Local information theoretic methods for smooth coefficients dynamic panel data models

Francesco Bravo*
University of York

November 2015

Abstract

This paper considers estimation and inference in semiparametric smooth coefficients dynamic panel data models. It proposes a class of local estimators that can be given an interesting information theoretic interpretation, and a number of test statistics that can be used to test for the (local) correct specification of the model and for the constancy of the smooth coefficients. The results of the paper are rather general as they allow for the three cases of "large N, small T", "small N, large T" and "large N, large T", for the possibility that some of the regressors might be correlated with the unobservable errors and for the possibility that some of the variables used in the estimation might not be directly observable. Simulations show that the proposed method have competitive finite sample properties.

Keywords: α-mixing, Cressie-Read discrepancy, Kernel estimation, Instrumental variables

*I am grateful to two referees for very useful comments and constructive suggestions that improved considerably the original version.

Address correspondence to: Department of Economics, University of York, York YO10 5DD, UK. E-mail: francesco.bravo@york.ac.uk. Web Page: https://sites.google.com/a/york.ac.uk/francescobravo/

1 Introduction

This paper considers estimation and inference for semiparametric dynamic panel data models. Panel data are particular type of longitudinal data very popular in both economics and finance, where they are used to control for individual heterogeneity and identify and measure effects that are simply not detectable in pure cross-section or pure time series models. Dynamic panel data models include lags of the dependent variable and are particularly useful to characterize, for example, dynamic (short, medium and long run) economic relationships and the dynamic implications of various financial policies. There is a vast literature on parametric panel data models, see for example Hsiao (2003) and Baltagi (2010). There is also a rapidly expanding literature on nonparametric and semiparametric panel data models. Examples include Henderson, Carroll and Li (2008) who considered a nonparametric fixed-effect panel data model, Henderson and Ullah (2005) and Lin and Carroll (2006) who both considered nonparametric random-effects panel data models. Li and Stengos (1996) and Baltagi and Li (2002) considered a partially linear dynamic panel data models with some regressors possibly being correlated with the unobservable errors, whereas Lee (2014) considered a nonparametric fixed-effect dynamic panel data model. Sun, Carroll and Li (2009) considered a smooth (or varying) coefficient fixed effect panel data model, while both Cai and Li (2008) and Tran and Tsionas (2010) considered smooth coefficients dynamic panel data models. Su and Ullah (2011) provide a recent review on nonparametric and semiparametric panel data models.

Smooth coefficient models, originally proposed by Cleveland, Grosse and Shyu (1991) and Hastie and Tibshirani (1993), include both pure nonparametric and partially linear regression model as special cases; they are very versatile and have been used, for example, in the context of generalized linear models and quasi-likelihood estimation (Cai, Fan and Li 2000), time series (Cai, Fan and Yao 2000) and longitudinal data (Fan and Wu 2008) - see Fan and Zhang (2008) for a recent review. This paper considers a smooth coefficients dynamic panel data model and proposes an estimation approach alternative to that proposed originally by Cai and Li (2008) and by Tran and Tsionas (2010). The former proposed a one step nonparametric generalized method of moment (NPGMM henceforth) estimator that is based on local linear estimation (Fan and Gijbels 1996), whereas the latter proposed a (typically more efficient) two step nonparametric GMM (2NPGMM henceforth) estimator that is based on local (constant) estimation.

This paper proposes a local estimation method for the unknown smooth coefficients parameters that is similar to that proposed by Tran and Tsionas (2010), but as opposed to the latter it does not require the additional estimation of a certain unknown matrix, which is one of the causes of the bias in local GMM estimation of nonparametric estimating equations models, see Bravo (2014) for more details. The proposed method jointly estimates the unknown parameters and a set of probability weights that reflect some auxiliary information characterizing the unknown distribution of the observations using a local version of the Cressie-Read (power) divergence discrepancy. Baggerly (1998) introduced the Cressie-Read discrepancy as a generalization of Owen's (1988) empirical likelihood method for identically and independently distributed observations; Bravo (2002) proposed a modified version of the Cressie-Read discrepancy for α-mixing processes. The proposed estimator is defined as the minimizer of the Cressie-Read discrepancy between the empirical distribution and a constrained multinomial distribution supported on the observations, where the constraint is an estimating equation that represents the available auxiliary information. Given that the Cressie-Read discrepancy can be interpreted as a generalized entropy measure it seems natural to call the resulting estimators nonparametric information theoretic (NPIT henceforth) estimators. Examples of NPIT estimators include the exponential tilting estimator of Kitamura and Stutzer (1997), defined as the minimizer of the Kullback-Liebler divergence (or relative entropy) between the empirical and a constrained multinomial distribution, which was used for example by Bravo (2005) to construct various specification tests in time series regressions. Another important example is the empirical likelihood estimator, which can be interpreted as the minimizer of the reverse Kullback-Liebler between the empirical and the constrained distribution. DiCiccio and Romano (1990) provided a detailed analysis of the connections between empirical and exponential likelihood with the Kullback-Liebler divergence in the context of constructing nonparametric confidence intervals. Associated with the NPIT estimator there are the estimated multinomial probabilities which can be used to construct an efficient estimator of the unknown distribution of the observations, and, as shown by Guggenberger, Ramalho and Smith (2012), to construct Pearson-type goodness of fit test statistics that can be used for inferences in the context of possibly unidentified estimating equations with time series data.

This paper makes three main contributions: first it establishes the asymptotic normality of the proposed NPIT estimator for the three possible scenarios of "large N, small T", in which only the cross section dimension of the panel grows as the sample sizes increases, of "small N, large T", in which only the time series dimension of the panel grows as the sample size increases, and of "large N, large T", in which both the cross section and time series dimensions grow as the sample size increases. This result is rather general since it is also valid when some or all of the regressors are possibly correlated with the unobservable errors, and when some (or all) of the variables used in the estimation, the so-called instruments in the econometric literature, are not directly observable but can be consistently estimated using either fully parametric or nonparametric methods. These two features are important because often in economic and financial applications correlation between regressors and unobservable errors is very likely, and optimal instruments are effectively unknown because they come in the form of conditional expectations of observable variables, see for example Baltagi and Li (2002). This result complements and extends that obtained by Cai and Li (2008) and by Tran and Tsionas (2010) because it considers the case of unobservable instruments and proposes a two step estimation procedure, in which the first step is used to estimate the instruments.

Second it considers the important issue of local correct specification and constancy of (part or all of) the smooth coefficients and proposes two general, easy to implement, test statistics. The first one is based on the Cressie-Read discrepancy criterion itself, whereas the second one uses estimated probabilities to construct statistics that are in the same spirit of Pearson's classical goodness of fit testing. The tests are local in nature, and are asymptotically distribution free being distributed either as a chi-squared random variable or as a nonstandard distribution that is independent of nuisance parameters, hence can be easily simulated. Interestingly these type of test statistics seem not to have been previously considered in the semiparametric panel data literature.

Finally the paper illustrates the finite sample properties of the proposed method using Monte Carlo simulations and compare them with those based on alternative NPGMM estimators. The results of the simulations are encouraging and suggest that the proposed estimators and test statistics have competitive finite sample properties.

The rest of the paper is organized as follows: next section introduces the statistical model and the nonparametric information theoretic estimator. Section 3 develops the asymptotic theory for both the estimators and the test statistics. Section 4 contains the results of the Monte Carlo study and some concluding remarks. All the proofs can be found in a supplementary Appendix.

The following notation is used throughout the paper: a prime indicates transpose, "$tr(\cdot)$" denotes the trace operator, "$\otimes$" denotes Kronecker product, and for any vector v $v^{\otimes 2} = vv'$.

2 The statistical model and the estimators

The smooth coefficients dynamic panel data model considered is

$$y_{it} = x_{it}'\beta_0(u_{it}) + \varepsilon_{it} \quad i = 1, ..., N; t = 1, ..., T, \tag{1}$$

where x_{it} and u_{it} are, respectively, a k and p dimensional vectors of observable regressors, ε_{it} is an unobservable error term and $\beta_0(\cdot)$ is a vector of unknown smooth functions. The vector x_{it} may contain lagged dependent values, typically only y_{it-1}, and a set of contemporaneous and possibly lagged regressors, say $\widetilde{x}_{it}$, while ε_{it} may contain an unobserved time-invariant random variable η_i, which represents unknown heterogeneity in the sample. It is assumed that η_i is uncorrelated with $\widetilde{x}_{it}$ and u_{it}, which excludes the fixed effect specification, and that the regressors $\widetilde{x}_{it}$ might exhibit nonzero correlation with the errors, that is $E(\varepsilon_{it}|\widetilde{x}_{is}) \neq 0$ $(s \leq t)$. Note also that by construction $E(\eta_i|y_{it-1}) \neq 0$. Model (1) encompasses many nonparametric and semiparametric panel data models: without the regressors x_{it}, (1) is a nonparametric random effect model, see Henderson and Ullah (2005), whereas with $x_{it}'\beta_0(u_{it}) = x_{1it}'\beta_{10} + x_{2it}'\beta_{20}(u_{it})$ (1) becomes a partially linear (possibly dynamic) model, see for example Li and Stengos (1996), Li and Ullah (1998) and Baltagi and Li (2002).

Because of the potential correlation between the unobserved heterogeneity variable η_i and the lagged dependent variables and possibly between the regressors $\widetilde{x}_{it}$ and the errors, any semiparametric least squares type of estimator of $\beta_0(\cdot)$ would be inconsistent. Instead, as in Cai and Li (2008) and Tran and Tsionas (2010), this paper assumes that there exists an l dimensional ($l \geq k$) vector of additional variables z_{it}, called instruments in the econometric literature, such that

$$E\left(z_{it}\varepsilon_{it}|u_{it}\right) = 0 \quad a.s.. \tag{2}$$

The restriction (2) provides the basis for the local estimation method of this paper. To be specific for a given point $u_{it} = u \in \mathbb{R}^p$, let π_{it} $(i = 1, ..., N; t = 1, ..., T)$ denote a set of unknown multinomial weights supported on the observations and let

$$\frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(NT\pi_{it}\right)^{\gamma+1} - 1\right] \tag{3}$$

denote the Cressie-Read discrepancy family, where $\gamma \in \mathbb{R}$ is a user specific parameter with the values $\gamma = 0$ and $\gamma = -1$ to be interpreted as limits. Then the local minimum Cressie-Read discrepancy estimator is defined as the solution of the following program

$$\min_{\beta,\pi_{it}} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\left[\left(NT\pi_{it}\right)^{\gamma+1} - 1\right]}{\gamma(\gamma+1)} | \sum_{i=1}^{N} \sum_{t=1}^{T} \pi_{it} = 1, \; \sum_{i=1}^{N} \sum_{t=1}^{T} \pi_{it} z_{it} \varepsilon_{it} K_h\left(u_{it} - u\right) = 0 \right\}, \tag{4}$$

where $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function in $\mathbb{R}^p$ and h is the bandwidth. By a Lagrange multiplier argument it is possible to show that for a fixed β the solution to (4) is

$$\widehat{\pi}_{it}^{CR}(u) = \frac{1}{NT}\left[\left(\widehat{\eta} + \widehat{\xi}' z_{it}\left(y_{it} - x_{it}'\beta\left(u_{it}\right)\right) K_h\left(u_{it} - u\right)\right)\right]^{\frac{1}{\gamma}}, \tag{5}$$

where the estimated Lagrange multipliers $\widehat{\eta}$ and $\widehat{\xi}$ are associated with the restrictions $\sum_{i=1}^{N} \sum_{t=1}^{T} \pi_{it} = 1$ and $\sum_{i=1}^{N} \sum_{t=1}^{T} \pi_{it} z_{it}\left(y_{it} - x_{it}'\beta\left(u_{it}\right)\right) K_h\left(u_{it} - u\right) = 0$, respectively. Inserting (5) into (3) gives the profile local Cressie-Read function

$$\Gamma^{CR}\left(\beta, \widehat{\lambda}, u\right) = -\sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\left(1 + \gamma\widehat{\lambda}(u)' z_{it}\left(y_{it} - x_{it}'\beta\left(u_{it}\right)\right) K_h\left(u_{it} - u\right)\right)^{\frac{\gamma+1}{\gamma}}}{\gamma+1}, \tag{6}$$

where $\widehat{\lambda}(u) = \widehat{\xi}(u) / (\gamma\widehat{\mu})$. Thus the nonparametric estimator

$$\widehat{\beta}(u) := \arg\min_{\beta} \Gamma^{CR}\left(\beta, \widehat{\lambda}, u\right) \tag{7}$$

can be interpreted as the minimizer of the local Cressie-Read discrepancy between the probability weights used by the empirical distribution function and those of a nonparametric likelihood consistent with the localized restriction (2) that is $E\left(z_{it}\varepsilon_{it}|u_{it}=u\right)=0$. For example the profile nonparametric empirical likelihood ($NPEL$) function (corresponding to the limit case $\gamma=-1$) and the exponential tilting ($NPET$) (corresponding to the limit case $\gamma=0$) are given, respectively, by

$$\begin{aligned}
\Gamma^{EL}\left(\beta,\widehat{\lambda},u\right) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\log\left(1-\widehat{\lambda}\left(u\right)'z_{it}\left(y_{it}-x_{it}'\beta\left(u_{it}\right)\right)\right)K_h\left(u_{it}-u\right), \qquad (8)\\
\Gamma^{ET}\left(\beta,\widehat{\lambda},u\right) &= -\sum_{i=1}^{N}\sum_{t=1}^{T}\exp\left(\widehat{\lambda}\left(u\right)'z_{it}\left(y_{it}-x_{it}'\beta\left(u_{it}\right)\right)\right)K_h\left(u_{it}-u\right),
\end{aligned}$$

and the resulting NPEL and NPET estimators are

$$\begin{aligned}
\widehat{\beta}\left(u\right) &:= \arg\min_{\beta}\Gamma^{EL}\left(\beta,\widehat{\lambda},u\right),\\
\widehat{\beta}\left(u\right) &:= \arg\min_{\beta}\Gamma^{ET}\left(\beta,\widehat{\lambda},u\right).
\end{aligned}$$

Note that (6) (and (8)) corresponds to the dual formulation of (4) (see Newey and Smith (2004)) which is very useful both in the analysis of the asymptotic properties of the local estimator $\widehat{\beta}\left(\cdot\right)$ and in its computation.

