data QR with FE (QRFE) estimator $(\widehat{\eta},\widehat{\alpha},\widehat{\beta})$ solves

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

with $\rho_\tau(u)=u(\tau-I(u<0))$ as in Koenker and Bassett (1978), z_{it} an auxiliary variable that identifies the FE, $\{\upsilon_k\}$ a set of weights for the τ_k–quantiles, $k=1,\ldots,q$, that are jointly estimated. The existence of the parameter η_i, whose dimension N is tending to infinity raises another problem, the incidental parameters problem.[8] The FE estimator of panel models can be severely biased because of the incidental parameter problems. Koenker (2004, Theorem 1) shows that, under some regularity conditions, notably that with $T\to\infty$ as $N\to\infty$ and $N^a/T\to 0$, for some $a>0$, QRFE is consistent and asymptotically Gaussian.[9]

When T is small relative to N, the QRFE estimator, as the *within* least squares estimator case, is inevitably biased in the presence of lagged dependent variables as regressors. In standard estimation of dynamic panel models the initial values of the dynamic process raise a problem, and consistency of the estimators depends critically on the assumptions about the initial conditions. However, the bias of the dynamic panel QR can be ameliorated through the use of IV, say w, that affect the determination of lagged y but are independent of innovations. Galvao (2009) discussed estimation and inference for model described in (2), proposing an IV strategy to estimate the model and reduce the dynamic bias. Based on Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991) the IV approach employs lagged (or lagged differences of the) regressors as instruments along with the QR IV framework proposed by Chernozhukov and Hansen (2006, 2008). The instrumental variables quantile regression with fixed effects (IVQRFE) estimator is then defined as

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

where $(\widehat{\eta}(\alpha),\widehat{\beta}(\alpha),\widehat{\gamma}(\alpha))$ solves

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

with $\|x\|_A=\sqrt{x'Ax}$, A being a normalization matrix, and w_{it} contains the valid instruments, which in general are lagged (or lagged differences of the) regressors. The final parameter estimates of interest are, thus

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

and

$$\widehat{\theta}=(\widehat{\theta}(\tau_1)',\ldots,\widehat{\theta}(\tau_q)')'.$$

Unfortunately, in estimation of dynamic panel data models IV become less informative when the autoregressive parameter increases toward one, and also when the variability of the FE increases. In these cases instruments from lagged (or lagged differences) of y tend to perform poorly. In addition, the estimation of the individual effects increases the variability of the estimators of the covariate effects, a process known as inflation effect. Therefore, we consider a penalizing strategy to improve the properties of the QR dynamic panel IV estimators. The penalty term reduces the bias and increases the efficiency of the IV estimators, because the regularization, or shrinkage, of the individual effects toward a common value can help to mollify the inflation caused by the presence of the unobserved individual heterogeneity. The underlying intuition is that there is no need to use instruments for the lagged y in the dynamic panel data model that does not have any FE. In fact, with shrinkage the model without IV, QRFE, performs as well as the IVQRFE model. Therefore, shrinking the FE may help to overcome the dynamic panel bias problem.

It is worth to mention that shrinkage models have been widely used for model selection, but this is not our purpose for introducing a penalty in the QRFE model. In particular, the l_1 penalty, also known as the Lasso regression introduced by Tibshirani (1996), has been proposed as a method that identifies the true model consistently (see Wang et al., 2007a, for a discussion on LAD-Lasso regression models). In our case, the shrinkage does not involve any model selection of the FE. The reason is that the *true* model contains FE, which are potentially correlated with all the regressors, and in particular, with the lag of the dependent variable. Therefore no FE can be excluded to estimate the model consistently. As a result using a penalty term invariably introduces some bias in the model, and therefore there is no "free lunch". However, Koenker (2004) shows that this bias is quite small in QRFE models, while there is a considerable gain in efficiency.

[8] The incidental parameters problem arises in the estimation of nonlinear and dynamic panel models which include individual specific effects to control for unobserved time invariant heterogeneity; see Neyman and Scott (1948) for a general discussion, Nickell (1981) for its implications in linear dynamic panel model, and Bester and Hansen (2009) for a recent discussion.

[9] Also see below assumptions (A1)–(A5) and Theorem 1.

The penalized QRFE (PQRFE) estimator $(\widehat{\eta}(\lambda),\widehat{\alpha}(\lambda),\widehat{\beta}(\lambda))$ solves

$$\min_{\eta,\beta,\gamma}\sum_{k=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\upsilon_k\rho_\tau(y_{it}-z_{it}'\eta-\alpha(\tau_k)y_{it-1}-x_{it}'\beta(\tau_k))+\lambda\sum_{i=1}^{N}|\eta_i|,$$

with a given $\lambda\geq 0$ penalty term that affects only the FE. Note that as $\lambda\to 0$ PQRFE becomes QRFE, while as $\lambda\to\infty$, $\widehat{\eta}_i\to 0$ for all $i=1,2,\ldots,N$ and we obtain the model purged of the FE. As in Knight and Fu (2000), by allowing λ to vary with the sample size, the asymptotic distribution of the PQRFE estimators can be obtained. Koenker (2004, Theorem 2) shows that for $\lambda_T/\sqrt{T}\to\lambda_0\geq 0$ PQRFE is asymptotically Gaussian. If $\lambda_0=0$ the PQRFE estimator is asymptotically unbiased, otherwise the estimators will be asymptotically biased and the bias will depend on the size of λ_0 and the sign of η_i.

Finally, from the availability of instruments, w_{it}, we introduce the penalized IVQRFE (PIVQRFE) as

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

where $(\widehat{\eta}(\alpha,\lambda),\widehat{\beta}(\alpha,\lambda),\widehat{\gamma}(\alpha,\lambda))$ solves

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

with $\|x\|_A=\sqrt{x'Ax}$. The final parameter estimates of interest are, thus,

$$\widehat{\theta}(\tau,\lambda)\equiv(\widehat{\alpha}(\tau,\lambda),\widehat{\beta}(\tau,\lambda))\equiv(\widehat{\alpha}(\tau,\lambda),\widehat{\beta}(\widehat{\alpha}(\tau,\lambda),\tau,\lambda))$$

and

$$\widehat{\theta}(\lambda)=(\widehat{\theta}(\tau_1,\lambda)',\ldots,\widehat{\theta}(\tau_q,\lambda)')'.$$

Note that, for $\lambda=0$, we have the IVQRFE estimator proposed in Galvao (2009). Thus, to reduce the dynamic bias in the QRFE estimator, we suggest to use a penalized panel data version of the IV QR method of Chernozhukov and Hansen (2006, 2008) along with lagged (or lagged differences of the) regressors as instruments, as in the dynamic panel data literature. Moreover, if we exclude the instruments this is the PQRFE estimator. Therefore, this model combines these two models in order to solve the weak instrument problem in the presence of high values of the autoregressive parameter. The choice of the shrinkage parameter λ will be investigated in the next section.

Now we briefly discuss the asymptotic properties of the PIVQRFE estimator. The existence of the parameter η_i, whose dimension N is tending to infinity, and the penalty term raise some new issues for the asymptotic analysis of the proposed estimator. We impose the following regularity conditions:

(A1) The y_{it} are independent across individuals, stationary, with conditional distribution functions F_{it}, and differentiable conditional densities, $0<f_{it}<\infty$, with bounded derivatives f_{it}' for $i=1,\ldots,N$ and $t=1,\ldots,T$.
(A2) For all τ, $(\alpha(\tau),\beta(\tau))\in$ int $\mathcal{A}\times\mathcal{B}$, and $\mathcal{A}\times\mathcal{B}$ is compact and convex.
(A3) $\max_{it}\|y_{it}\|=O(\sqrt{NT})$; $\max_{it}\|x_{it}\|=O(\sqrt{NT})$; $\max_{it}\|w_{it}\|=O(\sqrt{NT})$.
(A4) $N^a/T\to 0$ for some $a>0$, and $\lambda_T/\sqrt{T}\to\lambda_0\geq 0$.
(A5) Denote $\Phi(\tau_k)=\mathrm{diag}(f_{it}(\xi_{it}(\tau_k)))$, 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 $\tilde{X}=[X',W']'$. Then, the following matrices are positive definite:

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

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

and

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

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 $J_{\beta\gamma}^{-1}$, and $H=\bar{J}_\gamma'A\bar{J}_\gamma$. Then, $J_{\beta\gamma}$, $J_\alpha'HJ_\alpha$ are invertible.

Condition (A1) is a standard assumption in QR literature and imposes a restriction on the density function of y_{it}. (A2) imposes compactness on the parameter space of $\alpha(\tau)$. Such assumption is needed since the objective function is not convex in α. Assumption (A3) imposes bound on the variables. Condition (A4) is exactly the same as in Koenker (2004) and restricts the rate at which N and T grow to infinity. We impose this condition in order to use a similar asymptotic approximation. Finally, Assumption (A5) states invertibility conditions for matrices in order to guarantee asymptotic normality. In this paper we assume identification of the parameters. In the IV QR model, as discussed in Chernozhukov and Hansen (2006), the identification is shown through the use of a version of Hadamard's theorem. It requires 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.

The recent literature analyzing bias in dynamic panel data models develops asymptotic theory where N and T are large. Under conditions (A1)–(A5) we show the asymptotic properties of the PIVQRFE estimator as N grows at a controlled rate relative to T. Assumption (A4), based on Koenker (2004), provides a lower bound for the relative rate of increasing 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 to the other parameters. Under assumption (A4) the convergence is fast enough to eliminate the inconsistency problem found for very small T and large N asymptotic approximations. An alternative view of this argument is that it applies to situations in which T tends to infinity, and N is fixed. The limiting distribution of PIVQRFE estimator for the dynamic panel model with FE is given in Theorem 1.

Theorem 1 (*Consistency and asymptotic normality*). *Under assumptions* (A1)–(A5),

$$\widehat{\vartheta}(\tau)\xrightarrow{p}\vartheta(\tau),$$

where $\vartheta=(\eta,\alpha,\beta)$. *Furthermore*,

$$\sqrt{NT}(\widehat{\theta}(\lambda)-\theta(\lambda))\xrightarrow{d}N(Bias(\lambda),\Omega),$$

where $\Omega=(K',L')'S(K',L')$, *S*, *H and all the elements with J are defined as in* (A5), $K=(J_\alpha' HJ_\alpha)^{-1}J_\alpha H$, $L=\bar{J}_\beta M$, $M=I-J_\alpha K$, $Bias(\lambda)=(K',L')'V\lambda_0 s$, *with* $V=(\upsilon_1 X'M_Z',\dots,\upsilon_q X'M_Z')$ *and s is an NT-vector* sgn (η_i).