3 Asymptotic results

This section contains the main result of the paper. As mentioned in the Introduction the results of this paper are valid for the three possible cases of "large N, small T","small N, large T" and "large N, large T". The latter two are particularly useful for economic and financial type of data since they typically exhibit temporal dependence. In terms of estimation, Theorems 1 and 2 consider the case where the instruments are observable; the results for the "large N, small T" and "large N, large T" cases complement those of Cai and Li (2008) and Tran and Tsionas (2010); the result for the "small N, large T" case is new. Theorem 3 is also new as it considers the case of unobservable instruments that can however be estimated either using a parametric or a nonparametric estimator. The theorem shows that there is no estimation effect coming from the first step estimation, that is the proposed two step NPIT (2NPIT henceforth) estimator has the same asymptotic distribution as that of Theorems 1 and 2. In terms of inference, this section considers two general classes of test statistics, calculated at either one specific point or at a set of finite number of points. It is shown that, under a (standard) undersmoothing condition the test statistics are asymptotic distribution free with either a standard asymptotic χ^2 calibration or a nonstandard asymptotic distribution that can easily simulated as it is nuisance parameter free. The tests are also shown to have power against local alternatives and to be consistent. Theorems 4-6 and Corollaries 5.1 and 6.1 consider the hypothesis of local correct specification of (2); Theorems 7-9 and Corollary 9.1 consider the hypothesis of local constancy of some or all of the smooth coefficients.

3.1 One step estimation

Assume that the instruments z_{it} are observable, and let

$$\Omega_0(u) = Var(z_{it}\varepsilon_{it}|u_{it} = u), \Sigma_0(u) = E(z_{it}x'_{it}|u_{it} = u),$$
$$\Omega_{1t}(u_{i1}, u_{it}) = E(z_{i1}z'_{it}\varepsilon_{i1}\varepsilon_{it}|u_{i1}, u_{it}).$$

Furthermore assume that

either

A1 $(y_{it}, x'_{it}, z'_{it}, u'_{it})_{i=1,t=1}^{N,T}$ are i.i.d. across i for fixed t, and are strictly stationary across t for fixed i,

A2 (i) $E(\varepsilon_{it}|z_{it}, u_{it}) = 0$ $a.s.$, $rank\{\Sigma(u)\} = k$ for all u, (ii) $E\|z_{it}x'_{it}\|^2 < \infty$, $E\left\|z_{it}^{\otimes 2}\right\|^2 < \infty$, $E\varepsilon_{it}^2 < \infty$,

A3 (i) for each t $\Omega_{1t}(u_1, u_2)$ and the joint density $f_{1t}(u_1, u_2)$ of u_{i1} and u_{it} are continuous at $u_1 = u, u_2 = u$, (ii) for each u $\Sigma(u)$, the marginal density $f(u)$ of u_{it} and the joint density $f(z, x, u)$ of z_{it}, x_{it} and u_{it} are positive, and $\sup_t \|\Omega_{1t}(u, u) f_{1t}(u)\| < \infty$ (iii) $\beta_0(u)$, $f(u)$, $f(z, x, u)$ are twice continuously differentiable at $u \in \mathbb{R}^p$,

A4 K is a symmetric, nonnegative and bounded second order kernel having compact support,

A5 $h \to 0$ and $Nh^p \to \infty$ as $N \to \infty$,

or

A1' $(y_{it}, x'_{it}, z'_{it}, u'_{it})_{i=1,t=1}^{N,T}$ are i.i.d. across i for fixed t, and are α-mixing with mixing coefficient $\alpha(k) = O(k^{-\tau})$ with $\tau = (2+\delta)(1+\delta)/\delta$ and $\delta > 0$ is defined in A6,

A5' $h \to 0$ and $Th^p \to \infty$ as $T \to \infty$,

A6 for the same $\delta > 0$ defined in A1' $E\left(\|z_{it}\varepsilon_{it}\|^{2(1+\delta)} |u_{it} = u\right)$ and $E\left(\|z_{it}x'_{it}\|^{2(1+\delta)} |u_{it} = u\right)$ are continuous at u,

A7 $T^{(\tau+1)/\tau}h^{p(2+\delta)/(1+\delta)} \to \infty$,

or

A5" $h \to 0$ and $NTh^p \to \infty$ as both $N \to \infty$, $T \to \infty$,

A7' $(NT)^{(\tau+1)/\tau} h^{p(2+\delta)/(1+\delta)} \to \infty$.

The above regularity conditions are fairly standard in the literature on semiparametric panel data models and cover the three possible cases of "large N, small T" (A1-A6), "small N and large T" (A1', A2-A4, A5', A6-A7) and "large N and large T" (A1', A2-A4, A5", A6, A7'). A1 and A1' exclude deterministic and stochastic trends; the rate assumption on the mixing coefficient in A1' is standard in the literature on semiparametric smooth coefficient models for time series, see for example Cai, Fan and Yao (2000). A2(i) implies (2), while the rank condition is sufficient to show the consistency of the NPIT estimator; A2(ii) contains mild moment assumptions on the regressors and the unobservable errors. A3 is a standard smoothness condition on the conditional covariance of the estimating equations of the smooth coefficients and on the marginal density and the joint density of the observable variables. A4 is standard in kernel estimation, but it could be replaced with a weaker one at the expense of a more involved proof. A6 is used to establish the asymptotic normality of the NPIT estimator. Finally the rate assumption in A7 and A7' are standard for local estimators with time series, see for example Cai (2003) and Cai and Li (2008).

Let $\nu_0 = \int K(v)^2 dv$, $\mu_2 = \int v^{\otimes 2} K(v)\, dv$;

Theorem 1 *Under A1-A6*

$$(NTh^p)^{1/2}\left(\widehat{\beta}(u) - \beta_0(u) - \frac{h^2}{2}B(u)\right) \xrightarrow{d} N\left(0, \frac{\nu_0}{f(u)}\Xi_0(u)^{-1}\right),$$

where

$$\begin{aligned} B(u) &= \Xi(u)\Sigma_0(u)'\Omega_0(u)^{-1}[B_1(u), ..., B_p(u)]', \quad \Xi_0(u) = \Sigma_0(u)'\Omega_0(u)^{-1}\Sigma_0(u), \\ B_j(u) &= E\left\{x_{it}z_{it}'\left[tr\left(f(u)\mu_2\frac{\partial^2\beta_0(u)}{\partial u'\partial u_j}\right) + 2\frac{\partial f(z_{it}, x_{it}, u_{it})}{f(z_{it}, x_{it}|u_{it} = u)\partial u_j}\frac{\partial\beta_0(u)}{\partial u'}\right]|u_{it} = u\right\}, \end{aligned}$$

for $j = 1, ..., p$.

Theorem 1 shows that the NPIT estimator has the same asymptotic variance and the same asymptotic mean squared error as that of the 2NPGMM estimator proposed by Tran and Tsionas (2010). Note also that as mentioned in the Introduction, the proposed estimator is typically more efficient than the NPGMM estimator of Cai and Li (2008).

An immediate consequence of the theorem is that the optimal bandwidth h^{opt} minimizing the asymptotic mean squared error is

$$h^{opt} = \left(\frac{1}{NT}\right)^{1/(p+4)}\left(\frac{p\nu_0}{f(u)}tr\left(\Xi_0(u)^{-1}\right)\|B(u)\|^{-2}\right)^{1/(p+4)},$$

which shows that the optimal convergence rate is of order $(NT)^{-4/(p+4)}$. Next theorem shows that the result of Theorem 1 holds also for the cases of finite N and $T \to \infty$ and both N and $T \to \infty$. Note that for the latter case the asymptotic distribution is obtained as T and $N \to \infty$ simultaneously, rather than sequentially, and without imposing any restrictions on the relative expansion rate of N and T. This differs from the case of dynamic fixed effect panel data models, where, because of the presence of the fixed effect itself, it is typically assumed that $\lim_{N,T\to\infty} N/T = c$, where $0 < c < \infty$, see for example Hahn and Kuersteiner (2002) and Lee (2014).

Theorem 2 *Under A1', A2-A4, A5', A6-A7, or under A1', A2-A4, A5" , A6, A7'*

$$(NTh^p)^{1/2}\left(\widehat{\beta}(u) - \beta_0(u) - \frac{h^2}{2}B(u)\right) \xrightarrow{d} N\left(0, \frac{\nu_0}{f(u)}\Xi_0(u)^{-1}\right).$$

3.2 Two step estimation

This section considers the case where the instruments are not directly observable but are unique (at least locally and/or possibly up to an additive constant) and can be consistently estimated. For example as in Baltagi and Li (2002) the instruments could take the form of a conditional expectation $z_{(j)it} = E\left(v_{(j)it}|w_{(j)it}\right)$, where for $j = 1, ..., l$ $v_{(j)it}$ and $w_{(j)it} \in \mathbb{R}^q$ are both observable and can contain, respectively, lagged values of the dependent variable and some of the regressors and u_{it}. For the parametric estimation case we assume that

$$z_{(j)it} = g\left(w_{(j)it}, \gamma\right)$$

for some known continuously differentiable function $g : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$, and that there exists a unique unknown parameter vector $\gamma_0 \in \Gamma$, such that $rank\left[E\left(\partial g\left(w_{(j)it}, \gamma_0\right)/\partial\gamma'\right)\right] = q$ $(j = 1, ..., l)$. In this case the estimated instruments are $\widehat{z}_{(j)it} = g\left(w_{(j)it}, \widehat{\gamma}\right)$. For the nonparametric estimation case, identification of the instruments follows by the uniqueness (up to a constant) of the conditional expectation and the condition $rank\left(\partial z_{(j)it}/\partial w_{(j)it}\right) = q$ $a.s.$ $(j = 1, ..., l)$. In this case the instruments are estimated using the leave one out kernel

$$\widehat{z}_{(j)it} = \sum_{1\leq m\neq i\leq NT} W_b\left(w_{(j)mt} - w_{(j)it}\right) v_{(j)it},$$

where $W_b(\cdot) = W(\cdot/b)/b$ is a kernel function in $\mathbb{R}^q$ and b is another bandwidth. Let

$$\begin{aligned} \Omega_{1t}^{\partial g}\left(u_{i1}, u_{it}\right) &= E\left(\frac{\partial g\left(w_{i1}, \gamma_0\right)}{\partial\gamma}\frac{\partial g\left(w_{it}, \gamma_0\right)}{\partial\gamma'}\varepsilon_{i1}\varepsilon_{it}|u_{i1}, u_{it}\right), \qquad (9)\\ \Omega_{1t}^{\partial gz}\left(u_{i1}, u_{it}\right) &= E\left(\frac{\partial g\left(w_{i1}, \gamma_0\right)}{\partial\gamma}z_{it}'\varepsilon_{i1}\varepsilon_{it}|u_{i1}, u_{it}\right); \end{aligned}$$

assume that

A3' (i) for each t $\Omega_{1t}(u_1, u_2)$ and the joint density $f_{1t}(u_1, u_2)$ of u_{i1} and u_{it} are continuous at $u_1 = u, u_2 = u$, (ii) for each u the marginal density $f(u)$ of u_{it} are positive, and $\sup_t \|\Omega_{1t}(u,u) f_{1t}(u)\| < \infty$, $\sup_t \left\|\Omega_{1t}^{\partial g}(u,u) f_{1t}(u)\right\| < \infty$, $\sup_t \left\|\Omega_{1t}^{\partial gz}(u,u) f_{1t}(u)\right\| < \infty$, (iii) $\beta_0(u)$, $f(u)$, $f(z,x,u)$ are twice continuously differentiable at $u \in \mathbb{R}^p$, (iv) for each w the marginal density $f(w)$ of w_{it} is positive,

A4' The kernels K and W are symmetric, nonnegative and bounded second order kernels with compact support,

A8 *either* (i) $\|\widehat{\gamma} - \gamma_0\| = O_p\left((NT)^{-1/2}\right)$, $E\sup_{\gamma\in\Gamma}\|[\partial g(w_{it}, \gamma)/\partial\gamma']\|^2 < \infty$ *or* (ii) $b \to 0$ and $NTb^q/\log(NT) \to \infty$ as $NT \to \infty$.

The following theorem shows that the 2NPIT estimator is asymptotically equivalent to the NPIT estimator.

Theorem 3 *Under conditions A1-A2, A3'–A5" or A1', A2, A3"-A5', A6-A7 the result of Theorems 1 and 2 holds.*

3.3 Inference

This section considers the important problem of testing for the local correct specification of (2) and for the constancy of the smooth coefficients $\beta(\cdot)$. Two types of test statistics are proposed: the first one is based on the profile Cressie-Read function (6), whereas the second one is based on the local estimated probabilities $\widehat{\pi}_{it}(\cdot)$ defined in (5).

The null hypothesis of correct local specification[1] at a point $u_{it} = u$ is

$$H_0 : E(z_{it}\varepsilon_{it}|u_{it} = u) = 0, \tag{10}$$

which can be tested using the local NPIT distance statistic $D^{CR}(\cdot)$

$$D^{CR}(u) = 2\left(\Gamma^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - \Gamma^{CR}\left(\widehat{\beta}, 0, u\right)\right).$$

Theorem 4 *Under the assumptions of Theorems 1, 2 or 3, if $NTh^{p+4} \to 0$, then under the null hypothesis (10)*

$$D^{CR}(u) \xrightarrow{d} \chi^2(l-k).$$

An alternative way to test (10) is to use the estimated probabilities (5) expressed in their dual formulation

$$\widehat{\pi}_{it}^{CR}(\beta, \lambda, u) = \frac{1}{NT}\left(1 + \gamma\lambda(u)' z_{it}\left(y_{it} - x_{it}'\beta(u_{it})\right) K_h(u_{it} - u)\right)^{\frac{1}{\gamma}}.$$

Since in the absence of the restriction (10) the estimated probabilities solutions to (4) are given by $\widehat{\pi}_{it}^{CR}(\beta, 0, u) = 1/(NT)$, it follows that the following two Pearson's goodness of fit type of statistics

$$\begin{aligned} P_1^{CR}(u) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2, \\ P_2^{CR}(u) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\left(NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2}{NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)} \end{aligned} \tag{11}$$

can be used to test (10).