Proof. The proof is given in the Appendix.

The implementation of the PIVQRFE procedure is given below. The *R* code is available from the authors upon request. For implementation we treat λ as fixed. In the next section we discuss its estimation. Define the objective function

$$Q_{NT}(\eta,\alpha,\beta,\gamma):=\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\upsilon_k\rho_\tau(y_{it}-z_{it}'\eta-\alpha(\tau_k)y_{it-1}-x_{it}'\beta(\tau_k)-w_{it}'\gamma(\tau_k))+\lambda\sum_{i=1}^{N}|\eta_i|. \tag{3}$$

For K=1 and given λ, the PIVQRFE estimator can be implemented as follows:

(1) For a given quantile of interest τ, and penalty parameter λ, define a grid of values $\{\alpha_j,j=1,\dots J;|\alpha|<1\}$, and estimate the $\tau-$QR defined in Eq. (3) to obtain coefficients $\widehat{\eta}(\alpha_j,\tau)$, $\widehat{\beta}(\alpha_j,\tau)$ and $\widehat{\gamma}(\alpha_j,\tau)$, that is, for a given value of the autoregressive structural parameter, say α, one estimates the ordinary panel QR to obtain

$$(\widehat{\eta}_i(\alpha_j,\tau),\widehat{\beta}(\alpha_j,\tau),\widehat{\gamma}(\alpha_j,\tau)):=\min_{\eta_i,\beta,\gamma}Q_{NT}(\tau,\eta_i,\alpha,\beta,\gamma). \tag{4}$$

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

$$\widehat{\alpha}(\tau)=\min_{\alpha}[\widehat{\gamma}(\alpha,\tau)']A[\widehat{\gamma}(\alpha,\tau)], \tag{5}$$

where A is a positive definite matrix.[10] The estimate $\widehat{\beta}(\tau)$ is then given by $\widehat{\beta}(\widehat{\alpha}(\tau),\tau)$, which leads to the estimates of interest

$$\widehat{\theta}(\tau)=(\widehat{\alpha}(\tau),\widehat{\beta}(\tau))=(\widehat{\alpha}(\tau),\widehat{\beta}(\widehat{\alpha}(\tau),\tau)). \tag{6}$$

For $K>1$ and given λ, 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 parameters estimates from the model without any instruments. The design matrix for the problem of estimating $K>1$ is as follows:

$$[\upsilon\otimes(I_N\otimes 1_T)\vdots\Upsilon\otimes Y_{-1}\vdots\Upsilon\otimes X\vdots\Upsilon\otimes W],$$

where I_N is an $N\times N$ identity matrix, ι_T is a $T\times 1$ vector of ones, Υ is a $K\times K$ diagonal matrix with the weights υ in the diagonal. The corresponding response vector is $\tilde{y}=((\upsilon\otimes y)'0_N')'$. However, as Koenker (2004) and Belloni and Chernozhukov (2009) observe, 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.

[10] 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 $\widehat{\gamma}(\alpha(\tau),\tau)$ otherwise.

3. Penalty selection

In this section we discuss two methods to select the penalty parameter, λ. As in any regularization problem, the selection of the tuning parameter is of fundamental interest. In practice, the tuning parameter λ is typically selected by minimizing the generalized cross validation. First, we follow a substantial body of the literature (see for instance Koenker et al., 1994; Wang et al., 2007a, b; Li and Zhu, 2008) and propose a Bayesian information criterion (BIC) criterion to choose the tuning parameter λ. We select the optimal λ by minimizing

$$BIC_1(p_\lambda)=\log\widehat{\sigma}_\lambda+\tfrac{1}{2}(NT)^{-1}p_\lambda \log NT, \tag{7}$$

where $\widehat{\sigma}_\lambda = NT^{-1}\sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it}-z_{it}'\widehat{\eta}(\lambda)-\widehat{\alpha}(\lambda)y_{it-1}-x_{it}'\widehat{\beta}(\lambda)-w_{it}'\widehat{\gamma}(\lambda))$, and p_λ is a measure of the effective dimension of the fitted model with penalty parameter λ. Machado (1993) considers similar criteria for QR and more general M-estimators of regression.

Second, we also consider the least square approximation (LSA) of Wang and Leng (2007).[11] They propose

$$BIC_2(p_\lambda)=(\widehat{\theta}_\lambda-\tilde{\theta})'\widehat{\Sigma}^{-1}(\widehat{\theta}_\lambda-\tilde{\theta})+(NT)^{-1}p_\lambda \log NT, \tag{8}$$

where $\widehat{\theta}_\lambda$ is the penalized estimate for a given λ and $\tilde{\theta}$ is the unpenalized estimate with covariance matrix $\widehat{\Sigma}$. In our application we set $\tilde{\theta}$ to be the QRFE and IVQRFE estimates for models with and without IV, respectively. We also use the respective unpenalized covariance matrix in the two cases. Moreover, we use $\widehat{\theta}_\lambda$ as the PQRFE and PIVQRFE estimates.

Note that the second term in both Eqs. (7) and (8) coincide, that is, they coincide in the l_1-penalty term. The first term of the BIC_1 criterion is based on the loss function of the quantile regression problem, $\rho_\tau(\cdot)$, while BIC_2 is based on a least-squares loss function.

The commonly used BIC criteria depends on the quantity p_λ which should be an informative measure of the complexity of a fitted model. We propose two different methods for computing this. Consider a sufficiently small number κ. First, we use p_λ as the dimension of the set $\{\alpha\cup\beta\cup\{i|\ |\eta_i|>\kappa\}\}$. This will be denoted as Method 1. Second, following Nychka et al. (1995) and Yuan (2006) we use the SURE divergence formula (Stein, 1981), $\sum_{i=1}^n \partial\widehat{y}_i/\partial y_i$ to estimate p_λ, where $\widehat{y}_i$ is the fitted model. This measures the sum of the sensitivity of each fitted value with respect to the corresponding observed value. As Meyer and Woodroofe (2000) and Li and Zhu (2008) observe, it turns out that in the case of l_1-norm QR case, the number of interpolated data points is a convenient estimate for the effective dimension of the model. The second method will be denoted as Method 2. In this case, p_λ is computed as $\sum_{i=1}^N \sum_{t=1}^T I[|\widehat{\xi}_{it}(\tau,\lambda)|<\kappa]$ where $\widehat{\xi}_{it}(\tau,\lambda)=y_{it}-z_{it}'\widehat{\eta}(\lambda)-\widehat{\alpha}(\tau,\lambda)y_{it-1}-x_{it}'\widehat{\beta}(\tau,\lambda)-w_{it}'\widehat{\gamma}(\tau,\lambda)$ is the $\tau-$quantile residual sequence for a given λ. Following Machado (1993) we use $\tau=0.5$ for the penalty selection. Monte Carlo results, presented in the next section, suggest that minimization of a BIC criterion produces a suitable estimator for λ for both methods.

4. Monte Carlo

4.1. Monte Carlo design

In this section we report the results of a Monte Carlo study which investigates the potential gains from exploiting the penalty term in the QR dynamic panel data model. We examine and compare the results with the penalized QRFE proposed by Koenker (2004), which has no IV, and with the IVQRFE proposed in Galvao (2009), which sets the penalty term equals to zero, and finally we introduce a penalty term in the latter model. A simple version of the model is considered where the response y_{it} is generated by the model

$$y_{it}=\eta_i+\alpha y_{it-1}+\beta x_{it}+\sigma_u u_{it}. \tag{9}$$

We do not exploit the fact that QR models are specially advantageous for heavy tailed innovations, and rather consider the simplest model where innovation term, u_{it}, is standard normal. Moreover, all the simulations are based on median or LAD regression. Although not reported, similar results are obtained for other quantiles.

The regressor x_{it} is generated according to

$$x_{it}=\mu_i+\zeta_{it}, \tag{10}$$

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

$$\zeta_{it}=\phi\zeta_{it-1}+\varepsilon_{it}+\theta\varepsilon_{it-1}, \tag{11}$$

with ε_{it} following a normal distribution. In all cases we set $\zeta_{i,-50}=0$ and generate ζ_{it} for $t = -49,-48,\ldots,T$, and discard the first 50 time series. 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 discarded the first 50 observations, using the observations t = 0 through T for

[11] We are grateful to an anonymous referee for pointing this out.

Table 1
Bias and RMSE for different autoregressive parameter values.

	α					
	0.8	0.85	0.9	0.95	0.98	0.99
QRFE						
$\widehat{\alpha}$						
Bias	−0.233	−0.241	−0.249	−0.245	−0.174	−0.127
RMSE	0.240	0.248	0.256	0.251	0.180	0.134
$\widehat{\beta}$						
Bias	0.007	0.002	−0.005	−0.011	−0.007	−0.003
RMSE	0.057	0.057	0.057	0.057	0.056	0.054
IVQRFE						
$\widehat{\alpha}$						
Bias	−0.009	−0.045	−0.068	−0.109	−0.121	−0.132
RMSE	0.145	0.150	0.153	0.166	0.179	0.191
$\widehat{\beta}$						
Bias	−0.002	−0.005	−0.009	−0.013	−0.015	−0.015
RMSE	0.057	0.056	0.056	0.059	0.062	0.063

Notes: Bias and RMSE using $\beta=0.2$, $\sigma_u=1$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. T=10 and N=50. Monte Carlo experiment based on 1000 replications.

estimation. The fixed effects, μ_i and α_i, are generated as

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

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

The above method of generating μ_i and η_i ensures that the random effects estimates are inconsistent because of the correlation that exists between the individual effects and the error term or the explanatory variables. In the simulations, we fix $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$, and $\theta=0.5$, and set the number of replications to 1000. We compare the estimators in terms of bias and root mean squared error (RMSE) for the median case. The IV used for the lagged dependent variable is the lag of the x covariate. Similar results are obtained with other instruments as well.