Theorem 5 *Under the same assumptions of Theorem 4*

$$P_1^{CR}(u), P_2^{CR}(u) \xrightarrow{d} \chi^2(l-k).$$

[1] It is important to emphasize the local nature of the hypothesis, meaning that the model could still be misspecified even if H_0 is true. I am indebited to a referee for pointing this out.

To investigate the power properties of $D^{CR}(\cdot)$ and $P_j^{CR}(\cdot)$ $(j=1,2)$ the following Pitman type alternative at the point $u_{it}=u$ is considered

$$H_a : E\left(z_{it}\varepsilon_{it} + \gamma_{NT}(u_{it})\,|u_{it}=u\right)=0, \tag{12}$$

for a continuous bounded function $\gamma_{NT}:\mathbb{R}^p \to \mathbb{R}^l$ that may depend on NT.

Corollary 5.1 *Under the same assumption of Theorem 4, if $NTh^{p+4}\to 0$ and $(NTh^p)^{1/2}\gamma_{NT}(u)\to \gamma(u)>0$ (for some $\|\gamma(u)\|<\infty$), then under the alternative hypothesis (12)*

$$D^{CR}(u), P_1^{CR}(u), P_2^{CR}(u) \xrightarrow{d} \chi^2(\kappa, l-k),$$

where $\chi^2(\kappa, l-k)$ is the noncentral chi-squared distribution with noncentrality parameter

$$\kappa = f(u)\gamma(u)'\left(\Omega_0(u)^{-1}\left(I-\Sigma_0(u)\Xi_0(u)^{-1}\Sigma_0(u)\right)\Omega_0(u)^{-1}\right)\gamma(u)/v_0.$$

If $NTh^{p+4}\to 0$ and $(NTh^p)^{1/2}\gamma_{NT}(u)\to\infty$, then under the alternative hypothesis (12)

$$D^{CR}(u), P_1^{CR}(u), P_2^{CR}(u) \xrightarrow{p} \infty.$$

Corollary (5.1) shows that the proposed tests have power against Pitman type alternatives and are consistent against any fixed alternatives of the form $\gamma_{NT}(\cdot)=\gamma(\cdot)$.

It is important to note that the test statistics of Theorems 4 and 5 are asymptotically valid at a single point u; if one wants to consider them over a fixed range of values of u, say $\{u_j\}_{j=1}^m$, they can be replaced by the following test statistics

$$\max_{1\le j\le m} D^{CR}(u_j), \max_{1\le j\le m} P_1^{CR}(u_j) \text{ and } \max_{1\le j\le m} P_2^{CR}(u_j). \tag{13}$$

Theorem 6 *Under the same assumptions of Theorem 4 for distinct $\{u_j\}_{j=1}^m$*

$$\max_{1\le j\le m} D^{CR}(u_j), \max_{1\le j\le m} P_1^{CR}(u_j), \max_{1\le j\le m} P_2^{CR}(u_j) \xrightarrow{d} \max_{1\le j\le m}\chi_j^2(l-k).$$

Notice that the distribution of Theorem 6 is nonstandard but it can be evaluated numerically or easily simulated since it does not depend on any nuisance parameters. Alternatively for m large enough one could use the fact that the asymptotic distribution of an appropriately scaled $\max_j \chi_j^2(p)$ random variable converges to a Gumbel distribution[2] (see Embrechts, Kluppelberg and Mikosch (1997, p.156)).

The power properties of the test statistics (13) are established in the next corollary.

[2] To be specific, if $\gamma\sim\Gamma(\alpha,\beta)$ (Gamma distribution with shape parameter α and scale parameter β), then $a_m\left(\max_j \gamma_j - b_m\right)\xrightarrow{d}\Lambda$ as $m\to\infty$, where , $a_m=\beta$, $b_m=\beta\left(\ln m+(\alpha-1)\ln\ln m-\ln\Gamma(\alpha)\right)$ and Λ is a Gumbel random variable, that is $\Pr(\Lambda\le x)=\exp(-\exp(-x))$. Given that a chi-squared with p degrees of freedom is a $\Gamma(p/2,2)$ random variable, it follows that

$$2\left(\max_j \chi_j^2(p) - 2\left(\ln m+(p-2)/2\ln\ln m-\ln\Gamma(p/2)\right)\right)\xrightarrow{d}\Lambda.$$

Corollary 6.1 *Under the same assumptions of Theorem 4 for distinct* $\{u_j\}_{j=1}^m$ *if* $NTh^{p+4} \to 0$ *and* $(NTh^p)^{1/2}\gamma_{NT}(u_j) \to \gamma(u_j) > 0$ *(for some* $\|\gamma(u_j)\| < \infty$*,* $j = 1, ..., m$*), then under the alternative hypothesis* (12) *at* $u_{it} = u_j$ $(j = 1, ..., m)$

$$\max_{1\leq j\leq m} D^{CR}(u_j), \max_{1\leq j\leq m} P_1^{CR}(u_j), \max_{1\leq j\leq m} P_2^{CR}(u_j) \xrightarrow{d} \max_{1\leq j\leq m} \chi_j^2(\kappa_j, l-k),$$

where

$$\kappa_j = f(u_j)\gamma(u_j)'\left(\Omega_0(u_j)^{-1}\left(I - \Sigma_0(u_j)\Xi_0(u_j)^{-1}\Sigma_0(u_j)\right)\Omega_0(u_j)^{-1}\right)\gamma(u_j)/v_0.$$

If $NTh^{p+4} \to 0$ *and* $(NTh^p)^{1/2}\gamma_n(u_j) \to \infty$ *(for some* $\|\gamma(u_j)\| < \infty$*,* $j = 1, ..., m$*), then under the alternative hypothesis* (12) *at* $u_{it} = u_j$ $(j = 1, ..., m)$

$$\max_{1\leq j\leq m} D^{CR}(u_j), \max_{1\leq j\leq m} P_1^{CR}(u_j), \max_{1\leq j\leq m} P_2^{CR}(u_j) \xrightarrow{p} \infty.$$

The null hypothesis of constancy of some (or all) of the smooth coefficients $\beta(\cdot)$ at $u_{it} = u$ can be expressed as

$$H_0 : \beta^{(p)}(u) = \beta^{(p)}, \tag{14}$$

where $\beta^{(p)}(\cdot)$ denotes the vector containing the first p $(1 \leq p \leq k)$ elements of $\beta(\cdot)$ $(p \leq k)$, so that for $p = k$ (14) implies that the whole smooth coefficients vector $\beta(\cdot)$ is assumed constant. Let

$$\widetilde{\beta}(u) = \arg\min_{\beta} \Gamma^{CR}\left(\beta, \widetilde{\lambda}, u\right) \text{ s.t. } \beta^{(p)}(u) - \beta^{(p)} = 0$$

denote the constrained estimator[3] and let

$$D_{(p)}^{CR}(u) = 2\left(\Gamma^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - \Gamma^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right)\right),$$

denote the resulting NPIT distance statistic.

Theorem 7 *Under the same assumption of Theorem 4, then under the null hypothesis* (14)

$$D_{(p)}^{CR}(u) \xrightarrow{d} \chi^2(p).$$

The null hypothesis (14) can also be tested using the same Pearson goodness of fit type of statistics based on comparing the unconstrained $\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)$ and constrained $\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right)$ estimated probabilities; let

$$\begin{aligned} P_3^{CR}(u) &= \sum_{i=1}^N \sum_{t=1}^T \left(NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right) - NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)\right)^2, \\ P_4^{CR}(u) &= \sum_{i=1}^N \sum_{t=1}^T \frac{\left(NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right) - NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)\right)^2}{NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)} \text{ or} \\ &= \sum_{i=1}^N \sum_{t=1}^T \frac{\left(NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right) - NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)\right)^2}{NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right)}. \end{aligned}$$

[3] Note that for $p = k$ the resulting constrained estimator does not depend on u, that is $\widetilde{\beta}(u) = \widetilde{\beta}$.

Theorem 8 *Under the same assumptions of Theorem 4, then under the null hypothesis (14)*

$$P_3^{CR}(u), P_4^{CR}(u) \xrightarrow{d} \chi^2(p).$$

As with the statistics $D^{CR}(\cdot)$, $P_1^{CR}(\cdot)$ and $P_2^2(\cdot)$ the following theorem allows for the possibility of testing the null hypothesis (14) at different points $\{u_j\}_{j=1}^m$.

Theorem 9 *Under the same assumptions of Theorem 4 for distinct $\{u_j\}_{j=1}^m$*

$$\max_{1\le j\le m} D_{(p)}^{CR}(u_j), \max_{1\le j\le m} P_3^{CR}(u_j), \max_{1\le j\le m} P_4^{CR}(u_j) \xrightarrow{d} \max_{1\le j\le m} \chi_j^2(p).$$

Finally to investigate the power properties of the test statistics $D_{(p)}^{CR}(\cdot)$, $P_3^{CR}(\cdot)$ and $P_4^{CR}(\cdot)$ and their max version it should be noted first that none of them can detect Pitman alternatives drifting at the parametric rate $(NT)^{-1/2}$. The test however will still be consistent for $\widetilde{\beta} \xrightarrow{p} \overline{\beta}$, where $\overline{\beta}$ is such that $\left\|E\left[z_{it}\left(y_{it}-x_{it}'\overline{\beta}\right)\right]\right\|>0$. To specify an alternative Pitman hypothesis we consider

$$H_a : \beta^{(p)}(u_{it}) = \beta^{(p)} + \gamma_{NT}^{(p)}(u_{it}) \ \ a.s., \tag{15}$$

for a continuous bounded function $\gamma_{NT}^{(p)} : \mathbb{R}^p \to \mathbb{R}^p$ that may depend on NT. The following corollary shows that the proposed tests have power against the Pitman alternatives given in (15) at the point $u_{it} = u$ and/or different points $\{u_j\}_{j=1}^m$, and are consistent against any fixed alternative.

Corollary 9.1 *Under the same assumptions of Theorems 7- 9, if $NTh^{p+4} \to 0$ and $(NTh^p)^{1/2}\gamma_{NT}^{(p)}(u) \to \gamma^{(p)}(u) > 0$ (for some $\left\|\gamma^{(p)}(u)\right\| < \infty$), then under the alternative hypothesis (15) at $u_{it} = u$*

$$D_{(p)}^{CR}(u), P_3^{CR}(u), P_4^{CR}(u) \xrightarrow{d} \chi^2(\kappa, p),$$

where $\kappa = \gamma^p(u)'\left(\Xi_0^{(pp)}(u)\right)^{-1}\gamma^p(u) f(u)/v_0$, and $\Xi_0^{(pp)}(u)$ is the upper left $p\times p$ block of the matrix $\Xi_0(u)^{-1}$ defined in Theorem 1. For distinct $\{u_j\}_{j=1}^m$

$$\max_{1\le j\le m} D_{(p)}^{CR}(u_j), \max_{1\le j\le m} P_3^{CR}(u_j), \max_{1\le j\le m} P_4^{CR}(u_j) \xrightarrow{d} \chi^2(\kappa_j, p),$$

where $\kappa_j = \gamma^p(u_j)'\left(\Xi_0^{(pp)}(u_j)\right)^{-1}\gamma^p(u_j) f(u_j)/v_0$.
If $NTh^{p+4} \to 0$ and $(NTh^p)^{1/2}\gamma_{NT}^{(p)}(u) \to \infty$ (for some $\left\|\gamma^{(p)}(u)\right\| < \infty$), then under the alternative hypothesis (15) at $u_{it} = u$

$$D_{(p)}^{CR}(u), P_3^{CR}(u), P_4^{CR}(u) \xrightarrow{p} \infty,$$

and for distinct $\{u_j\}_{j=1}^m$

$$\max_{1\le j\le m} D_{(p)}^{CR}(u_j), \max_{1\le j\le m} P_3^{CR}(u_j), \max_{1\le j\le m} P_4^{CR}(u_j) \xrightarrow{p} \infty.$$

4 Monte Carlo evidence

This section uses a dynamic panel data model with a random effect component to both illustrate the finite sample performance of the proposed estimators and test statistics and compare them with those based on the two step nonparametric GMM (2NPGMM) approach. The 2NPGMM estimator is defined as

$$\widehat{\beta}(u) = \arg\min_{\beta} J\left(\beta, \widehat{\Omega}, u\right), \tag{16}$$

where

$$\begin{aligned} J\left(\beta, \widehat{\Omega}, u\right) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(y_{it} - x_{it}'\beta\left(u_{it}\right)\right) K_h\left(u_{it} - u\right)\right)' \widehat{\Omega}(u)^{-1} \times \\ &\quad \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(y_{it} - x_{it}'\beta\left(u_{it}\right)\right) K_h\left(u_{it} - u\right)\right), \end{aligned}$$

$$\widehat{\Omega}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \overline{\varepsilon}_{it}^{2} z_{it}^{\otimes 2} K_h\left(u_{it} - u\right) \widehat{f}(u) \int K^2(v)\, dv,$$

with $\overline{\varepsilon}_{it} = y_{it} - x_{it}'\overline{\beta}(u)$ for a preliminary consistent estimator $\overline{\beta}(\cdot)$ and $\widehat{f}(\cdot)$ is a kernel estimator. The 2NPGMM test statistics for both the hypotheses of local correct specification (10) and smooth coefficient constancy (14) are defined, respectively, as

$$\begin{aligned} D^{GMM}(u) &= NTJ\left(\widehat{\beta}, \widehat{\Omega}, u\right), \\ D_{(p)}^{GMM}(u) &= NT\left(J\left(\widetilde{\beta}, \widetilde{\Omega}, u\right) - J\left(\widehat{\beta}, \widehat{\Omega}, u\right)\right), \end{aligned}$$

where

$$\widetilde{\Omega}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{\varepsilon}_{it}^{2} z_{it}^{\otimes 2} K_h\left(u_{it} - u\right) \widehat{f}(u) \int K^2(v)\, dv,$$