4.2. Monte Carlo results

First, we evaluate the bias induced by the values of the autoregressive coefficient. Table 1 shows the bias and RMSE of the QRFE and IVQRFE estimators. The autoregressive parameter α varies in {0.8, 0.85, 0.9, 0.95, 0.98, 0.99}. The simulations show that bias in the estimates of $\widehat{\alpha}$ and $\widehat{\beta}$ increase when α goes toward one. For $\alpha=0.8$, the IVQRFE model is able to significantly reduce the bias observed in the QRFE model. However, for high values of the autoregressive parameter, the bias and RMSE are similar (or higher) than those in the model with no IV. This result is consistent with the fact that IV become less informative for high values of α.

Second, we study the properties of QRFE and IVQRFE for different panel sizes. Table 2 reports simulations for $T \in \{5,10,20\}$ and $N \in \{50,100\}$ with a fixed autoregressive parameter $\alpha=0.98$. The experiments show that IVQRFE is able to significantly reduce the bias and has gains over QRFE for T= (5,10). Moreover, both bias and RMSE are monotonically decreasing on T for both N dimensions. These results are in line with least squares simulations (see for instance Arellano and Bond, 1991) that show that the dynamic bias exacerbates in short panels. Third, we study the properties of QRFE and IVQRFE for different values of the signal-to-noise ratio by considering different values of σ_u^2. Table 3 reports simulations for $\sigma_u^2 \in \{1,3,5\}$. The results show that both the bias and RMSE of $\widehat{\alpha}$ are increasing in σ_u^2.[12]

Next we consider the penalized version of the models varying the tuning parameter (λ). We fix the autoregression parameter to $\alpha=0.98$. Table 4 shows the bias and RMSE in the Koenker (2004) penalized FE model in QR (PQRFE) and the penalized instrumental variables dynamic quantile regression (PIVQRFE). Selected values of the penalty parameter λ are shown. The simulations show striking features of the penalty. First, small increments in the penalty term produce considerable gains in terms of bias and RMSE. Second, these gains are surprisingly high in the PQRFE model (i.e. without

[12] We are in debt to an anonymous referee for suggesting these experiments.

Table 2
Bias and RMSE for different panel sizes.

	N=50			N=100		
	T=5	T=10	T=20	T=5	T=10	T=20
QRFE						
$\widehat{\alpha}$						
Bias	−0.517	−0.167	−0.059	−0.515	−0.167	−0.059
RMSE	0.530	0.172	0.062	0.521	0.169	0.060
$\widehat{\beta}$						
Bias	−0.016	−0.005	−0.002	−0.018	−0.003	−0.002
RMSE	0.118	0.052	0.030	0.082	0.037	0.021
IVQRFE						
$\widehat{\alpha}$						
Bias	−0.216	−0.118	−0.045	−0.167	−0.079	−0.030
RMSE	0.278	0.176	0.088	0.230	0.127	0.059
$\widehat{\beta}$						
Bias	−0.012	−0.014	−0.010	−0.014	−0.008	−0.006
RMSE	0.126	0.058	0.036	0.084	0.041	0.024

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_u=1$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. Monte Carlo experiment based on 1000 replications.

Table 3
Bias and RMSE for different signal-to-noise ratios.

σ_u^2	1	3	5
QRFE			
$\widehat{\alpha}$			
Bias	−0.174	−0.306	−0.327
RMSE	0.180	0.312	0.332
$\widehat{\beta}$			
Bias	−0.011	−0.011	−0.014
RMSE	0.057	0.156	0.266
IVQRFE			
$\widehat{\alpha}$			
Bias	−0.121	−0.210	−0.224
RMSE	0.179	0.265	0.280
$\widehat{\beta}$			
Bias	−0.015	−0.028	−0.020
RMSE	0.065	0.167	0.282

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. T=10 and N=50. Monte Carlo experiment based on 1000 replications.

IV). For this case, a penalty term of $\lambda\approx 0.15$ produces the smallest bias and RMSE, and there is small changes for higher values of the penalty term. For the IV case, the optimal penalty is larger ($\lambda\approx 2$), and the gains in reducing the bias and RMSE, for the autoregressive coefficient, relative to the model with $\lambda=0$, are very substantial. For the exogenous variable coefficient, the bias and RMSE when $\lambda=2$ are slightly higher than those for the model without penalization. In order to illustrate the effect of the penalty on different panel sizes, Fig. 1 plots the RMSE for the PIVQRFE models in Table 2 for the same λ values considered here. The figure shows that reductions in RMSE are mostly in short panels. These are particularly large for T=(5,10), while, in contrast, for T=20, the calculated RMSE is flat. We also plot the effect of λ on the RMSE for different values of σ_u^2 (see Fig. 2). The graph shows that RMSE reduces monotonically with λ, for all three values of σ_u^2.

We also develop an experiment to assess the finite sample performance of the *BIC* criteria in estimating λ. We use the location model described in Eq. (9), with $\alpha=0.98$ and $\beta=0.2$, and estimate the model for a grid of λ. Then, we select $\widehat{\lambda}$ by minimizing $BIC(p_\lambda)$ for each model (BIC_1 and BIC_2 with Methods 1 and 2 of computing p_λ), and report the coefficients for this selection, i.e. $\widehat{\alpha}(0.5,\widehat{\lambda})$ and $\widehat{\beta}(0.5,\widehat{\lambda})$. We consider three tolerance parameter cases, $\kappa\in\{0.001,0.01,0.1\}$ in order to assess the sensibility of the models to the p_λ value. Table 5 presents the bias and RMSE results for $\widehat{\alpha}$ and $\widehat{\beta}$, together with the mean and standard error of the selected $\widehat{\lambda}$. In this case we consider 500 repetitions because this method is quite computationally

Table 4
Bias and RMSE for different penalty parameter values.

	λ										
	0	0.1	0.15	0.2	0.25	0.3	1	1.5	2	3	4
PQRFE											
$\widehat{\alpha}$											
Bias	−0.163	−0.012	0.011	0.018	0.022	0.024	0.029	0.029	0.030	0.030	0.030
RMSE	0.169	0.029	0.018	0.020	0.022	0.024	0.029	0.030	0.030	0.030	0.030
$\widehat{\beta}$											
Bias	−0.004	0.007	0.004	0.003	−0.001	−0.003	−0.015	−0.020	−0.026	−0.033	−0.033
RMSE	0.053	0.054	0.052	0.051	0.050	0.050	0.048	0.046	0.048	0.052	0.055
PIVQRFE											
$\widehat{\alpha}$											
Bias	−0.123	−0.104	−0.095	−0.091	−0.097	−0.089	−0.047	−0.020	−0.010	−0.013	−0.014
RMSE	0.180	0.168	0.149	0.146	0.153	0.148	0.105	0.064	0.052	0.053	0.052
$\widehat{\beta}$											
Bias	−0.014	−0.007	−0.005	−0.002	0.000	0.003	0.027	0.039	0.047	0.056	0.082
RMSE	0.061	0.058	0.058	0.059	0.056	0.057	0.073	0.078	0.087	0.105	0.134

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_u=1$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. T=10 and N=50. Monte Carlo experiment based on 1000 replications.

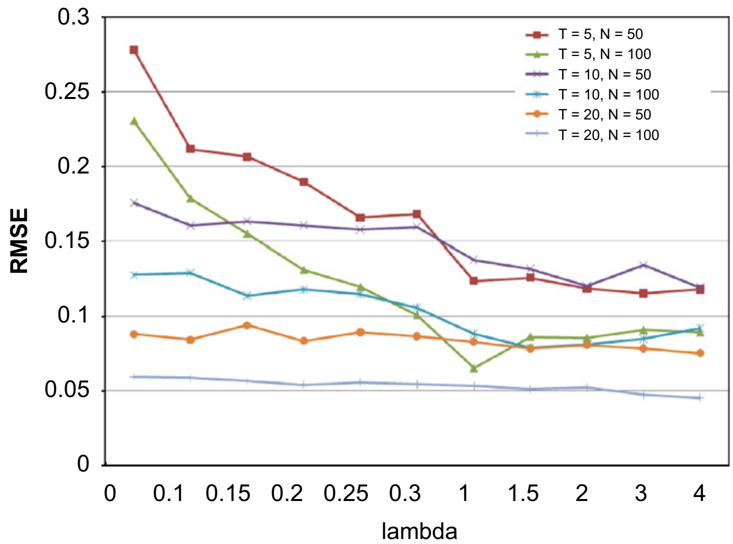

Fig. 1. RMSE and tuning parameter λ for different panel sizes—PIVQRFE.

demanding. We note that the choice of p_λ (Methods 1 or 2), which depends on the tolerance parameter κ, has some effect on the estimators. If κ is too small, only a few observations exactly provides discrepancies from zero with such a small value, and therefore the model produces a very small shrinkage. If κ is large, the model produces a very large shrinkage.

Consider first the BIC_1 criterion. Regarding Method 1, the results show that PQRFE has a better performance than PIVQRFE. Overall, the former selects a λ value smaller than the latter, with small variations across different values of κ. Moreover, bias and RMSE are slightly above the best cases found in Table 4. Consider now Method 2. The simulations show that the PQRFE model has an excellent performance, with comparable bias than the best results in Table 4 for both α and β. Note, however, that the RMSE is higher for α than in any λ−value simulation in Table 4 (except for the case with $\lambda=0$). The selected λ varies according to the κ value. In general there is a marked trade-off between bias and RMSE. The smallest bias for α (-0.0079) is obtained with $\kappa=0.01$, with an associated average λ of 0.371. However, this also corresponds to the highest RMSE (0.051). Moreover, the smallest RMSE correspond to $\kappa=0.001$, with an associated $\lambda=0.0302$, although this has the highest bias. The high standard deviation in $\widehat{\lambda}$ is due to the number of cases where the selection of λ is zero. As expected the selected λ depends on the tolerance parameter κ. For $\kappa=0.001$, $\widehat{\lambda}$ is very small due to the considerable

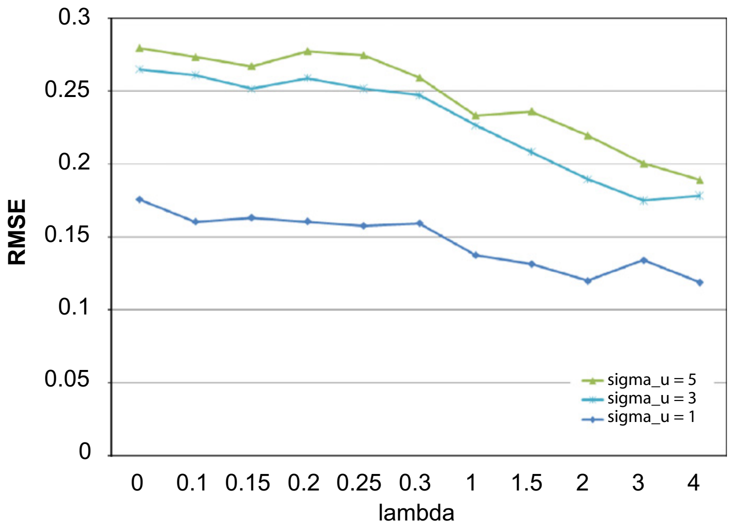

Fig. 2. RMSE and tuning parameter λ for different signal-to-noise ratios—PIVQRFE.

Table 5
Penalty selection—BIC criterion.

Model	PQRFE	PIVQRFE	PQRFE	PIVQRFE	PQRFE	PIVQRFE	PQRFE	PIVQRFE
	BIC_1				BIC_2			
	Method 1		Method 2		Method 1		Method 2	
$\kappa=0.1$								
$\widehat{\lambda}$								
Mean	0.077	0.315	0.926	0.684	3.246	3.245	2.343	2.355
Std. err.	(0.043)	(0.407)	(1.000)	(0.616)	(0.515)	(0.835)	(0.727)	(0.920)
$\widehat{\alpha}$								
Bias	0.032	−0.038	0.015	−0.038	0.029	−0.025	0.020	−0.068
RMSE	0.032	0.061	0.036	0.067	0.030	0.083	0.045	0.137
$\widehat{\beta}$								
Bias	−0.0035	0.017	−0.0056	0.055	−0.062	0.015	−0.054	0.018
RMSE	0.047	0.073	0.052	0.100	0.072	0.082	0.068	0.086
$\kappa=0.01$								
$\widehat{\lambda}$								
Mean	0.089	0.329	0.371	0.797	3.141	3.335	1.634	1.489
Std. err.	(0.415)	(0.453)	(0.719)	(0.503)	(0.447)	(0.531)	(1.025)	(1.116)
$\widehat{\alpha}$								
Bias	−0.018	−0.033	−0.0079	−0.039	0.029	−0.026	0.0045	−0.071
RMSE	0.035	0.054	0.051	0.068	0.029	0.077	0.075	0.138
$\widehat{\beta}$								
Bias	−0.0027	0.018	−0.0026	0.026	−0.060	0.016	−0.043	0.0066
RMSE	0.052	0.075	0.032	0.081	0.072	0.088	0.063	0.075
$\kappa=0.001$								
$\widehat{\lambda}$								
Mean	0.057	0.284	0.0302	0.444	3.010	3.280	1.583	0.901
Std. err.	(0.311)	(0.357)	(0.197)	(0.431)	(0.407)	(0.621)	(0.788)	(1.045)
$\widehat{\alpha}$								
Bias	−0.025	−0.031	−0.026	−0.047	0.029	−0.026	0.0024	−0.089
RMSE	0.032	0.049	0.032	0.078	0.029	0.076	0.040	0.155
$\widehat{\beta}$								
Bias	−0.0038	0.014	0.0003	0.020	−0.058	0.018	−0.049	0.005
RMSE	0.052	0.070	0.050	0.080	0.070	0.090	0.066	0.072

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_u=1$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. T=10 and N=50. Monte Carlo experiment based on 500 replications.

Table 6
BIC_1 criterion, different panel sizes.

	Method 1			Method 2		
	N=50, T=5	N=50, T=10	N=50, T=20	N=50, T=5	N=50, T=10	N=50, T=20
PQRFE						
$\widehat{\lambda}$						
Mean	0.028	0.041	0.002	0.040	0.284	0.110
Std. err.	(0.192)	(0.198)	(0.001)	0.160	(0.613)	(0.469)
$\widehat{\alpha}$						
Bias	−0.111	−0.016	−0.006	−0.100	−0.013	−0.004
RMSE	0.135	0.033	0.008	0.133	0.033	0.011
$\widehat{\beta}$						
Bias	0.0008	0.0007	0.0006	0.0004	−0.002	−0.0003
RMSE	0.124	0.051	0.030	0.125	0.050	0.030
PIVQRFE						
$\widehat{\lambda}$						
Mean	0.114	0.325	0.565	0.230	0.394	0.857
Std. err.	(0.149)	(0.423)	(0.694)	(0.194)	(0.518)	(0.848)
$\widehat{\alpha}$						
Bias	−0.133	−0.037	−0.010	−0.123	−0.043	−0.012
RMSE	0.175	0.058	0.029	0.186	0.070	0.032
$\widehat{\beta}$						
Bias	0.027	0.021	0.012	0.078	0.028	0.020
RMSE	0.154	0.071	0.044	0.205	0.082	0.050

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_u=1$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. $\kappa=0.01$. Monte Carlo experiment based on 1000 replications.

number of cases when 0 is chosen. PIVQRFE models tend to select a higher value of λ. In this case the bias and RMSE for α are similar across specifications, although they are bigger than the smallest values in Table 4.

The BIC_2 criterion chooses a larger $\widehat{\lambda}$, although with a larger standard deviation. Method 1 selects a value above 3 for all κ while Method 2 varies depending on the tolerance parameter. In general Method 1 gives better results than Method 2 for estimating α. The results for PQRFE, Method 1, show a similar bias for all κ and a very small RMSE. Overall these values are smaller than the corresponding estimates for BIC_1. The results for PIVQRFE, Method 1, show smaller bias but larger RMSE for the autoregressive parameter than the corresponding estimator using BIC_1.

Finally, we evaluate the performance of the BIC_1 criteria for different panel sizes (Table 6) and signal-to-noise ratios (Table 7), using a fixed $\kappa=0.01$.[13] Table 6 shows that the PQRFE estimator choice of λ varies depending on the model. The average of the selected λ increases from T=5 to 10, but it reduces drastically for T=20. This is consistent with the fact that the bias reduces with the time dimension, and with it the usefulness of the shrinkage method. For the PQRIVFE estimator, on the contrary, the selected λ increases with the time dimension. In all cases, the calculated bias and RMSE of $\widehat{\alpha}$ are smaller than the corresponding values in Table 2. The results in Table 7 show that Method 1 works better than Method 2 for the PIVQRFE estimator. The main reason is that, for fixed κ, Method 2 selects the penalty term in terms of the absolute values of the residuals, which in turn, depends on the variance of the error term in the DGP. As a result in Method 1, a higher σ_u^2 corresponds to a smaller $\widehat{\lambda}$, while the contrary occurs in Method 1.

5. Application: target capital structure adjustment

In this section we apply our proposed estimator to a model describing firms partial adjustment towards target capital structure. Since Modigliani and Miller (1958)'s irrelevance proposition, researchers have investigated firms' decisions about how to finance their operations. In particular, they studied whether the irrelevance proposition is consistent with the available data, or, instead capital market imperfections make firm value dependent on capital structure. In this case, firms would select target debt–equity ratios, trading off the costs and benefits of leverage. Recently, there is a growing literature in estimating "target leverage" models that include the speed of capital structure adjustment. Fama and French (2002) estimate that firms adjust between 7% and 18% each year, while Roberts (2002) documents adjustment speeds approaching 100% for some industries. Welch (2004), Leary and Roberts (2005), and Flannery and Rangan (2006) also estimate quite different adjustment speeds. It has become clear that firm-specific factors affect capital structure (Lemmon

[13] Simulations for BIC_2, other κ values, and for other panel sizes (N=100) are available from the authors upon request.

Table 7
BIC_1 criterion, different signal-to-noise ratios.

σ_u^2	Method 1			Method 2		
	1	3	5	1	3	5
PQRFE						
$\widehat{\lambda}$						
Mean	0.051	0.0048	0.003	0.271	0.035	0.004
Std. err.	(0.269)	(0.097)	(0.095)	(0.575)	(0.225)	(0.100)
$\widehat{\alpha}$						
Bias	−0.015	−0.144	−0.230	−0.013	−0.141	−0.230
RMSE	0.033	0.150	0.236	0.034	0.150	0.236
$\widehat{\beta}$						
Bias	0.002	0.005	0.0005	−0.0001	0.005	−0.001
RMSE	0.051	0.162	0.257	0.052	0.162	0.257
PIVQRFE						
$\widehat{\lambda}$						
Mean	0.324	0.380	1.01	0.429	0.204	0.133
Std. err.	(0.403)	(0.593)	(1.06)	(0.520)	(0.348)	(0.246)
$\widehat{\alpha}$						
Bias	−0.037	−0.125	−0.137	−0.041	−0.139	−0.211
RMSE	0.059	0.160	0.196	0.070	0.185	0.256
$\widehat{\beta}$						
Bias	0.019	0.050	0.100	0.028	0.034	0.016
RMSE	0.070	0.191	0.302	0.085	0.192	0.297

Notes: Bias and RMSE using $\alpha=0.98$, $\beta=0.2$, $\sigma_\mu=1$, $\sigma_\eta=1$, $\phi=0.6$ and $\theta=0.5$. T=10 and N=50. $\kappa=0.01$. Monte Carlo experiment based on 1000 replications.

et al., 2008), but the econometric difficulties associated with dynamic panel data have made it difficult to achieve consensus on the importance of these factors.

A regression specification used to estimate the trade-off leverage behavior must allow each firm's target debt ratio to vary over time, and must recognize that deviations from target leverage are not necessarily offset quickly. Both of these requirements are satisfied in the model with partial (incomplete) adjustments toward a target leverage ratio that depends on firm characteristics. As in Flannery and Rangan (2006), we model the possibility that target leverage might differ across firms or over time by specifying a target capital ratio as

$$MDR^*_{it+1} = \eta_i + X_{it}\beta, \tag{12}$$

where MDR^*_{it+1} is firm i's desired debt ratio[14] at $t+1$, X_{it} is a vector of observable firm characteristics related to the costs and benefits of operating with various leverage ratios, β is a coefficient vector, and η_i represents unobservable firm characteristics. Under the trade-off hypothesis, $\beta \neq 0$, and the variation in MDR^*_{it+1} should be nontrivial.