$\widetilde{\varepsilon}_{it} = y_{it} - x_{it}'\widetilde{\beta}(u)$ and $\widetilde{\beta}(\cdot)$ is the constrained 2NPGMM estimator defined as

$$\widetilde{\beta}(u) = \arg\min_{\beta} J\left(\beta, \widehat{\Omega}, u\right) \text{ s.t. } \beta^{(p)}(u) - \beta^{(p)} = 0.$$

The asymptotic equivalence between $D^{GMM}(\cdot)$, $D_{(p)}^{GMM}(\cdot)$ and the corresponding NPIT statistics $D^{CR}(\cdot)$, $D_{(p)}^{CR}(\cdot)$ implies that

$$\begin{aligned} & D^{GMM}(u) \xrightarrow{d} \chi^2(l-k), \quad D_{(p)}^{GMM}(u) \xrightarrow{d} \chi^2(p), \\ & \max_{1\le j\le m} D^{GMM}(u_j) \xrightarrow{d} \max_{1\le j\le m} \chi_j^2(l-k), \quad \max_{1\le j\le m} D_{(p)}^{GMM}(u_j) \xrightarrow{d} \max_{1\le j\le m} \chi_j^2(p). \end{aligned} \tag{17}$$

The Monte Carlo design is similar to that considered by Tran and Tsionas (2010), that is

$$y_{it} = \beta_{10}(u_{it}) y_{it-1} + \beta_{20}(u_{it}) x_{it} + \eta_i + \varepsilon_{it}, \tag{18}$$

where

$$\beta_{10}\left(u_{it}\right)=\exp\left(-\left(0.5u_{it}-2.5\right)^2\right),\ \beta_{20}\left(u_{it}\right)=\sin\left(2\pi u_{it}\right),$$

u_{it} is i.i.d. $U[2,4]$, the uniform distribution between 2 and 4, x_{it} is i.i.d. $U[0,3]$, ε_{it} is i.i.d. $N\left(0,\sigma_\varepsilon^2\right)$, η_i is i.i.d. $N\left(0,\sigma_\eta^2\right)$. The simulations consider the two most commonly used (in empirical work) members of the nonparametric Cressie-Read discrepancy, namely nonparametric empirical likelihood (NPEL) and nonparametric exponential tilting (NPET) (both defined in (8)). As in Tran and Tsionas (2010) two sets of instruments are considered: $z_{it}=\left[y_{it-2},u_{it-1},x_{it},x_{it-1}\right]'$ and the optimal (unobserved) instruments

$$z_{it}=\left[E\left(y_{it-1}|u_{it-1}\right),E\left(y_{it-1}|u_{it-2}\right),E\left(y_{it-1}|u_{it-1},u_{it-2}\right),x_{it}\right]'$$

(see Baltagi and Li (2002)). The unknown smooth coefficients $\beta_{j0}(\cdot)$ $(j=1,2)$, density $f(\cdot)$ and optimal instruments are estimated using the Epanechnikov kernel with bandwidth chosen by least squares cross-validation. Tables 1 and 2 report the mean square error (MSE) of the two estimators $\widehat{\beta}_j(\cdot)$ for two combinations of the variances σ_ε^2 and σ_η^2, sample sizes N, T using both the observed instruments z_{it} and the optimal (estimated) instruments

$$\widehat{z}_{it}=\left[\widehat{E}\left(y_{it-1}|u_{it-1}\right),\widehat{E}\left(y_{it-1}|u_{it-2}\right),\widehat{E}\left(y_{it-1}|u_{it-1},u_{it-2}\right),x_{it}\right]',$$

respectively. The results are based on 5000 replications, which implies that the Monte Carlo standard error is approximately 0.003.

Tables 1 and 2 approx here

The results of Tables 1 and 2 suggest that both the NPEL and the NPET estimators perform better than the 2NPGMM estimator. As expected, the estimators based on the optimal instruments are characterized by a smaller MSE than those based on the observed instruments. Note also that increasing the time dimension results in estimators with a slightly lower MSE. Between the NPEL and the NPET estimator, the former seems to have an edge over the latter, which is consistent with the theoretical findings of Bravo (2014).

The finite sample properties of the test statistics of Section 3.3 are investigated considering only the case of optimal instruments with the null hypothesis specified as

$$H_0:\beta_{10}(u)=\beta_{10}=0.3,$$

versus a sequence of alternatives indexed by $\delta=[0.0.2,0.4,0.6,0.8,1]$

$$H_1=\beta_{10}+\delta\left(\beta_{10}(u)-\beta_{10}\right).$$

Table 3 reports the finite sample size (corresponding to $\delta=0$) at a 0.01 and 0.05 nominal level for the NPEL $D_{(p)}^{EL}(\cdot)$, NPET $D_{(p)}^{ET}(\cdot)$ and 2NPGMM $D_{(p)}^{GMM}(\cdot)$ statistics, and for the two

Pearson type statistics $P_3^{EL}(\cdot)$ and $P_3^{ET}(\cdot)$[4] obtained, respectively, as a by-product of the local empirical likelihood and exponential tilting estimation used to compute $D_{(p)}^{EL}(\cdot)$ and $D_{(p)}^{ET}(\cdot)$. The test statistics are computed at the points $u = 2.5$ and $u = 3.5$ and for two sample sizes: $N = 100, T = 5$ and $N = 100, T = 50$, using 5000 replications and bandwidth fixed at $h = h^{ave}$, where h^{ave} is the average of the 5000 bandwidths used to obtain Table 2.

Figures 1- 4 show the size adjusted power ($\delta = [0.2, 0.4, 0.6, 0.8, 1]$) for the five test statistics considered in Table 3 obtained using 1000 replications for each value of δ.

Figures 1-4 approx here

Table 3 and Figures 1-4 illustrate that the NPIT statistics perform well and are superior to the 2NPGMM statistic both in terms of size and power. NPEL and NPET have similar finite sample properties with the exponential tilting having a slight overall edge in terms of power. Interestingly the local Pearson's goodness of fit type of statistics seem to be characterized by slightly better finite sample properties than those based on the local distance statistics. In particular Table 3 suggests that the Pearson's goodness of fit statistics are the only one with a statistically insignificant (at the 0.05 level) size distortion. It also suggests that the 2NPGMM statistics is always characterized by a statistically significant size distortion.

Table 4 and Figure 5-6 report, respectively, the finite sample size and power of the statistics $\max_j D_{(p)}^{EL}(\cdot)$, $\max_j D_{(p)}^{ET}(\cdot)$, $\max_j D_{(p)}^{GMM}(\cdot)$, $\max_j P_3^{EL}(\cdot)$, $\max_j P_3^{ET}(\cdot)$ evaluated at $\{u_j\}_{j=1}^{10}$ where $u_j = 2 + 0.15j$.

Figures 5-6 approx here

Table 4 and Figures 5-6 confirm the findings of Table 3 and Figures 1-4 as they suggest that the tests based on NPEL and NPET have better finite sample properties than those based on 2NPGMM with the exponential tilting having an edge over the empirical likelihood. Note that in this case also the NPEL and NPET statistics have a statistically significant size distortion.

Overall the results of the simulations are encouraging and suggest that the NPIT approach can be a valid alternative to the 2NPGMM approach that has been used for smooth coefficients dynamic panel data models. NPIT estimators seem to be characterized by a smaller MSE while NPIT test statistics are typically less size distorted and more powerful than those based on 2NPGMM.

5 Supplemental appendix

Throughout the Appendix "CMT", "CLT" and "LLN" denote Continuous Mapping Theorem, Central Limit Theorem, and Law of Large Numbers, respectively. C denotes an arbitrary

[4] The results for the statistics $P_4^{EL}(\cdot)$ and $P_4^{ET}(\cdot)$ are similar to those of $P_3^{EL}(\cdot)$ and $P_3^{ET}(\cdot)$ and thus are not reported.

positive constant that may differ from line to line, $n = NT$ and finally unless otherwise stated $\sum =: \sum_{i=1}^{N} \sum_{t=1}^{T}$.

Proof of Theorem 1. Suppose that for a given u, $\widetilde{\beta}(u) \xrightarrow{p} \beta_0(u)$; let $\widetilde{\varepsilon}_{it} = y_{it} - x_{it}'\widetilde{\beta}(u)$, and note that

$$\widetilde{\varepsilon}_{it} = \varepsilon_{it} + x_{it}'\left(\widetilde{\beta}(u) - \beta_0(u)\right). \tag{19}$$

The same arguments of Cai and Li (2008, Proposition 2(i)) show that

$$\begin{aligned}
\left\| \frac{h^p}{n} \sum \left(z_{it}^{\otimes 2} \varepsilon_{it} x_{it}' \left(\widetilde{\beta}(u) - \beta_0(u)\right) K_h(u_{it} - u)^2 \right) \right\| &= o_p(1), \\
\left\| \frac{h^p}{n} \sum \left(z_{it} x_{it}' \left(\widetilde{\beta}(u) - \beta_0(u)\right) K_h(u_{it} - u) \right)^{\otimes 2} \right\| &\leq \\
\left\| \widetilde{\beta}(u) - \beta_0(u) \right\|^2 \left\| \frac{h^p}{n} \sum \left(z_{it} x_{it}' K_h(u_{it} - u) \right)^{\otimes 2} \right\| &= o_p(1) O_p(1), \\
\left\| \frac{h^p}{n} \sum \left(z_{it} \varepsilon_{it} K_h(u_{it} - u) \right)^{\otimes 2} - f(u) \Omega_0(u) \nu_0 \right\| &= o_p(1),
\end{aligned}$$

hence

$$\left\| \frac{h^p}{n} \sum \left(z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) \right)^{\otimes 2} - \frac{h^p}{n} \sum \left(z_{it} \varepsilon_{it} K_h(u_{it} - u) \right)^{\otimes 2} \right\| = o_p(1), \tag{20}$$

and therefore by the triangle inequality

$$\left\| \frac{h^p}{n} \sum \left(z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) \right)^{\otimes 2} - f(u) \Omega_0(u) \nu_0 \right\| = o_p(1). \tag{21}$$

By a second order Taylor expansion about $\lambda = 0$ and (21) we have that

$$\begin{aligned}
0 &\leq \frac{1}{n} \Gamma^{CR}\left(\widetilde{\beta}, \lambda, u\right) - \frac{1}{n} \Gamma^{CR}\left(\widetilde{\beta}, 0, u\right) = -\lambda(u)' \frac{1}{n} \sum z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) - \\
&\quad \frac{1}{2} \lambda(u)' \frac{1}{n} \sum \left(z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) \right)^{\otimes 2} \lambda(u) \\
&= -\lambda(u)' \frac{1}{n} \sum z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) - \frac{1}{2h^p} \lambda(u)' \Omega(z) f(u) \nu_0 \lambda(u) + o_p(1),
\end{aligned}$$

so that by the quadratic approximation lemma (Fan and Gijbels 1996) the maximizer $\widehat{\lambda}(u)$ of $\Gamma^{CR}\left(\widetilde{\beta}, \lambda, u\right)$ is given by

$$\widehat{\lambda}(u) = -\left(\Omega_0(u) f(u) \nu_0\right)^{-1} \frac{h^p}{n} \sum z_{it} \widetilde{\varepsilon}_{it} K_h(u_{it} - u) + o_p(1). \tag{22}$$

Using (19) the triangle inequality and the CLT applied to $\sum z_{it} \varepsilon_{it} K_h(u_{it} - u)/n$ (see Cai and Li (2008, Theorem 2)) imply that

$$\left\| \widehat{\lambda}(u) \right\| \leq C \left\| \frac{h^p}{n} \sum z_{it} \varepsilon_{it} K_h(u_{it} - u) \right\| + o_p(1) = O_p\left(\frac{n}{h^p}\right)^{-1/2} + o_p(1). \tag{23}$$

Let $\theta_n = -\left(n/h^p\right)^{-1/2}\rho\theta$, where $\|\theta\| = 1$ and $\rho = O_p(1)$; note that

$$\begin{aligned}
\max_{i,t}\left\|z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right\| &\leq \sum \frac{1}{h^p}\left\|z_{it}\varepsilon_{it}K\left(u_{it}-u\right)\right\| \\
&\leq n^{1/2(1+\delta)}\left(\frac{1}{nh^p}\sum\left\|z_{it}\varepsilon_{it}K\left(u_{it}-u\right)\right\|^{2(1+\delta)}\right)^{1/2(1+\delta)} \\
&= O_p\left(n^{1/2(1+\delta)}\right)
\end{aligned} \tag{24}$$

by Jensen's inequality and a standard kernel calculation that shows that

$$\frac{1}{nh^p}\sum\left\|z_{it}\varepsilon_{it}K\left(u_{it}-u\right)\right\|^{2(1+\delta)} = O_p(1).$$

Similarly $\max_{i,t}\left\|z_{it}x_{it}'K_h\left(u_{it}-u\right)\right\| = O_p\left(n^{1/2(1+\delta)}\right)$, hence

$$\begin{aligned}
&\max_{i,t}\left|\theta_n' z_{it}\widetilde{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right| \leq \max_{i,t}\left|\theta_n' z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right| + \\
&\left\|\widetilde{\beta}(u)-\beta_0(u)\right\|\max_{i,t}\left\|\theta_n' z_{it}x_{it}'K_h\left(u_{it}-u\right)\right\| = o_p(1).
\end{aligned} \tag{25}$$

Note also that by (19), (20) and A3, for any unit vector θ

$$\begin{aligned}
&\sigma_{\max}\left(\frac{h^p}{n}\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\right) + o_p(1) \geq \theta'\frac{h^p}{n}\sum\left(z_{it}\widetilde{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\theta \\
\geq\ &\sigma_{\min}\left(\frac{h^p}{n}\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\right) + o_p(1) > 0 + o_p(1),
\end{aligned} \tag{26}$$

where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ denote, respectively, largest and smallest eigenvalues and $\widetilde{\varepsilon}_{it}$ is defined in (19).