A standard partial adjustment model is given by

$$MDR_{it+1} - MDR_{it} = \delta(MDR^*_{it+1} - MDR_{it}) + u_{it+1}. \tag{13}$$

Each year, the typical firm closes a proportion δ of the gap between its actual and its desired leverage levels. Substituting (12) into (13) and rearranging gives the estimable model

$$MDR_{it+1} = \delta\eta_i + X_{it}(\delta\beta) + MDR_{it}(1-\delta) + u_{it}, \tag{14}$$

where the coefficient δ is the adjustment speed toward the target. Eq. (14) implies that managers take "action" or "steps" to close the gap between where they are (MDR_{it}) and where they wish to be ($X_{it}\beta + \eta_i$). The specification further implies that the firm's actual debt ratio eventually converges to its target debt ratio, and that the long-run impact of X_{it} on the capital ratio is given by its estimated coefficient, divided by δ.

To investigate the heterogeneity associated with the speed of adjustment we use the QR models discussed in Section 2. Thus, we estimate the following model:

$$Q_{MDR_{it}}(\tau|\eta_i, MDR_{it-1}, X_{it-1}) = \eta_i + \alpha(\tau)MDR_{it-1} + X'_{it-1}\beta(\tau), \tag{15}$$

[14] Here $MDR_{it} = D_{it}/(D_{it} + S_{it}P_{it})$, where D_{it} denotes the book value of firm i's interest-bearing debt at time t, S_{it} denotes the number of common shares outstanding at time t, and P_{it} is the price per share at time t.

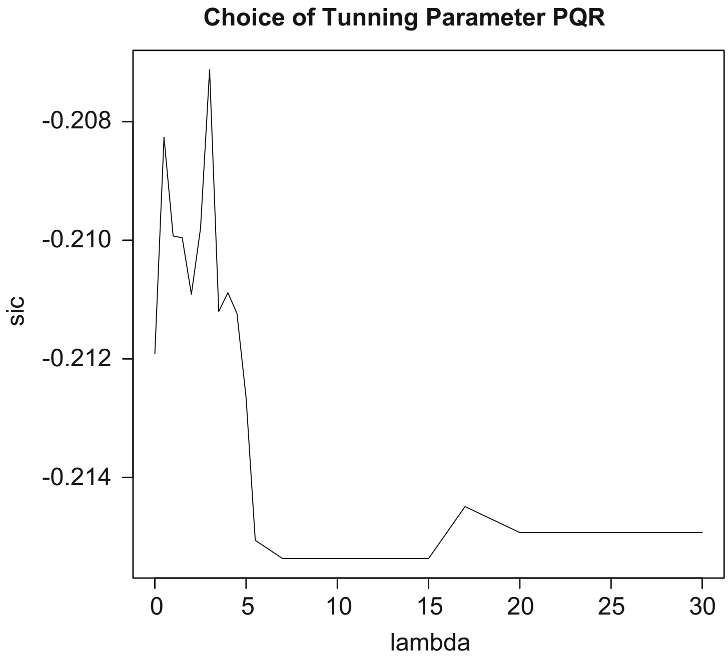

Fig. 3. Choice of the tuning parameter λ for PQRFE.

where $\alpha(\tau)=(1-\delta(\tau))$, and X_{it} is a vector containing several covariates. We estimate the model for all the deciles and we impose the fixed effects to be independent of the quantile of interest, τ, by estimating a weighted QR model (using the same weights for all quantiles). We use data from the Compustat Industrial Annual dataset. The sample consists of annual CRSP/Compustat data from the years 1971 through 2005.[15]

The first step to estimate the model is to use the proposed BIC_1 criterion to select the parameter λ. We choose the appropriate tuning parameter for the median case in the PQRFE and PIVQRFE models using Method 1 and $\kappa=0.01$.[16] Fig. 3 plots the calculated BICs for different PQRFE models. The BIC values drops dramatically after 5 and the minimum is observed for $\lambda=13$. Fig. 4 plots the BICs against several λ's for the PIVQRFE model. The picture shows a great decrease in BIC from $\lambda=0$ to 3, then BIC increases, and finally after $\lambda=20$ it becomes approximately constant. The estimate of λ that minimizes the BIC is $\widehat{\lambda}=3$. We will use both values of λ for both models.

We compute the speed of adjustment, $\widehat{\delta}$, from the estimated coefficient $\widehat{\alpha}$ for the deciles in Eq. (15) with $\widehat{\delta}=1-\widehat{\alpha}$. In the IV models we use MDR_{it-2} and MDR_{it-3} as instruments for MDR_{it-1}. For given quantile of interest τ the coefficient on lagged MDR implies that firms in that corresponding conditional quantile function close such percentage the gap between current and desired leverage within one year. For instance, in the median case firms close $(\widehat{\delta}(0.5)*100)$ percent of the gap between current and desired leverage in one year, where $\widehat{\delta}(0.5)$ is the estimated coefficient for the median.

The estimates for $\widehat{\delta}$ for both models PQRFE and PIVQRFE, using $\lambda=0$, $\widehat{\lambda}=3$ and $\widehat{\lambda}=13$, are presented in Table 8. The estimates for the speed of convergence at the median are very similar for the models without penalization ($\lambda=0$) showing a speed of about 25%. When a penalty of $\widehat{\lambda}=3$ is used the estimates are 23% and 19% for PQRFE and PIVQRFE, respectively. For a penalty of $\widehat{\lambda}=13$ the estimates are similar: 21% and 19%. Thus, the PQRFE model slightly overestimates the speed of convergence, δ, relatively to the PIVQRFE estimates. This pattern is also observed for other quantiles as well. The differences are exacerbated for small point estimates of δ (which correspond to high values of α, the autoregressive parameter), in particular, for $\tau=0.9$. In this case, while for the case with no penalization QRFE and IVQRFE are very similar, there is a considerable difference for $\widehat{\lambda}=3$, although small differences for the highest penalty value. These results are

[15] Following standard practice, we exclude financial firms (SIC codes 6000–6999) and regulated utilities (SIC codes 4900–4999). We omit firm-years with a negative book value of equity or missing data for long-term debt, debt in current liabilities, or any of the leverage factors. The log of total assets is the one variable that is not a ratio and it is deflated to 1983 dollars with the consumer price index from the Bureau of Labor Statistics. The final sample includes 19,140 firm years from 638 firms with at most 30 years of data. The panel is unbalanced with the smallest time series span being 19 years and the largest 30 years. The vector of covariates is X=(*EBITTA, MB, DEPTA, LnTA, FATA, RDDum, RDTA*). The definition of the variables is as follows. *MDR*: market debt ratio = book value of (short-term plus long-term) debt (items [9]+[34])/market value of assets (items [9]+[34]+[199]∗[25]). *EBITTA*: profitability: earnings before interest and taxes (items [18]+[15]+[16]), as a proportion of total assets (item [6]). *MB*: market to book ratio of assets: book liabilities plus market value of equity (items [9]+[34]+[10]+[199]∗[25]) divided by book value of total assets (item [6]). *DEPTA*: depreciation (item [14]) as a proportion of total assets (item [6]). *LnTA*: log of asset size, measured in 1983 dollars (item [6]∗1,000,000, deflated by the consumer price index). *FATA*: fixed asset proportion: property, plant, and equipment (item [14]/total assets, item [6]). *RDDUm*: dummy variable equal to one if firm did not report R&D expenses. *RDTA*: R&D expenses (item [46]) as a proportion of total assets (item [6]).

[16] Similar results are obtained with $\kappa\in\{0.1,0.001\}$.

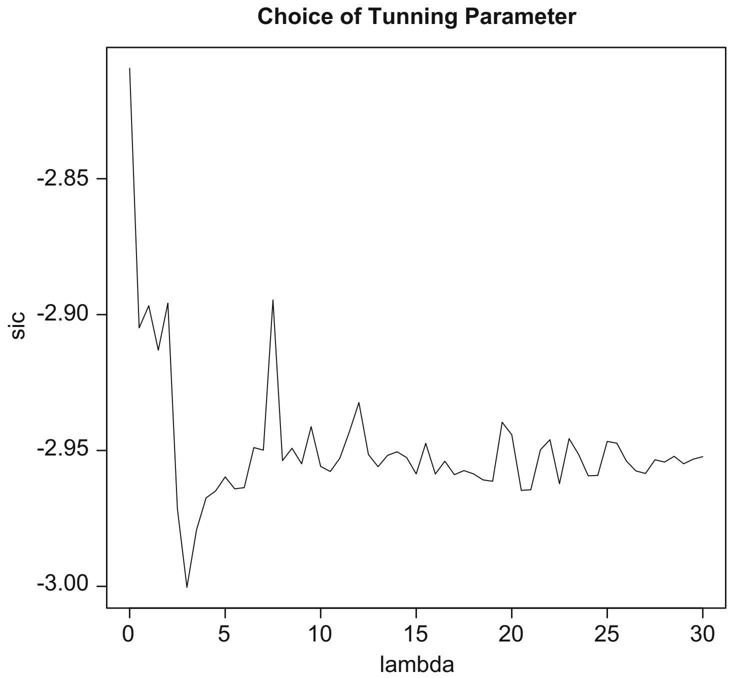

Fig. 4. Choice of the tuning parameter λ for PIVQRFE.

consistent with the previous Monte Carlo section where the IVQRFE estimator has a negative bias in the presence of weak instruments. Moreover, it is important to note that there is a strong heterogeneity across the quantiles: for $\lambda=0$, the adjustment speeds lie between 50% the first decile (in both QRFE and PIVQRFE cases) and 11% and 9% the last decile for QRFE and IVQRFE, respectively. For the case of $\widehat{\lambda}=3$, the estimated speed of convergence ranges from about 49% and 44% to 10% and 3% for PQRFE and IVQRFE, respectively.

While the simple QR panel data estimates have some attractive features, they fail to control for other potential bias in the model. The IV panel regression with unobserved FE is more appropriate if firms have relatively stable, unobserved variables affecting their leverage targets. Tables 9 and 10 report estimates for all coefficients in the FE IV panel data model as in Eq. (15) for $\lambda=0$ and $\widehat{\lambda}=3$, respectively. We construct 95% confidence intervals for the coefficients using pair-wise bootstrap methods and the results are presented in Figs. 5 and 6.

As discussed previously, there is evidence of strong heterogeneity across the deciles. The IV estimates show adjustment speeds that lie between 44% for the first decile and 10% to the last one for $\lambda=0$, and even smaller rates of adjustment speed that range from 44% to 3% for $\widehat{\lambda}=3$. In the median case, the results for the speed of convergence presented here (24% for $\lambda=0$ and 19% for $\widehat{\lambda}=3$) are smaller than the 34.4% found by Flannery and Rangan (2006) for the mean case. However, our estimates show the important fact that the speed vary substantially across quantiles, showing important heterogeneity of the firms. The confidence intervals for $\lambda=0$ show that zero lies inside the confidence bands for the *MDR* coefficient of the median and lower quantiles. On the other hand, for the case $\widehat{\lambda}=3$, the *MDR* coefficient is statistically different from zero for all quantiles. These results suggest that the model with $\widehat{\lambda}=3$ is better adjusted. The exogenous variables, X_{it}, represent the target debt ratio. They have coefficients with appropriate signs, with exception of *LnTA*. For the no penalty case, all the determinants of target leverage variables zero lie inside confidence interval showing evidence. However, for the case when $\widehat{\lambda}=3$, we can see that only for some of the variables zero lie inside the confidence interval, *MB*, *LnTA*, *FA* and *RDDUM*.

Finally, we briefly evaluate the models in terms of their prediction errors. In order to compare the models we remove the last time series observation for each firm from the dataset. These observations are used to evaluate the models using Eq. (15) and data from $t=1,\ldots,T-1$. We use the same tuning parameter as before, that is for the PQRFE $\widehat{\lambda}=13$, and for PIVQRFE $\widehat{\lambda}=3$. We also evaluate both models imposing $\lambda=0$, i.e. using the unpenalized estimators. We compute prediction for MDR_{iT} for each firm i and period T at the median ($\tau=1/2$) as

$$\widehat{MDR}_{iT}=\widehat{\eta}_i+\widehat{\alpha}MDR_{iT-1}+X'_{iT-1}\widehat{\beta}.$$

Using the vector of forecasts we compute the root mean square error, that is, $(1/N)\sum_{i=1}^{N}(\widehat{MDR}_{iT}-MDR_{iT})^2$. The mean error prediction for PIVQRFE is 0.3050, and when we set $\lambda=0$ for the IV model we obtain error prediction for IVQRFE as 0.3079. Thus, the penalized model performs better than the unpenalized model in terms of prediction. For the PQRFE the error is 0.3063, which is larger than the error for the PIVQRFE model. For the QRFE we obtain 0.3093, which corresponds to the largest error prediction. These patterns determine that gains can be obtained in prediction using penalized models with and without IV. Furthermore, the combination of shrinkage and IV produces the best model.

Table 8
Estimates of $\hat{\delta}$ for PQRFE and PIVQRFE.

τ	$\lambda = 0$		$\hat{\lambda} = 3$		$\hat{\lambda} = 13$	
	QRFE	IVQRFE	PQRFE	PIVQRFE	PQRFE	PIVQRFE
0.1	0.505	0.443	0.494	0.437	0.473	0.494
0.2	0.413	0.369	0.397	0.401	0.377	0.374
0.3	0.347	0.290	0.329	0.316	0.311	0.341
0.4	0.286	0.236	0.274	0.213	0.254	0.222
0.5	0.257	0.241	0.231	0.194	0.208	0.191
0.6	0.206	0.185	0.190	0.164	0.169	0.074
0.7	0.169	0.158	0.165	0.154	0.130	0.091
0.8	0.146	0.107	0.121	0.088	0.098	0.117
0.9	0.111	0.102	0.089	0.027	0.063	0.058

Table 9
Partial adjustment model results for IVQRFE–$\lambda = 0$.

τ	MDR	EBIT	MB	DEP	lnTA	FA	RDDUM	RDTA
0.1	0.5567	−0.0689	−0.0163	−0.4357	0.0152	0.0708	−0.0043	−0.0445
0.2	0.6311	−0.0761	−0.0095	−0.3249	0.0120	0.0567	0.0032	−0.0457
0.3	0.7100	−0.0733	−0.0074	−0.2796	0.0109	0.0491	0.0057	−0.0291
0.4	0.7640	−0.0741	−0.0060	−0.2413	0.0101	0.0436	0.0079	−0.0286
0.5	0.7591	−0.0789	−0.0052	−0.2168	0.0097	0.0445	0.0096	−0.0240
0.6	0.8150	−0.0860	−0.0048	−0.2351	0.0096	0.0471	0.0128	−0.0117
0.7	0.8422	−0.1005	−0.0046	−0.2309	0.0095	0.0493	0.0195	−0.0204
0.8	0.8927	−0.1329	−0.0056	−0.2784	0.0099	0.0667	0.0347	−0.0199
0.9	0.8977	−0.2192	−0.0080	−0.2536	0.0111	0.1159	0.0834	−0.0680

Table 10
Partial adjustment model results for PIVQRFE–$\hat{\lambda} = 3$.

τ	MDR	EBIT	MB	DEP	lnTA	FA	RDDUM	RDTA
0.1	0.5634	−0.0673	−0.0169	−0.3914	0.0172	0.0912	0.0515	−0.0524
0.2	0.5988	−0.0678	−0.0092	−0.2837	0.0140	0.0687	0.0563	−0.0484
0.3	0.6741	−0.0755	−0.0073	−0.2464	0.0130	0.0656	0.0601	−0.0394
0.4	0.7866	−0.0679	−0.0052	−0.2145	0.0121	0.0557	0.0609	−0.0318
0.5	0.8056	−0.0719	−0.0044	−0.1933	0.0117	0.0581	0.0623	−0.0268
0.6	0.8364	−0.0790	−0.0043	−0.2014	0.0113	0.0616	0.0658	−0.0119
0.7	0.8457	−0.0930	−0.0039	−0.1985	0.0113	0.0630	0.0715	−0.0158
0.8	0.9119	−0.1254	−0.0052	−0.2190	0.0117	0.0812	0.0865	−0.0226
0.9	0.9731	−0.2098	−0.0081	−0.1941	0.0132	0.1360	0.1334	−0.0725

6. Conclusions

We explored penalized dynamic quantile regression estimators for longitudinal data with fixed effects, where the penalty involves l_1 shrinkage of the fixed effects. The penalty term reduces the bias and increases the efficiency of the estimators, because the regularization, or shrinkage, of the individual effects toward a common value can help to mollify the inflation caused by the presence of the unobserved individual heterogeneity and the dynamic panel bias. The Monte Carlo experiments show that the gains from using penalized version of the model are substantial.

In addition, we discuss two possible BIC estimators for the parameter that controls the degree of shrinkage. In practice, given the estimates of the model, select the value of penalty that minimizes the model selection criterion as estimate of the tuning parameter. This avoids a more computationally intensive cross-validation approach. The Monte Carlo simulations show that these methods for selecting the penalty work reasonably well.

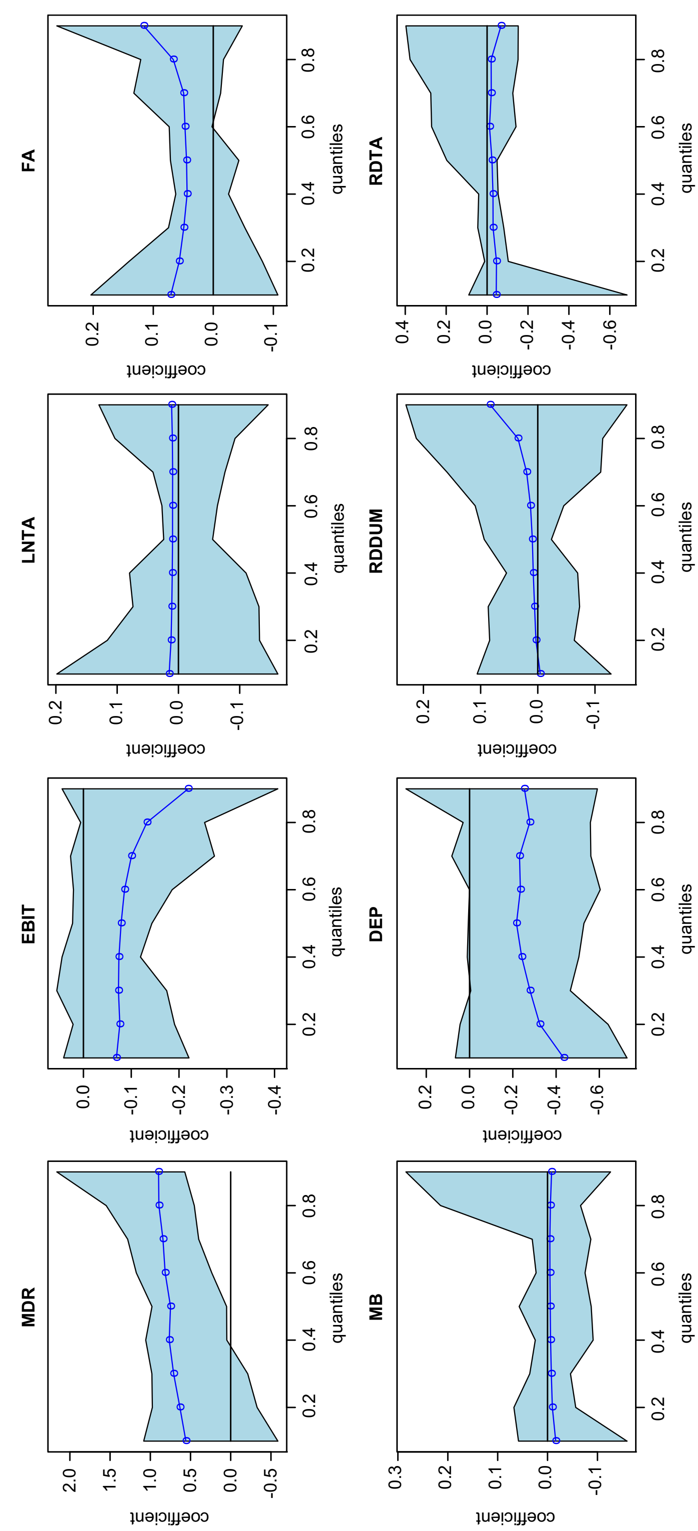

Fig. 5. Partial adjustment model confidence bands for IVQRFE ($\lambda = 0$).