Let $\widehat{\beta}(u)$ denote the local minimizer of $\Gamma^{ECR}(\beta,\lambda,u)$,

$$\widehat{\varepsilon}_{it} = y_{it} - x_{it}'\widehat{\beta}(u)$$

denote the resulting residual and assume that $\left\|\widehat{\beta}(u)-\beta_0(u)\right\| = o_p(1)$. Using (26) and as in the proof of Lemma A3 of Newey and Smith (2004), a Taylor expansion about $\theta_n = 0$ shows that

$$\begin{aligned}
\frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},\theta_n,u\right) &= \frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},0,u\right) - \theta_n'\frac{1}{n}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right) - \\
&\quad \frac{1}{2}\theta_n'\frac{1}{n}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\theta_n \\
&\geq \frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},0,u\right) - \left(\frac{h^p}{n}\right)^{1/2}\rho\left\|\frac{1}{n}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)\right\| - C\rho^2\left(\frac{1}{n}\right),
\end{aligned} \tag{27}$$

which implies

$$\begin{aligned}
&\frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},0,u\right) + \rho\left\|\left(\frac{h^p}{n}\right)^{1/2}\frac{1}{n}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)\right\| - O_p\left(\frac{1}{n}\right) \leq \\
&\frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},\theta_n,u\right) \leq \frac{1}{n}\Gamma^{CR}\left(\widehat{\beta},\widehat{\lambda},u\right) \leq \frac{1}{n}\Gamma^{CR}\left(\beta_0,0,u\right) + O_p\left(\left(\frac{h^p}{n}\right)\right).
\end{aligned}$$

Rearranging (27) it follows that

$$\rho\left\|\frac{1}{n}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)\right\|\leq o_p\left(\left(\frac{h^p}{n}\right)^{1/2}\right)+O_p\left(\left(\frac{1}{nh^p}\right)^{1/2}\right)\to 0,$$

which, given (19) with $\widehat{\beta}(\cdot)$ replacing $\widetilde{\beta}(\cdot)$, implies

$$\left\|\frac{1}{n}\sum\left(z_i x_{it}' K_h\left(u_{it}-u\right)\right)\right\|\left\|\widehat{\beta}(u)-\beta_0(u)\right\|=o_p(1).$$

By the rank condition A2(i) it then follows that $\left\|\widehat{\beta}(u)-\beta_0(u)\right\|=o_p(1)$. The asymptotic distribution of $\widehat{\beta}(\cdot)$ is obtained by a standard mean value expansion. By the consistency of $\widehat{\lambda}(\cdot)$ and $\widehat{\beta}(\cdot)$ the first order conditions $0=\partial\Gamma^{ECR}\left(\widehat{\beta},\widehat{\lambda},u\right)/\partial\left(\lambda',\beta'\right)'$ are satisfied with probability approaching 1, hence expanding about 0 and $\beta_0(\cdot)$ we have

$$\begin{aligned}0 &= -\begin{bmatrix}\frac{1}{n}\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)+b_n(u)\\ 0\end{bmatrix}+\\ &\quad \frac{1}{n}\begin{bmatrix}\sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\lambda^{\otimes 2}} & \sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\lambda\partial\beta'}\\ \sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\beta\partial\lambda'} & \sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\beta^{\otimes 2}}\end{bmatrix}\begin{bmatrix}\widehat{\lambda}(u)\\ \widehat{\beta}(u)-\beta_0(u)\end{bmatrix},\end{aligned}$$

where

$$b_n(u)=\frac{1}{n}\sum x_{it}z_{it}'\left(\beta\left(u_{it}\right)-\beta_0(u)\right)K_h\left(u_{it}-u\right), \tag{28}$$

and $\overline{\beta}=:\overline{\beta}(u)$, $\overline{\lambda}=:\overline{\lambda}(u)$ are the mean values. By (25) with $\theta_n=\overline{\lambda}$, $\max_{it}\left|\overline{\lambda}' z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right|=o_p(1)$, where $\overline{\varepsilon}_{it}=y_{it}-x_{it}'\overline{\beta}(u)$ is the mean value residual, hence as in Newey and Smith (2004)

$$\max_{i,t}\left|\left(1+\gamma\overline{\lambda}(u)' z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\frac{1}{\gamma}-j}-1\right|=o_p(1)\ \text{ for } j=0,1; \tag{29}$$

the triangle inequality and (29) show that

$$\begin{aligned}\left\|\frac{1}{n}\sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\lambda^{\otimes 2}}\right\| &\leq \max_{i,t}\left|\left(1+\gamma\overline{\lambda}(u)' z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\frac{1}{\gamma}-1}-1\right|\times\\ &\quad\left\|-\frac{1}{n}\sum\left(z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\right\|+\\ &\quad\left\|-\frac{1}{n}\sum\left(z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\right\|\\ &= \left\|-\frac{1}{n}\sum\left(z_{it}\overline{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}\right\|+o_p(1),\end{aligned}$$

hence by (21)

$$\left\|\frac{h^p}{n}\sum\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\overline{\lambda},u\right)}{\partial\lambda^{\otimes 2}}+f(u)\,\Omega_0(u)\,\nu_0\right\|=o_p(1). \tag{30}$$

Similarly

$$
\begin{aligned}
\left\| \frac{1}{n} \sum \frac{\partial^2 \Gamma^{CR}\left(\overline{\beta}, \overline{\lambda}, u\right)}{\partial \lambda \partial \beta'} \right\| \quad \leq \quad & \max_{i,t} \left| \left(1 + \gamma \overline{\lambda}\left(u\right)' z_{it} \overline{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} - 1 \right| \times \\
& \left\| \frac{1}{n} \sum \left(\overline{\lambda}' z_{it} \overline{\varepsilon}_{it} z_{it} x_{it}' K_h \left(u_{it} - u\right)^2 \right) \right\| + \\
& \left\| \frac{1}{n} \sum \left(\overline{\lambda}' z_{it} \overline{\varepsilon}_{it} z_{it} x_{it}' K_h \left(u_{it} - u\right)^2 \right) \right\| + \\
& \max_{i,t} \left| \left(1 + \gamma \overline{\lambda}\left(u\right)' z_{it} \overline{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}} - 1 \right| \times \\
& \left\| \frac{1}{n} \sum \left(z_{it} x_{it}' K_h \left(u_{it} - u\right)\right) \right\| + \left\| \frac{1}{n} \sum \left(z_{it} x_{it}' K_h \left(u_{it} - u\right)\right) \right\|,
\end{aligned}
$$

and by the Cauchy-Schwarz inequality and the same arguments used to establish (19) and (21) it follows that

$$
\begin{aligned}
\frac{\left\|\overline{\lambda}\right\|}{n} \sum \left\| z_{it} \overline{\varepsilon}_{it} z_{it} x_{it}' K_h \left(u_{it} - u\right)^2 \right\| \quad & \leq \quad \left\|\overline{\lambda}\right\| \left(\frac{1}{n} \sum \left\| z_{it} \overline{\varepsilon}_{it} K_h \left(u_{it} - u\right)\right\|^2 \right)^{1/2} \\
\left(\frac{1}{n} \sum \left\| z_{it} x_{it}' K_h \left(u_{it} - u\right)\right\|^2 \right)^{1/2} \quad & = \quad o_p\left(1\right) O_p\left(1\right),
\end{aligned}
$$

$$
\left\| \frac{1}{n} \sum \left(z_{it} x_{it}' K_h \left(u_{it} - u\right)\right) - f\left(u\right) \Sigma_0\left(u\right) \right\| = o_p\left(1\right),
$$

hence

$$
\left\| \frac{1}{n} \sum \frac{\partial^2 \Gamma^{ECR}\left(\overline{\beta}, \overline{\lambda}, u\right)}{\partial \lambda \partial \beta'} - f\left(u\right) \Sigma_0\left(u\right) \right\| = o_p\left(1\right). \tag{31}
$$

Finally similar arguments can be used to show that

$$
\left\| \sum \frac{\partial^2 \Gamma^{CR}\left(\overline{\beta}, \overline{\lambda}, u\right)}{\partial \beta^{\otimes 2}} \right\| = o_p\left(1\right). \tag{32}
$$

Combining (30)-(32) and the CMT imply

$$
\begin{aligned}
\left(nh^p\right)^{1/2} \left[\begin{array}{c} \widehat{\lambda}\left(u\right)/h^p \\ \widehat{\beta}\left(u\right) - \beta_0\left(u\right) \end{array} \right] \quad = \quad & \left[\begin{array}{cc} -f\left(u\right) \Omega_0\left(u\right) \nu_0 & \Sigma_0\left(u\right) \\ \Sigma_0\left(u\right)' & 0 \end{array} \right]^{-1} \times \\
& \left(nh^p\right)^{1/2} \left[\begin{array}{c} \frac{1}{n} \sum \left(z_{it} \varepsilon_{it} K_h \left(u_{it} - u\right)\right) + b_n\left(u\right) \\ 0 \end{array} \right] + o_p\left(1\right).
\end{aligned} \tag{33}
$$

By a standard kernel calculation

$$
\begin{aligned}
E\left(b_n(u)\right) &= \int x_{it}z_{it}'\left[\frac{\partial\beta_0(u)}{\partial u'}\left(u_{it}-u\right)+\frac{1}{2}\sum_{j=1}^{p}\frac{\partial^2\beta_0(u)}{\partial u'\partial u_j}\left(u_{it}-u\right)\left(u_{it}-u\right)_j\right]\times \\
& K_h\left(u_{it}-u\right)f\left(z_{it},x_{it},u_{it}\right)dz_{it}dx_{it}du_{it} = \\
& \frac{h^2}{2}E\left\{x_{it}z_{it}'\left[tr\left(f(u)\mu_2\frac{\partial^2\beta_0(u)}{\partial u'\partial u_j}\right)+2\frac{\partial f\left(z_{it},x_{it},u_{it}\right)}{f\left(z_{it},x_{it}|u_{it}=u\right)\partial u_j}\frac{\partial\beta_0(u)}{\partial u'}\right]|u_{it}=u]\right\} \\
& +O\left(h^3\right).
\end{aligned}
$$

The asymptotic normality of $(h^p)^{1/2}\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)/n^{1/2}$ can be established using Lyapunov CLT, since A6 can be used to verify the Lyapunov condition - see also Cai and Li (2008, Theorem 2), and the result follows by the CMT. ■

Proof of Theorem 2. The proof of the theorem is the same as that of Theorem 1 with the exception of the CLT used. The first result (i.e. the small N large T) is obtained following closely Cai (2003). For a unit vector θ let $(v_{it})_{t=1}^{T}=\left\{(h^p)^{1/2}\theta' z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right\}_{t=1}^{T}$, which for each i is a stationary α-mixing sequence. Using Proposition 2(ii) of Cai and Li (2008) it is possible to show that $Var\left(\sum_{t=1}^{T}v_{it}/T^{1/2}\right)=\theta' f(u)\Omega_0(u)\nu_0\theta$, hence by the i.i.d. assumption

$$
Var\left(\frac{1}{(NT)^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\right)=\theta' f(u)\Omega_0(u)\nu_0\theta:=\sigma^2(u). \tag{34}
$$

To show the asymptotic normality the indices $1,...,T$ are partitioned using Doob's small-block large block technique into $2q_T+1$ subsets with the large block of size $r=:r_T=\lfloor(nh^p)^{1/2}\rfloor$ and the small one of size $s=:s_T=\lfloor(nh^p)^{1/2}/\log T\rfloor$, $q=:q_T=\lfloor T/(r+s)\rfloor$ where $\lfloor\cdot\rfloor$ is the integer part function and note that $s/r\to 0$, $r/T\to 0$ and $(T/r)\alpha(s)\to 0$. For $0\le j\le q$ let $V_{ij,1}=\sum_{t=j(r+s)+1}^{j(r+s)+r}v_{it}$, $V_{ij,2}=\sum_{t=j(r+s)+r+1}^{(j+1)(r+s)}v_{it}$, $V_{iq}=\sum_{t=q(r+s)+1}^{T}v_{it}$ so that

$$
\frac{1}{n^{1/2}}\sum v_{it}=\frac{1}{n^{1/2}}\sum_{i=1}^{N}\left(\sum_{j=0}^{q-1}\left(V_{ij,1}+V_{ij,2}\right)+V_{iq}\right)=:\frac{U_{n1}+U_{n2}+U_{n3}}{n^{1/2}}.
$$

The same arguments used by Cai (2003) show that $E\left(U_{n2}^2/n\right)=\sum_{i=1}^{N}Var\left(\sum_{j=0}^{q-1}V_{ij,2}/T^{1/2}\right)/N=o(1)$ and $E\left(U_{n3}^2/n\right)=\sum_{i=1}^{N}Var\left(V_{iq}/T^{1/2}\right)/N=o(1)$, hence $U_{n2}=o_p\left(n^{1/2}\right)$ and $U_{n3}=o_p\left(n^{1/2}\right)$. Furthermore for any $1\le i\le N$ and $\iota=(-1)^{1/2}$

$$
\left|E\exp\left(\iota t\sum_{j=0}^{q-1}V_{ij,1}\right)-\prod_{j=0}^{q-1}E\exp\left(\iota tV_{ij,1}\right)\right|\le 16\alpha(T/r)\alpha(s)\to 0 \tag{35}
$$

by Lemma 1.1 of Volkonskii and Rozanov (1959). Note that by A1'

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=0}^{q-1}E\left(V_{ij,1}\right)^2=\frac{qr}{T}\frac{1}{r}Var\left(\sum_{t=1}^{r}v_{it}\right)\to\sigma^2(u), \tag{36}
$$

where $\sigma^2(u)$ is defined in (34). Finally as shown by Cai (2003) $E\left[V_{i1,1}^2 I\left(|V_{i1,1}| \geq \epsilon\sigma(u)T^{1/2}\right)\right] = O\left(T^{-\delta/2}r^{2(2+\delta)}h^{-p(2+\delta)\delta/2(1+\delta)}\right)$ for every $\epsilon > 0$, hence

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=0}^{q-1} E\left(V_{ij,1}^2 I\left(|V_{i1,1}| \geq \epsilon\sigma(u)T^{1/2}\right)\right) = O\left(T^{\delta/4}h^{-p[1+2/(1+\delta)]\delta/4}\right) \to 0 \tag{37}$$

by A7. Thus (35) – (37) imply the Lindeberg-Feller CLT and the result follows by CMT.