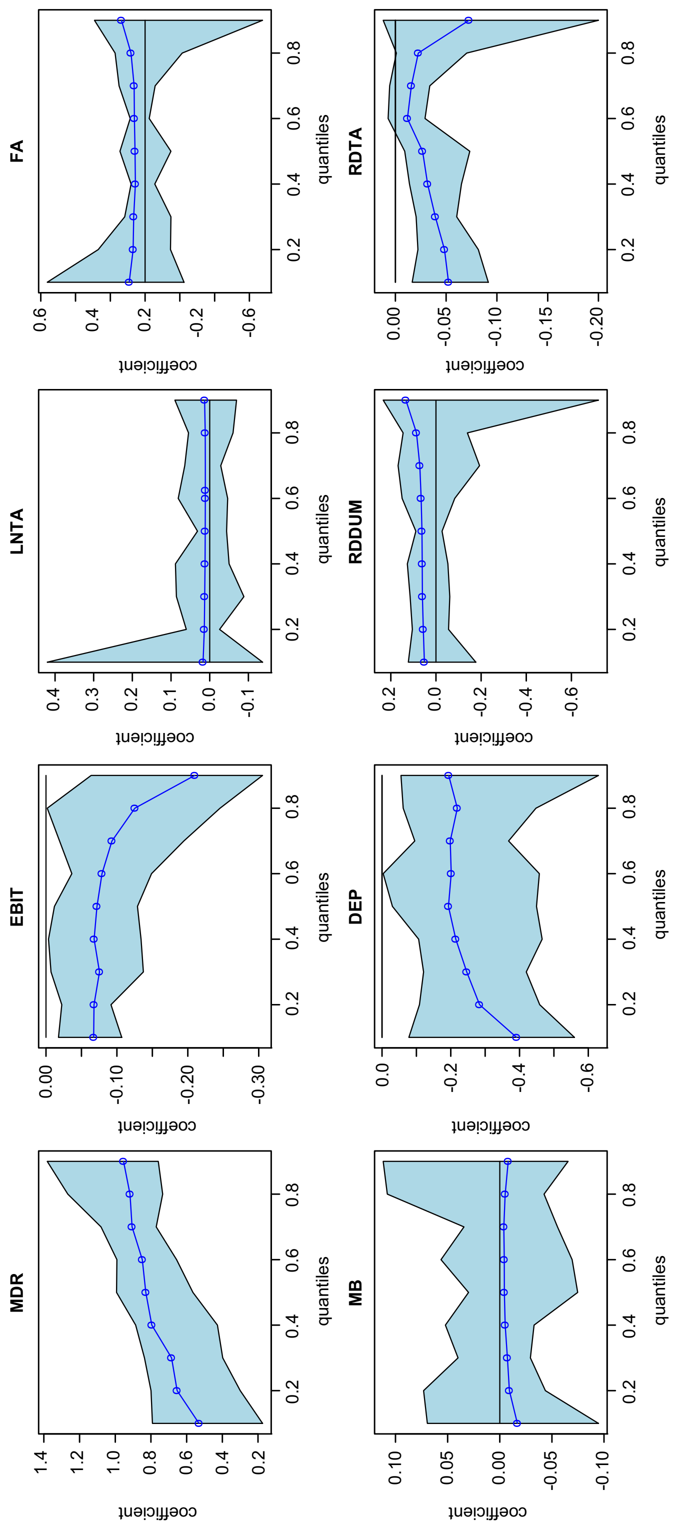

Fig. 6. Partial adjustment model confidence bands for PIVQRFE when $\hat{\lambda}=3$.

Acknowledgments

The authors would like to express their appreciation to an anonymous referee for detailed comments and very constructive suggestions and to Anil Bera, Xuming He, Ted Juhl, Roger Koenker, Steve Portnoy, and participants in the seminars at University of Illinois at Urbana-Champaign and 2009 Meetings of the Midwest Econometrics Group for helpful comments and discussions. All the remaining errors are ours.

Appendix A

In this appendix we present the proof for Theorem 1.

Proof. Consistency is a simple application of Theorem 2 of Koenker (2004), which shows consistency of the PQRFE estimator under assumptions (A1)–(A5), together with the IV proof in Chernozhukov and Hansen (2006). Since we are assuming identification of the true parameters with the IV, consistency follows. More interesting results correspond to the asymptotic distribution, together with the asymptotic bias and variance, which depend on the penalty term. The proof is similar to Theorem 2 of Koenker (2004). Consider the following model:

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

From availability of instruments, w_{it}, the objective function is

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

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

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

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

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

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,

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

For fixed $\delta_\beta,\delta_\gamma$, $\sup_{\tau\in t}\|\alpha_n(\tau)-\alpha(\tau)\| \xrightarrow{p} 0$, and $C_1>0$,

$$\sup_{\|\delta\|<C_1}\|g_i(\delta_{\eta_i},\delta_\alpha,\delta_\beta,\delta_\gamma)-g_i(0,0,0,0)-E[g_i(\delta_\eta,\delta_\alpha,\delta_\beta,\delta_\gamma)-g_i(0,0,0,0)]\| = o_p(1).$$

Expanding we have

$$E[g_i(\delta_\eta,\delta_\alpha,\delta_\beta,\delta_\gamma)-g_i(0,0,0,0)]$$
$$= E\left(-\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k\psi_\tau\left(y_{it}-\xi_{it}-\frac{\delta_\eta}{\sqrt{T}}-\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}-\frac{\delta_\beta}{\sqrt{NT}}x_{it}-\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right)-\frac{\lambda_T}{\sqrt{T}}\operatorname{sgn}\left(\eta_i-\frac{\delta_{\eta_i}}{\sqrt{T}}\right)\right.$$
$$\left.+\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k\psi_\tau(y_{it}-\xi_{it})+\frac{\lambda_T}{\sqrt{T}}\operatorname{sgn}(\eta_i)\right)$$
$$= -\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k\left[\tau-F\left(\xi_{it}+\frac{\delta_\eta}{\sqrt{T}}+\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right)\right]-\frac{\lambda_T}{\sqrt{T}}\left[\operatorname{sgn}\left(\eta_i-\frac{\delta_{\eta_i}}{\sqrt{T}}\right)-\operatorname{sgn}(\eta_i)\right]$$

$$= \frac{1}{\sqrt{T}}\left[\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k f_{it}(\xi_{it}(\tau))\left[\frac{\delta_\eta}{\sqrt{T}}+\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right]\right]-\frac{\lambda_T}{\sqrt{T}}\left[\mathrm{sgn}\left(\eta_i-\frac{\delta_{\eta_i}}{\sqrt{T}}\right)-\mathrm{sgn}(\eta_i)\right]+R_{it}.$$

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

$$\widehat{\delta}_\eta = -(\overline{f}_i)^{-1}\left[\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k f_{it}(\xi_{it}(\tau))\left[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right]-\frac{1}{\sqrt{T}}\sum_{k=1}^{K}\sum_{t=1}^{T}\upsilon_k\psi_\tau(y_{it}-\xi_{it}(\tau))+\lambda_0 s\right]+R_{it},$$

where $\overline{f}_i=T^{-1}\sum_k\sum_t\upsilon_k f_{it}$, R_{it} is the remainder term for each i, and noticing that $(\lambda_T/\sqrt{T})\mathrm{sgn}(\eta_i)\to\lambda_0 s$ with $s_0=\mathrm{sgn}(\eta_i)$.

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

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

where $X_{it}=(w'_{it},x'_{it})'$, and $\delta_\alpha(\alpha_n)=\sqrt{NT}(\alpha_n(\tau)-\alpha(\tau))$, $\delta_\beta(\alpha_n)=\sqrt{NT}(\widehat{\beta}(\alpha_n,\tau)-\beta)$, and $\delta_\gamma(\alpha_n)=\sqrt{NT}(\widehat{\gamma}(\alpha_n,\tau)-0)$.

According to Koenker (2004)

$$\sup_{\|\delta\|<C_2}\|g_i(\delta_\alpha,\delta_\beta,\delta_\gamma)-g_i(0,0,0)-E[g_i(\delta_\alpha,\delta_\beta,\delta_\gamma)-g_i(0,0,0)]\|=o_p(1)$$

and at the minimizer, $G(\widehat{\delta}_\alpha,\widehat{\delta}_\beta,\widehat{\delta}_\gamma)=o((NT)^{-1})$.

Expanding, as above,

$$\begin{aligned}
&E[G(\delta_\alpha,\delta_\beta,\delta_\gamma)-G(0,0,0)]\\
&\quad=\frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\left(\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}+\frac{\widehat{\delta}_\eta}{\sqrt{T}}\right)+o_p(1)\\
&\quad=\frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\left[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right.\\
&\qquad-\frac{1}{\sqrt{T}}\overline{f}_i^{-1}\left[T^{-1/2}\sum_k\sum_t\upsilon_k f_{it}(\xi_{it}(\tau))\left[\frac{\delta_\alpha}{\sqrt{NT}}y_{it-1}+\frac{\delta_\beta}{\sqrt{NT}}x_{it}+\frac{\delta_\gamma}{\sqrt{NT}}w_{it}\right].\right.\\
&\qquad\left.\left.-T^{-1/2}\sum_k\sum_t\upsilon_k\psi_\tau(y_{it}-\xi_{it}(\tau))+\lambda_0 s_0\right]+R_{it}\right]+o_p(1)\\
&\quad=\frac{1}{NT}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\left(y_{it-1}-\overline{f}_i^{-1}T^{-1}\sum_k\sum_t\upsilon_k f_{it}y_{it-1}\right)\delta_\alpha\\
&\qquad+\frac{1}{NT}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\left(x_{it}-\overline{f}_i^{-1}T^{-1}\sum_k\sum_t\upsilon_k f_{it}x_{it}\right)\delta_\beta\\
&\qquad+\frac{1}{NT}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\left(w_{it}-\overline{f}_i^{-1}T^{-1}\sum_k\sum_t\upsilon_k f_{it}w_{it}\right)\delta_\gamma\\
&\qquad-\frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\overline{f}_i^{-1}T^{-1/2}\sum_k\sum_t\upsilon_k\psi(y_{it}-\xi_{it})\\
&\qquad+\frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}\overline{f}_i^{-1}\lambda_0 s_0-\frac{1}{\sqrt{NT}}\sum_k\sum_i\sum_t\upsilon_k X_{it}f_{it}R_{it}/\sqrt{T}+o_p(1),
\end{aligned}$$

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 $G(\widehat{\delta}_\alpha,\widehat{\delta}_\beta,\widehat{\delta}_\gamma)=0$ and then $E[G(\delta_\alpha,\delta_\beta,\delta_\gamma)-G(0,0,0)]=G(0,0,0)$; let $\Phi(\lambda)=\mathrm{diag}(f_{it}(\xi_{it}(\tau)))$, $P_Z=Z(Z'\Phi(\lambda)Z)^{-1}Z\Phi(\lambda)$, Ψ_τ be an NT-vector $(\psi_\tau(y_{it}-\xi_{it}(\tau)))$, and s is an NT-vector $\mathrm{sgn}(\eta_i)$. Define $\delta_\vartheta=(\delta_\beta,\delta_\gamma)$, then we have

$$\begin{aligned}
&\upsilon(X'M_Z\Phi M_Z Y_{-1})\delta_\alpha+\upsilon(X'M_Z\Phi M_Z X)\delta_\vartheta-\upsilon X'P_Z\Psi-\upsilon X'P_Z\lambda_0 s=-\upsilon X'\Psi+R_{nT},\\
&\quad J_\vartheta\widehat{\delta}_\vartheta=[-(\upsilon X'M_Z\Psi-\upsilon X'P_Z\lambda_0 s-R_{nT})-J_\alpha\delta_\alpha],\\
&\quad \widehat{\delta}_\vartheta=J_\vartheta^{-1}[-(X'M_Z\Psi-X'P_Z\lambda_0 s-R_{nT})-J_\alpha\delta_\alpha],
\end{aligned}$$

where

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

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

$$\widehat{\delta}_\gamma = \bar{J}_\gamma[-(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s - R_{nT}) - J_\alpha \delta_\alpha], \tag{16}$$

$$\widehat{\delta}_\beta = \bar{J}_\beta[-(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s - R_{nT}) - J_\alpha \delta_\alpha]. \tag{17}$$

The remainder term R_{nT} has a dominant component that comes from the Bahadur representation of the η's. By (A1)–(A5), we have for a generic constant C_3,

$$R_{nT} = T^{-1/4} \frac{KC_3}{\sqrt{N}} \sum_{i=1}^{N} R_{i0} + o_p(1).$$

Using the same argument as in Koenker (2004), 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 (A4) ensure that the contribution of the remainder is negligible. Thus, (16) and (17) simplify to

$$\widehat{\delta}_\gamma = \bar{J}_\gamma[-(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s) - J_\alpha \delta_\alpha],$$

$$\widehat{\delta}_\beta = \bar{J}_\beta[-(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s) - J_\alpha \delta_\alpha].$$

By consistency, $wp \to 1$,

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

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

$$\widehat{\delta}_\alpha = -[J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} [J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma (\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s)].$$

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

$$\begin{aligned}\widehat{\delta}_\beta &= -\bar{J}_\beta[(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s) - J_\alpha [J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma (\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s)] \\ &= -\bar{J}_\beta[(I - J_\alpha [J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma)(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s)].\end{aligned}$$

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

$$\begin{aligned}\widehat{\delta}_\gamma &= -\bar{J}_\gamma[(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s) - J_\alpha [J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma (\upsilon X' M_Z \Psi)] \\ &= -\bar{J}_\gamma[(I - J_\alpha [J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma J_\alpha]^{-1} J_\alpha' \bar{J}_\gamma A \bar{J}_\gamma)(\upsilon X' M_Z \Psi - \upsilon X' P_Z \lambda_0 s)].\end{aligned}$$

By condition (A5), using the fact that $\bar{J}_\gamma J_\alpha$ is invertible

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

Let $\Psi_k = \text{diag}(\psi_{\tau_k}(y_{it} - \xi_{it}(\tau_k)))$. Noticing that $\Psi_k 1_{NT} 1_{NT}' \Psi_l = (\tau_k \wedge \tau_l - \tau_k \tau_l) I_{NT}$, and that conditions (A1)–(A5) imply a CLT. Thus, neglecting the remainder term, and using the definition of δ_α and δ_β we have

$$\sqrt{NT}(\widehat{\theta}(\lambda) - \theta(\lambda)) \xrightarrow{d} N(Bias(\lambda), \Omega),$$

where $\Omega = (K', L')' S (K', L')$, S, H and all the elements with J are defined as in (A5), $K = (J_\alpha' H J_\alpha)^{-1} J_\alpha H$, $L = \bar{J}_\beta M$, $M = I - J_\alpha K$, $Bias(\lambda) = (K', L')' V \lambda_0 s$, with $V = (\upsilon_1 X' M_Z', \ldots, \upsilon_q X' M_Z')$ and s is an NT-vector sgn (η_i). □

References

Abrevaya, J., Dahl, C.M., 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., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5–27.

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

Anderson, T.W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18, 47–82.

Arellano, M., 2003. Panel Data Econometrics. Oxford University Press, New York.

Arellano, M., Bond, S., 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., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 68, 29–51.

Arellano, M., Hahn, J., 2006. A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. CEMFI, Mimeo.

Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: some recent developments. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics, and Econometrics, Ninth World Congress, vol. III. Cambridge University Press, pp. 381–409.

Belloni, A., Chernozhukov, V., 2009. L1-penalized quantile regression in high-dimensional sparse models. Mimeo.
Bester, C.A., Hansen, C., 2009. A penalty function approach to bias reduction in non-linear panel models with fixed effects. Journal of Business and Economic Statistics 27, 29–51.
Blundell, R., Bond, S.R., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.
Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effects models. Journal of Econometrics 132, 491–525.
Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: a robust inference approach. Journal of Econometrics 142, 379–398.
Donoho, D., Chen, S., Saunders, M., 1998. Atomic decomposition by basis pursuit. SIAM Journal of Scientific Computing 20, 33–61.
Fama, E.F., French, K.R., 2002. Testing trade-off and pecking order predictions about dividends and debt. Review of Financial Studies 15, 1–34.
Flannery, M.J., Rangan, K.P., 2006. Partial adjustment toward capital structures. Journal of Financial Economics 79, 469–506.
Galvao, A.F., 2009. Essays on quantile regression for dynamic panel data models. University of Illinois at Urbana-Champaign, Ph.D. Dissertation.
Hahn, J., 1999. How informative is the initial condition in the dynamic panel model with fixed effects? Journal of Econometrics 93, 309–326
Hahn, J., Hausman, J., Kuersteiner, G., 2007. Long difference instrumental variables estimation for dynamic panel models with fixed effects. Journal of Econometrics 140, 547–617.
Hahn, J., Kuersteiner, G.M., 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., Kuersteiner, G.M., 2003. Bias reduction for dynamic nonlinear panel models with fixed effects. UC Davis, Mimeo.
Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72, 1295–1319.
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: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, pp. 179–185.
Hsiao, C., 2003. Analysis of Panel Data, second ed. Cambridge University Press, Cambridge, MA.
Jiang, J., 1998. Asymptotic Properties of the Empirical BLUP and BLUE in Mixed Linear Models. Statistica Sinica 8, 861–885.
Knight, K., 2001. Comparing conditional quantile estimators: first and second order considerations. Mimeo.
Knight, K., Fu, W., 2000. Asymptotics for Lasso-type estimators. The Annals of Statistics 28, 755–770.
Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.
Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–49.
Koenker, R., Hallock, K., 2000. Quantile regression an introduction. Manuscript, University of Illinois at Urbana-Champaign.
Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81, 673–680.
Leary, M.T., Roberts, M.R., 2005. Do firms rebalance their capital structure? The Journal of Finance 60, 2575–2619.
Lemmon, M.L., Roberts, M.R., Zender, J.F., 2008. Back to the beginning: persistence and the cross-sectional distribution of capital structure. The Journal of Finance 63, 1575–1608.
Li, Y., Zhu, J., 2008. L_1-norm quantile regression. Journal of Computational and Graphical Statistics 17, 1–23.
Lindley, D., Smith, A., 1972. Bayes estimates for the linear model. Journal of the Royal Statistical Society (B) 34, 1–41.
Machado, J.A.F., 1993. Robust model selection and M-estimation. Econometric Theory 9, 478–493.
Maddala, G.S., Hu, W., 1995. The pooling problem. In: Matyas, L., Sevestre, P. (Eds.), Econometrics of Panel Data. Kluwer Academic Publishers, Boston, MA, pp. 307–322.
Maddala, G.S., Li, H., Srivastava, V.K., 2001. A comparative study of different shrinkage estimators for panel data models. Annals of Economics and Finance 2, 1–30.
Maddala, G.S., Trost, R.P., Li, H., Joutz, F., 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. Journal of Business and Economic Statistics 15, 90–100.
Meyer, M., Woodroofe, M., 2000. On the degrees of freedom in shape-restricted regression. The Annals of Statistics 28, 1083–1104.
Modigliani, F., Miller, M., 1958. The cost of capital, corporation finance, and the theory of investment. American Economic Review 48, 655–669.
Neyman, J., Scott, E.L., 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.
Nychka, D., Gray, G., Haaland, P., Martin, D., O'Connell, M., 1995. A nonparametric regression approach to syringe grading for quality improvement. Journal of the American Statistical Association 90, 1171–1178.
Roberts, M.R., 2002. The dynamics of capital structure: an empirical analysis of a partially observable system. Duke Working paper.
Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statistical Science 6, 15–32.
Stein, C.M., 1981. Estimation of the mean of a multivariate normal distribution. The Annals of Statistics 9, 1135–1151.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B 58, 267–288.
Wang, H., Leng, C., 2007. Unified LASSO estimation by least squares approximation. Journal of the American Statistical Association 102, 1039–1048.
Wang, H., Li, G., Jiang, G., 2007a. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. Journal of Business and Economic Statistics 25, 347–355.
Wang, H., Li, R., Tsai, C.-L., 2007b. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94, 553–568.
Welch, I., 2004. Capital structure and stock returns. Journal of Political Economy 112, 106–131.
Yuan, M., 2006. GACV for quantile smoothing splines. Computational Statistics and Data Analysis 5, 813–829.