For the large N and large T case consider the doubly indexed sequence

$$\{v_{it}\}_{i,t=1}^{n} = \left\{(h^p)^{1/2}\theta' z_{it}\varepsilon_{it}K_h(u_{it}-u)\right\}_{i,t=1}^{n},$$

which is independent across i and stationary $\alpha-$mixing across t, and note that both (35) and (36) are still valid for $N,T \to \infty$. The joint asymptotic normality as $N,T \to \infty$ is established applying Theorem 2 of Phillips and Moon (1999) and verifying the generalized Lindeberg condition

$$\frac{1}{\sigma_N^2(u)}\sum_{i=1}^{N} E\left(V_{ij,1}^2 I\left(|V_{i1,1}| \geq \epsilon\sigma_N(u)\right)\right) \to 0, \tag{38}$$

where $\sigma_N^2(u) = Var\left(\sum_{i=1}^{N}\sum_{j=0}^{q-1} V_{ij,1}/T^{1/2}\right)$. By (36) $\sigma_N^2(u) = O(N)$ and

$$\sup_{1\leq i\leq N} E\left(\sum_{j=1}^{q-1} V_{ij,1}^2/T^{1/2}\right) < \infty,$$

hence Theorem 23.10 of Davidson (1994) implies that (38) holds. Thus by Theorem 2 of Phillips and Moon (1999) $\sum v_{it}/n^{1/2} \xrightarrow{d} N(0,\sigma^2(u))$ and the result follows by CMT. ■

Proof of Theorem 3. For the parametric case $\widehat{z}_{it} = g(w_{it},\widehat{\gamma})$ note that

$$\frac{h^p}{n}\sum\left(\widehat{z}_{it}\varepsilon_{it}K_h(u_{it}-u)\right)^{\otimes 2} = \frac{h^p}{n}\sum\left(z_{it}\varepsilon_{it}K_h(u_{it}-u)\right)^{\otimes 2} + \tag{39}$$
$$2\frac{h^p}{n}\sum\left((\widehat{z}_{it}-z_{it})z_{it}'\left(\varepsilon_{it}K_h(u_{it}-u)\right)^2\right) + \frac{h^p}{n}\sum\left((\widehat{z}_{it}-z_{it})\left(K_h(u_{it}-u)\right)^{\otimes 2}\right).$$

As in Owen (1990), A8(i) and an application of the Borel-Cantelli lemma gives

$$\max_{i,t}\sup_{\gamma\in\Gamma}\left\|\frac{\partial g(w_{it},\gamma)}{\partial\gamma'}\right\| = o_p\left(n^{1/2}\right),$$

hence a mean value expansion and A8(i) show that

$$\max_{i,t}\|\widehat{z}_{it}-z_{it}\| \leq \max_{it}\left\|\frac{\partial g(w_{it},\overline{\gamma})}{\partial\gamma'}\right\|\|\widehat{\gamma}-\gamma_0\| = o_p(1), \tag{40}$$

where $\overline{\gamma}$ is the mean value, hence using (40) in (39) yields

$$\begin{aligned}
&\left\| \frac{h^p}{n} \sum \left(\left(\widehat{z}_{it} - z_{it} \right) z_{it}' \left(\varepsilon_{it} K_h \left(u_{it} - u \right) \right)^2 \right) \right\| \leq \max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\| \times \\
&\left\| \frac{h^p}{n} \sum \left(z_{it} \left(\varepsilon_{it} K_h \left(u_{it} - u \right) \right)^2 \right) \right\| = o_p\left(1\right), \\
&\left\| \frac{h^p}{n} \sum \left(\left(\widehat{z}_{it} - z_{it} \right) \varepsilon_{it} K_h \left(u_{it} - u \right) \right)^{\otimes 2} \right\| \leq \max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\|^2 \times \\
&\left\| \frac{h^p}{n} \sum \left(\varepsilon_{it} K_h \left(u_{it} - u \right) \right)^2 \right\| = o_p\left(1\right),
\end{aligned}$$

hence

$$\frac{h^p}{n} \sum \left(\widehat{z}_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) \right)^{\otimes 2} = \frac{h^p}{n} \sum \left(z_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) \right)^{\otimes 2} + o_p\left(1\right);$$

therefore by triangle inequality

$$\left\| \frac{h^p}{n} \sum \left(\widehat{z}_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) \right)^{\otimes 2} - f\left(u\right) \Omega_0\left(u\right) \nu_0 \right\| = o_p\left(1\right). \tag{41}$$

Let $\widetilde{\varepsilon}_{it} = y_{it} - x_{it}' \widetilde{\beta}\left(u\right)$ for any consistent estimator $\widetilde{\beta}\left(u\right)$; by triangle inequality and similarly to (24)

$$\begin{aligned}
&\max_{i,t} \left| \theta_n' \widehat{z}_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right| \leq \max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\| \max_{i,t} \left| \theta_n' \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right| + \\
&\max_{i,t} \left| \theta_n' z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right| \leq \max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\| \left\| \theta_n \right\| \max_{i,t} \left| \varepsilon_{it} K_h \left(u_{it} - u \right) \right| + \\
&\left\| \widetilde{\beta}\left(u\right) - \beta_0\left(u\right) \right\| \max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\| \left\| \lambda_n \right\| \max_{i,t} \left\| x_{it} K_h \left(u_{it} - u \right) \right\| = o_p\left(1\right).
\end{aligned} \tag{42}$$

Using the same Taylor expansion argument as that of Theorem 1 it can be shows that the 2NPIT estimator $\widehat{\beta}\left(u\right)$ is consistent. To establish the asymptotic normality note that

$$\begin{aligned}
&\frac{1}{n} \sum \left(\widehat{z}_{it} \widehat{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right) = \frac{1}{n} \sum \left(\widehat{z}_{it} - z_{it} \right) \varepsilon_{it} K_h \left(u_{it} - u \right) + \\
&\frac{1}{n} \sum z_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) + \frac{1}{n} \sum \left(\widehat{z}_{it} - z_{it} \right) x_{it}' \left(\widehat{\beta}\left(u_{it}\right) - \beta_0\left(u\right) \right) K_h \left(u_{it} - u \right) + b_n\left(u\right),
\end{aligned} \tag{43}$$

where $b_n\left(u\right)$ is defined in (28). Since

$$\begin{aligned}
&\left(nh^p\right)^{1/2} \left\| \frac{1}{n} \sum \left(\widehat{z}_{it} - z_{it} \right) \varepsilon_{it} K_h \left(u_{it} - u \right) \right\| \leq \\
&\max_{i,t} \left\| \widehat{z}_{it} - z_{it} \right\| \left(nh^p\right)^{1/2} \left\| \frac{1}{n} \sum \varepsilon_{it} K_h \left(u_{it} - u \right) \right\| = o_p\left(1\right) O_p\left(1\right),
\end{aligned} \tag{44}$$

and similarly for

$$\left\| \frac{1}{n} \sum \left(\widehat{z}_{it} - z_{it} \right) x_{it}' \left(\beta\left(u_{it}\right) - \beta_0\left(u\right) \right) K_h \left(u_{it} - u \right) \right\| = o_p\left(\left(nh^p\right)^{-1/2} \right),$$

the conclusion follows by the same arguments as those used in Theorems 1 or 2. For the nonparametric case note first that by Masry (1996) $\sup_{i,t}\|\widehat{z}_{it}-z_{it}\| = o_p(1)$ hence as in (41) and (42)

$$\begin{aligned}\left\|\frac{h^p}{n}\sum\left(\widehat{z}_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)^{\otimes 2}-f\left(u\right)\Omega_0\left(u\right)\nu_0\right\| &= o_p\left(1\right), \qquad (45)\\ \max_{i,t}\left|\lambda_n'\widehat{z}_{it}\widetilde{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right| &= o_p\left(1\right),\end{aligned}$$

and the consistency of the 2NPIT estimator follows as before. By a standard kernel calculation

$$\left(\widehat{z}_{it}-z_{it}\right)=\sum_{j\neq i,t}W_{nb}\left(w_{jt}-w_{it}\right)v_{jt}+o_p\left(1\right), \qquad (46)$$

where $W_{nb}\left(w_{jt}-w_{it}\right)=W_b\left(w_{jt}-w_{it}\right)/\left[\left(n-1\right)f\left(w_{it}\right)\right]$; note that by A1 (or A1') if $i\neq i'$ and $j\neq j'$ the terms involved in the following summation

$$\begin{aligned}&nh^pVar\left(\frac{1}{n}\sum_{i,t}\sum_{j\neq i}W_{nb}\left(w_{jt}-w_{it}\right)v_{jt}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)= \qquad (47)\\ &\frac{h^p}{n}Cov\left(\sum_{i,t}\sum_{j\neq i}W_{nb}\left(w_{jt}-w_{it}\right)v_{jt}\varepsilon_{it}K_h\left(u_{it}-u\right)\right.\\ &\left.\sum_{i',t'}\sum_{j'\neq i'}W_{nb}\left(w_{j't'}-w_{i't'}\right)v_{j't'}\varepsilon_{i't'}K_h\left(u_{i't'}-u\right)\right),\end{aligned}$$

are 0, hence it suffices to consider only the two cases $i=i'$ and $j=j'$. For T finite and $t=t'$ by conditioning first on w_{it} and then on u_{it} and a standard kernel calculation show that

$$\begin{aligned}&\left\|\frac{h^p}{n}\sum_{i,t}\sum_{j\neq i}\sum_{j'\neq i}Cov\left(W_b\left(w_{jt}-w_{it}\right)v_{jt}\varepsilon_{it}K_h\left(u_{it}-u\right),W_b\left(w_{j't}-w_{it}\right)v_{j't}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)\right\|\leq\\ &b^2\left\|f\left(u\right)E\left(Var\left(v_{it}\varepsilon_{it}|w_{it}\right)|u_{it}=u\right)v_0\right\|+\\ &h^pb^2T\left\|f\left(u\right)E\left[Cov\left(v_{it}\varepsilon_{it},v_{is}\varepsilon_{is}|w_{it},w_{is}\right)|u_{it}=u\right]v_0\right\|=O\left(b^2\right)\end{aligned}$$

and similarly for $t\neq t'$; for the case $j=j'$ and $t=t'$ noting that for $\left(u_{it}-u\right)/b=v$ by A5' $\left(u_{i't}-u\right)/h=v+o\left(1\right)$ it follows that

$$\begin{aligned}&\left\|\frac{h^p}{n}\sum_{i,t',t}\sum_{j\neq i}\sum_{j\neq i'}Cov\left(W_b\left(w_{jt}-w_{it}\right)v_{jt}\varepsilon_{it}K_h\left(u_{it}-u\right),W_b\left(w_{jt}-w_{i't}\right)v_{j't}\varepsilon_{it}K_h\left(u_{i't}-u\right)\right)\right\|\leq\\ &b^2\left\|\int Var\left(v_{it}\varepsilon_{it}|w_{1it}=w_1,u_{it}=u\right)w_0dw_{it}v_0\right\|+\\ &h^pb^2T\left\|\int\left[Cov\left(v_{it}\varepsilon_{it},v_{is}\varepsilon_{is}|w_{it},w_{is},u_{it}=u\right)|w_0\right]dw_{it}dw_{is}v_0\right\|,\end{aligned}$$

where $w_0 = \int W(v)^2 dv$.

For $T \to \infty$ and $i = i'$, $t = t'$ the Cauchy-Schwarz inequality applied to (47) shows that

$$\left\| \frac{h^p}{n} \sum_{i,t} \sum_{j \neq i} \sum_{j' \neq i} Cov\left(W_b\left(w_{jt} - w_{it}\right) v_{jt}\varepsilon_{it} K_h\left(u_{it} - u\right), W_b\left(w_{j't} - w_{it}\right) v_{j't}\varepsilon_{it} K_h\left(u_{it} - u\right)\right) \right\|^2 \leq$$
$$\sum_t \alpha(t) f(w) \left|v_0 E\left[Var\left(\varepsilon_{it} | w_{it}\right) | w_{it} = w\right]\right|^{1/2} \times$$
$$\left\|Var\left(v_{jt} E\left[W_b\left(w_{jt} - w_{it}\right) W_b\left(w_{j't} - w_{it}\right) | w_{it}\right]\right)\right\| \leq Cb^2 \sum_t \alpha(t) = O\left(b^2\right)$$

and similarly for $t \neq t'$. For the case $j = j'$ and $t = t'$

$$\left\| \frac{h^p}{n} \sum_{i,t',t} \sum_{j \neq i} \sum_{j \neq i'} Cov\left(W_b\left(w_{jt} - w_{it}\right) v_{jt}\varepsilon_{it} K_h\left(u_{it} - u\right), W_b\left(w_{jt} - w_{i't}\right) v_{j't}\varepsilon_{it} K_h\left(u_{i't} - u\right)\right) \right\|^2 \leq$$
$$b^2 \left\| \frac{1}{n(n-1)} \sum_{i,j,t} \sum_{j \neq i} \alpha(t) \int v_0 Var\left(v_{jt}\varepsilon_{it} | w_{it}\right) w_0 dw_{it} \right\| \times$$
$$\left\| \frac{1}{n(n-1)} \sum_{i,i',t} \sum_{j \neq i} \sum_{j \neq i'} \alpha(t) \int Var\left(v_{jt}\varepsilon_{it} | W_b\left(w_{jt} - w_{it}\right) W_b\left(w_{jt} - w_{i't}\right)\right) w_0 dw_{jt} \right\|$$

and similarly for $t \neq t'$. Hence it follows that

$$\left\| nh^p Var\left(\frac{1}{n} \sum_{i,t} \sum_{j \neq i} W_{nb}\left(w_{jt} - w_{it}\right) v_{jt}\varepsilon_{it} K_h\left(u_{it} - u\right)\right) \right\| = o(1)$$

and

$$\left\| \frac{1}{n} \sum \left(\widehat{z}_{it} - z_{it}\right) \varepsilon_{it} K_h\left(u_{it} - u\right) \right\| = o_p\left(\left(nh^p\right)^{-1/2}\right). \tag{48}$$

Using similar arguments it is possible to show that

$$\left\| nh^p Var\left(\frac{1}{n} \sum_{i,t} \sum_{j \neq i} W_{nb}\left(w_{jt} - w_{it}\right) v_{jt} x_{it}'\left(\beta\left(u_{it}\right) - \beta_0(u)\right) K_h\left(u_{it} - u\right)\right) \right\| = o_p(1),$$

hence

$$\left(\frac{h^p}{n}\right)^{1/2} \sum \left(\widehat{z}_{it}\widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) = \frac{1}{n} \sum z_{it}\varepsilon_{it} K_h\left(u_{it} - u\right) + \widehat{b}_n(u) + o_p(1),$$

and the result follows again by the same arguments as those used in the proofs of Theorems 1 or 2. ■

Proof of Theorem 4. By a second order Taylor expansion about $\lambda = 0$ with Lagrange reminder $\overline{\lambda} =: \overline{\lambda}(u)$ - that is $\overline{\lambda}$ is on the line joining 0 and $\widehat{\lambda}$- it follows that

$$
\begin{aligned}
& \Gamma^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - \Gamma^{CR}\left(\widehat{\beta}, 0, u\right) = -\widehat{\lambda}(u)' \sum z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u) + \\
& \frac{1}{2}\widehat{\lambda}(u)' \frac{1}{n} \sum \frac{\partial^2 \Gamma^{CR}\left(\widehat{\beta}, \overline{\lambda}, u\right)}{\partial \lambda^{\otimes 2}} \widehat{\lambda}(u) \\
= \; & \frac{h^p}{n} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u))' (f(u)\Omega(u)\nu_0)^{-1} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u)) + o_p(1),
\end{aligned}
$$

where the second equality follows using (21) (with $\widehat{\varepsilon}_{it}$ replacing $\widetilde{\varepsilon}_{it}$) and (22). Since

$$
\begin{aligned}
& \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u)) = \sum (z_{it}\varepsilon_{it} K_h(u_{it}-u)) - \sum (z_{it}x_{it}' K_h(u_{it}-u)) \\
& \left(\Sigma_0(u)' \Omega_0(u)^{-1} \Sigma_0(u)\right)^{-1} \Sigma_0(u)' \Omega_0(u)^{-1} \sum (z_{it}\varepsilon_{it} K_h(u_{it}-u)) + o_p\left(\left(\frac{h^p}{n}\right)^{1/2}\right),
\end{aligned}
$$

it follows that

$$
\begin{aligned}
D^{CR}(u) \;=\; & \frac{h^p}{n} \sum (z_{it}\varepsilon_{it} K_h(u_{it}-u))' (f(u)\Omega(u)\nu_0)^{-1/2} M_0(u) \times \\
& (f(u)\Omega(u)\nu_0)^{-1/2} \sum (z_{it}\varepsilon_{it} K_h(u_{it}-u)) + o_p(1),
\end{aligned} \tag{49}
$$

where

$$
\begin{aligned}
M_0(u) \;=\; & I - (f(u)\Omega(u)\nu_0)^{-1/2} \Sigma_0(u) (\Xi_0(u) f(u)/\nu_0)^{-1} \\
& \Sigma_0(u)' (f(u)\Omega(u)\nu_0)^{-1/2},
\end{aligned}
$$

and the conclusion follows by a standard result on the distribution of quadratic forms in normal vectors with idempotent matrices, see e.g. Theorem 7.2 of Rao (1973). For the case of the estimated instruments $\widehat{z}_{it}$ using (41), (42) or (45) it follows that

$$
\begin{aligned}
& \Gamma^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - \Gamma^{CR}\left(\widehat{\beta}, 0, u\right) = -\widehat{\lambda}(u)' \sum \widehat{z}_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u) + \\
& \frac{1}{2}\widehat{\lambda}(u)' \sum (\widehat{z}_{it}\overline{\varepsilon}_{it} K_h(u_{it}-u))^{\otimes 2} \widehat{\lambda}(u) \\
= \; & \frac{h^p}{n} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u))' (f(u)\Omega(u)\nu_0)^{-1} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u)) + \\
& O_p(1) \left(\frac{h^p}{n}\right)^{1/2} \sum (\widehat{z}_{it} - z_{it}) \widehat{\varepsilon}_{it} K_h(u_{it}-u) + \\
& \frac{1}{2}\widehat{\lambda}(u)' \frac{h^p}{n} \sum ((\widehat{z}_{it} - z_{it}) \widehat{\varepsilon}_{it} K_h(u_{it}-u))^{\otimes 2} \widehat{\lambda}(u) \\
= \; & \frac{h^p}{n} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u))' (f(u)\Omega(u)\nu_0)^{-1} \sum (z_{it}\widehat{\varepsilon}_{it} K_h(u_{it}-u)) + o_p(1)
\end{aligned} \tag{50}
$$

by (44) or (48), hence the result follows as in (49). ■

Proof of Theorem 5. By a mean value expansion about $\lambda = 0$

$$\begin{aligned} \frac{\partial \widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)}{\partial \lambda} &= \frac{1}{n} - \frac{1}{n}\left(1 + \gamma \overline{\lambda}\left(u\right)' z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma} - 1} \\ & \quad \widehat{\lambda}\left(u\right)'\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) \end{aligned}$$

where $\overline{\lambda} =: \overline{\lambda}\left(u\right)$ is the mean value. By (25) and (29) it follows that

$$\frac{\partial \widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)}{\partial \lambda} = \frac{1}{n} - \frac{1}{n}\widehat{\lambda}\left(u\right)'\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(\frac{1}{n}\right), \tag{51}$$

hence $n\widehat{\pi}_{it}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1 = \widehat{\lambda}\left(u\right)'\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(1\right)$ and thus

$$\begin{aligned} &\sum\left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2 = \widehat{\lambda}\left(u\right)' \sum\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\otimes 2} \widehat{\lambda}\left(u\right) + o_p\left(1\right) = \\ &\frac{h^p}{n}\sum\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)'\left(f\left(u\right)\Omega\left(u\right)\nu_0\right)^{-1}\sum\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(1\right), \end{aligned} \tag{52}$$

so that the result follows as in the proof of Theorem 4. The second result follows noting that

$$\sum \frac{\left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2}{n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)} = \sum\left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2\left(1 + o_p\left(1\right)\right) \tag{53}$$

as $\max_{i,t}\left|\widehat{\lambda}\left(u\right)'\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)\right| = o_p\left(1\right)$. For the case of estimated instruments $\widehat{z}_{it}$,

$$\frac{\partial \widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)}{\partial \lambda} = \frac{1}{n} - \frac{1}{n}\widehat{\lambda}\left(u\right)'\left(\widehat{z}_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(\frac{1}{n}\right),$$

and by (50)

$$\begin{aligned} \sum\left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2 &= \frac{h^p}{n}\sum\left(\widehat{z}_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)'\left(f\left(u\right)\Omega\left(u\right)\nu_0\right)^{-1} \\ & \quad \sum\left(\widehat{z}_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(1\right) \\ &= \frac{h^p}{n}\sum\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)'\left(f\left(u\right)\Omega\left(u\right)\nu_0\right)^{-1} \\ & \quad \sum\left(z_{it} \widehat{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right) + o_p\left(1\right) \end{aligned}$$

and the conclusion follows by the same arguments as those used in the proof of Theorem 4. The conclusion for $\sum\left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - 1\right)^2 / n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)$ follows by (52) and (53). ■

Proof of Corollary 5.1. Under the local Pitman alternative and $\left(nh^p\right)^{1/2}\gamma_n\left(u\right) \to \gamma\left(u\right) > 0$ a standard kernel calculation and the same arguments as those used in the proofs of Theorems

1 and 2 imply that $\sum \left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u\right)\right)/\left(nh^p\right)^{1/2} \xrightarrow{d} N\left(\gamma\left(u\right)f\left(u\right),\Omega_0\left(u\right)v_0 f\left(u\right)\right)$, hence as in (49) the result for $D^{CR}\left(u\right)$ follows by standard results on the distribution of quadratic forms in nonzero mean normal vectors with idempotent matrices, see e.g. Theorem 7.2 of Rao (1973). The consistency under the condition $\left(nh^p\right)^{1/2}\gamma_n\left(u\right)\to\infty$ is a direct consequence of the previous conclusion. The result for $P_j^{CR}\left(u\right)$ $(j=1,2)$ follows by (52) and (53), which imply that $P_j^{CR}\left(u\right)=D^{CR}\left(u\right)+o_p\left(1\right)$. ■

Proof of Theorem 6. It is first shown that for any two distinct u_j and u_k for $1\leq j,k\leq m$

$$\frac{h^p}{n}Cov\left(\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right),\sum\left(z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right)=o\left(1\right). \tag{54}$$

For T finite, iterated expectations and a standard kernel calculation show that

$$Cov\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),\left(z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right)=\Omega_{1t}\left(u_{i1},u_{is}\right)f\left(u_j,u_k\right),$$

hence by A1

$$\begin{aligned}\left\|\frac{h^p}{n}Cov\left(\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right),\sum\left(z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right)\right\| &= h^pTO\left(1\right) \\ &\to 0.\end{aligned} \tag{55}$$

For $T\to\infty$ let d_n be an integer such that $d_nh^p\to 0$; then by

$$\begin{aligned}&h^p\sum\left\|\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right),\sum\left(z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right\|= \\ &h^p\sum_{s=1}^{d_n}\left\|Cov\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right\|+ \\ &h^p\sum_{s=d_n+1}^{T}\left\|Cov\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right\|\leq d_nh^p+\left(h^{-p\frac{\gamma}{2+\gamma}}\right)\sum\alpha\left(s\right)^{\frac{\gamma}{2+\gamma}}\to 0\end{aligned} \tag{56}$$

by (55), $E\left\|z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right\|^{2+\gamma}=O\left(h^{-p(1+\gamma)}\right)$, A7 and an application of Davidov's inequality (Hall and Heyde 1980, p. 278) that shows that

$$\begin{aligned}\left\|Cov\left(\left(z_{i1}\varepsilon_{i1}K_h\left(u_{i1}-u_j\right),z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right)\right\| &\leq C\alpha\left(s\right)^{\frac{\gamma}{2+\gamma}}\left\|E\left(\left\|z_{i1}\varepsilon_{i1}K_h\left(u_{i1}-u_j\right)\right\|^{2+\gamma}\right)\right\|^{\frac{1}{2+\gamma}} \\ &\quad \left\|E\left(\left\|z_{is}\varepsilon_{is}K_h\left(u_{is}-u_j\right)\right\|^{2+\gamma}\right)\right\|^{\frac{1}{2+\gamma}}.\end{aligned}$$

Thus by (54), the same CLTs used in the proofs of Theorems 1 and 2 can be used to show that

$$\begin{aligned}&\left(\frac{h^p}{n}\right)^{1/2}\begin{bmatrix}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_1\right)\right)\\ \vdots\\ \sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_m\right)\right)\end{bmatrix}\xrightarrow{d} \\ &N\left(\begin{array}{c}0,diag\left[f\left(u_1\right)\Omega_0\left(u_1\right)\nu_0-P_0\left(u_1\right),...,\\ f\left(u_m\right)\Omega_0\left(u_m\right)\nu_0-P\left(u_m\right)\right]\end{array}\right),\end{aligned} \tag{57}$$

where $diag\left[\cdot\right]$ indicates a diagonal matrix and

$$P_0\left(u\right) = \Sigma_0\left(u\right)\left(\Xi_0\left(u\right)f\left(u\right)/\nu_0\right)^{-1}\Sigma_0\left(u\right)'. \tag{58}$$

The result for $\max_j D^{CR}\left(u_j\right)$ follows by (49), (57) and the CMT, which imply that

$$\max_j\left(\frac{h^p}{n}\right)\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_j\right)\right)'\left(f\left(u_j\right)\Omega\left(u_j\right)\nu_0\right)^{-1}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_j\right)\right)\xrightarrow{d}$$
$$\max_j\chi_j^2\left(l-k\right).$$

The result for $\max_j P_k^{CR}\left(u_j\right)$ $(k=1,2)$ follows similarly using (52) and (53). For the estimated instruments $\widehat{z}_{it}$ we have

$$\begin{aligned}
&Cov\left(\widehat{z}_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),\widehat{z}_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)=\\
&Cov\left(\left(\widehat{z}_{it}-z_{it}\right)\varepsilon_{it}K_h\left(u_{it}-u_j\right),\left(\widehat{z}_{is}-z_{is}\right)\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)+\\
&2Cov\left(\left(\widehat{z}_{it}-z_{it}\right)\varepsilon_{it}K_h\left(u_{it}-u_j\right),\widehat{z}_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)+\\
&Cov\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),z_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right),
\end{aligned}$$

and for the parametric case

$$\begin{aligned}
&\left\|Cov\left(\widehat{z}_{it}\varepsilon_{is}K_h\left(u_{it}-u_j\right),\widehat{z}_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right\|=\\
&\left\|\widehat{\beta}-\beta_0\right\|^2\left\|\Omega_{1t}^{\partial g}\left(u_{i1},u_{is}\right)f\left(u_j,u_k\right)\right\|=o_p\left(1\right),
\end{aligned}$$

and similarly for the second term, whereas for the nonparametric case, (46) and a standard kernel calculation shows that

$$\begin{aligned}
&\left\|Cov\left(\widehat{z}_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right),\widehat{z}_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right\|=\\
&\left\|\Omega_{1t}\left(u_{i1},u_{is}\right)f\left(u_j,u_k\right)\right\|+O\left(b^2\right),
\end{aligned}$$

and similarly for the second term; thus by either (55) or (56)

$$\left\|\frac{h^p}{n}Cov\left(\sum\left(\widehat{z}_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right),\sum\left(\widehat{z}_{is}\varepsilon_{is}K_h\left(u_{is}-u_k\right)\right)\right)\right\|\to 0,$$

and the result follows using (49), (52), (53), (57) and the CMT. ■

Proof of Corollary 6.1. The same arguments as those used in the proof of Corollary 5.1 and Theorem 6 show that under the local Pitman alternative and $\left(nh^p\right)^{1/2}\gamma_n\left(u_j\right)\to\gamma\left(u_j\right)>0$ for $j=1,...,m$

$$\sum\left(z_{it}\varepsilon_{it}K_h\left(u_{it}-u_j\right)\right)/\left(nh^p\right)^{1/2}\xrightarrow{d}N\left(\gamma\left(u_j\right)f\left(u_j\right),\Omega_0\left(u_j\right)v_0f\left(u_j\right)\right)$$

and for any two distinct u_j and u_k for $1\le j,k\le m$

$$\frac{h^p}{n}Cov\left(\sum\left[\left(z_{it}\varepsilon_{it}-\gamma_n\left(u_j\right)\right)K_h\left(u_{it}-u_j\right)\right],\sum\left[\left(z_{is}\varepsilon_{is}-\gamma_n\left(u_k\right)\right)K_h\left(u_{is}-u_k\right)\right]\right)=o\left(1\right),$$

and

$$\left(\frac{h^p}{n}\right)^{1/2}\left[\begin{array}{c}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_1\right)\right)\\ \vdots\\ \sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_m\right)\right)\end{array}\right]\overset{d}{\to}$$
$$N\left(\begin{array}{c}\left(I-P_0\left(u_1\right)\right)\gamma\left(u_1\right)f\left(u_1\right),...,\left(I-P_0\left(u_m\right)\right)\gamma\left(u_m\right)f\left(u_m\right),\\ diag\left[f\left(u_1\right)\Omega_0\left(u_1\right)\nu_0-P_0\left(u_1\right),...,\right.\\ \left.f\left(u_m\right)\Omega_0\left(u_m\right)\nu_0-P\left(u_m\right)\right],\end{array}\right),$$

where the matrix $P_0\left(\cdot\right)$ is defined in (58). Then the same result on the distribution of quadratic forms in nonzero mean normal vectors used in Corollary 5.1 and the CMT imply that

$$\max_j\left(\frac{h^p}{n}\right)\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_j\right)\right)'\left(f\left(u_j\right)\Omega\left(u_j\right)\nu_0\right)^{-1}\sum\left(z_{it}\widehat{\varepsilon}_{it}K_h\left(u_{it}-u_j\right)\right)\overset{d}{\to}$$
$$\max_j\chi_j^2\left(\kappa_j,l-k\right),$$

where $\kappa_j=f\left(u_j\right)\gamma\left(u_j\right)'\left(\Omega_0\left(u_j\right)^{-1}\left(I-\Sigma_0\left(u_j\right)\Xi_0\left(u_j\right)^{-1}\Sigma_0\left(u_j\right)\right)\Omega_0\left(u_j\right)^{-1}\right)\gamma\left(u_j\right)/v_0$; the result $\max_j D^{CR}\left(u_j\right)$ follows by (49) and (57), whereas that for $\max_j P_k^{CR}\left(u_j\right)$ $(k=1,2)$ follows by (52), (53) which imply that $\max_j P_k^{CR}\left(u_j\right)=\max_j D^{CR}\left(u_j\right)+o_p\left(1\right)$. The consistency under the condition $\left(nh^p\right)^{1/2}\gamma_n\left(u_j\right)\to\infty$ is a direct consequence of the previous conclusion. ■

Proof of Theorem 7. A second-order Taylor expansion about $\widehat{\beta}\left(u\right)$ with Lagrange reminder $\overline{\beta}:=\overline{\beta}\left(u\right)$ shows that

$$D_{(p)}^{CR}\left(u\right)=\left(\widehat{\beta}\left(u\right)-\widetilde{\beta}\left(u\right)\right)'\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\widetilde{\lambda}\left(\beta^{(p)}\right)\right)}{\partial\beta^{\otimes2}}\left(\widehat{\beta}\left(u\right)-\widetilde{\beta}\left(u\right)\right)+o_p\left(1\right)$$

as $\partial\Gamma^{CR}\left(\widehat{\beta},\widetilde{\lambda}\right)/\,\partial\beta=o_p\left(1\right)$ by definition, where the notation $\widetilde{\lambda}\left(\beta^{(p)}\right)$ emphasizes the dependence of $\widetilde{\lambda}$ on the constraint. Then by the chain rule

$$\frac{\partial^2\Gamma^{CR}\left(\overline{\beta},\widetilde{\lambda}\left(\beta^{(p)}\right)\right)}{\partial\beta^{\otimes2}}=-\frac{1}{n}\sum\left(1-\gamma\widetilde{\lambda}\left(\beta^{(p)},u\right)'z_{it}\widetilde{\varepsilon}_{it}K_h\left(u_{it}-u\right)\right)^{\frac{1}{\gamma}-1}\tag{59}$$
$$\left[x_{it}z_{it}'\widetilde{\lambda}\left(\beta^{(p)}\right)^{\otimes2}z_{it}'x_{it}K_h\left(u_{it}-u\right)^2+\right.$$
$$\left.x_{it}z_{it}'\widetilde{\lambda}\left(\beta^{(p)}\right)z_{it}\widetilde{\varepsilon}_{it}'\frac{\partial\widetilde{\lambda}\left(\beta^{(p)}\right)}{\partial\beta'}z_{it}\widetilde{\varepsilon}_{it}K_h\left(u_{it}-u\right)^2-x_{it}z_{it}'\frac{\partial\widetilde{\lambda}\left(\beta^{(p)}\right)}{\partial\beta'}K_h\left(u_{it}-u\right)\right]$$

and using the same arguments as those used in the proof of Theorem 1 or 2

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} x_{it} z_{it}' \widetilde{\lambda}\left(\beta^{(p)}\right)^{\otimes 2} z_{it}' x_{it} K_h\left(u_{it} - u\right)^2 \right\| = o_p(1),$$

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} x_{it} z_{it}' \widetilde{\lambda}\left(\beta^{(p)}\right) \times \right.$$
$$\left. z_{it} \widetilde{\varepsilon}_{it}' \frac{\partial \widetilde{\lambda}\left(\beta^{(p)}\right)}{\partial \beta'} z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)^2 \right\| = o_p(1),$$

whereas

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} x_{it} z_{it}' \frac{\partial \widetilde{\lambda}\left(\beta^{(p)}\right)}{\partial \beta'} K_h\left(u_{it} - u\right) \right. \tag{60}$$
$$\left. - \Sigma_0(u) f(u) \frac{\partial \widetilde{\lambda}\left(\beta^{(p)}\right)}{\partial \beta'} \right\| = o_p(1).$$

Furthermore by differentiating with respect to β the first order condition

$$0 = -\frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}} z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)$$

we have

$$\begin{aligned} 0 &= -\partial \left(\frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}} z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right) \right) / \partial \beta = \\ & -\frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} \left(z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\otimes 2} \frac{\partial \widetilde{\lambda}\left(\beta^{(p)}, u\right)}{\partial \beta'} - \\ & \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} \left(z_{it} x_{it}' K_h\left(u_{it} - u\right)\right) - \\ & \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda}\left(\beta^{(p)}, u\right)' z_{it} \widetilde{\varepsilon}_{it} K_h\left(u_{it} - u\right)\right)^{\frac{1}{\gamma}-1} z_{it}^{\otimes 2} \widetilde{\lambda}\left(\beta^{(p)}, u\right) x_{it}' K_h\left(u_{it} - u\right), \end{aligned}$$

which yields

$$\frac{\partial \widetilde{\lambda}\left(\beta^{(p)}, u\right)}{\partial \beta'} = -\left(\Omega_0(u) \nu_0\right)^{-1} \Sigma_0(u) + o_p(1). \tag{61}$$

Thus using (60), (61) and the triangle inequality yield

$$\left\| \frac{\partial^2 \Gamma^{CR}\left(\overline{\beta}, \widetilde{\lambda}\left(\beta^{(p)}\right)\right)}{\partial \beta^{\otimes 2}} - f(u) \Sigma_0(u)' \left(\Omega_0(u) \nu_0\right)^{-1} \Sigma_0(u) \right\| = o_p(1),$$

so that

$$D_{(p)}^{CR}(u) = \left(\widehat{\beta}(u) - \widetilde{\beta}(u)\right)' f(u)\,\Sigma_0(u)'\left(\Omega_0(u)\,\nu_0\right)^{-1}\Sigma_0(u) \left(\widehat{\beta}(u) - \widetilde{\beta}(u)\right) + o_p(1). \tag{62}$$

The proof of Theorem 1 or 2 shows that

$$(nh^p)^{1/2}\left(\widehat{\beta}(u) - \beta_0(u)\right) = \Xi_0(u)^{-1}\Sigma_0(u)' \times (f(u)\,\Omega(u))^{-1}\left(\frac{h^p}{n}\right)^{1/2}\sum z_{it}\varepsilon_{it}K_h(u_{it}-u) + o_p(1)$$

and by a Lagrange multiplier argument it is easy to see that

$$(nh^p)^{1/2}\left(\widetilde{\beta}(u) - \beta_0(u)\right) = \left[I_k - \Xi_0(u)^{-1}R'\left(R\Xi_0(u)^{-1}R'\right)^{-1}R\right]\Xi_0(u)^{-1}\Sigma_0(u)' \times (f(u)\,\Omega_0(u))^{-1}\left(\frac{h^p}{n}\right)^{1/2}\sum z_{it}\varepsilon_{it}K_h(u_{it}-u) + o_p(1),$$

where $R = \left[I_p, O_{p\times(k-p)}\right]$ and $O_{p\times(k-p)}$ is a $p\times(k-p)$ matrix of zeros. Then some algebra shows that

$$D_{(p)}^{CR}(u) = \frac{h^p}{n}\sum\left(z_{it}\varepsilon_{it}K_h(u_{it}-u)\right)'\Lambda_0(u)' f(u)\,\Sigma_0(u)'\left(\Omega_0(u)\,\nu_0\right)^{-1} \times \Sigma_0(u)\,\Lambda_0(u)\sum z_{it}\varepsilon_{it}K_h(u_{it}-u) + o_p(1), \tag{63}$$

where

$$\Lambda_0(u) = \Xi_0(u)^{-1}R'\left(R\Xi_0(u)^{-1}R'\right)^{-1}R\Xi_0(u)^{-1}\Sigma_0(u)' f(u)\,\Omega_0(u)^{-1}.$$

By CMT and the same arguments used in the proofs of Theorems 1 and 2

$$\Lambda_0(u)\left(\frac{h^p}{n}\right)^{1/2}\sum z_{it}\varepsilon_{it}K_h(u_{it}-u) \xrightarrow{d} N\left(0, \frac{v_0}{f(u)}\Xi_0(u)^{-1}R'\left(R\Xi_0(u)^{-1}R'\right)^{-1}R\Xi_0(u)^{-1}\right), \tag{64}$$

and the result follows as in the proof of Theorem 4 by CMT, noting that

$$\Xi_0(u)^{-1/2}R'\left(R\Xi_0(u)^{-1}R'\right)^{-1}R\Xi_0(u)^{-1/2}$$

is symmetric and idempotent with $rank\left(\Xi_0(u)^{-1/2}R'\left(R\Xi_0(u)^{-1}R'\right)^{-1}R\Xi_0(u)^{-1/2}\right) = k$. For the estimated instruments case $\widehat{z}_{it}$, note that by triangle inequality and the same arguments as those used in the proof of Theorem 3

$$\left\|\frac{1}{n}\sum\left(1-\gamma\widetilde{\lambda}\left(\beta^{(p)},u\right)' z_{it}\widetilde{\varepsilon}_{it}K_h(u_{it}-u)\right)^{\frac{1}{\gamma}-1} x_{it}\widehat{z}_{it}'\widetilde{\lambda}\left(\beta^{(p)}\right)^{\otimes 2}\widehat{z}_{it}'x_{it}K_h(u_{it}-u)^2\right\| \leq$$

$$\left\|\widehat{z}_{it}-z_{it}\right\|^2\left\|\widetilde{\lambda}\left(\beta^{(p)}\right)\right\|^2\left\|\frac{1}{n}\sum\left(1-\gamma\widetilde{\lambda}\left(\beta^{(p)},u\right)' z_{it}\widetilde{\varepsilon}_{it}K_h(u_{it}-u)\right)^{\frac{1}{\gamma}-1} x_{it}^{\otimes 2}K_h(u_{it}-u)^2\right\| +$$

$$\left\|\frac{1}{n}\sum\left(1-\gamma\widetilde{\lambda}\left(\beta^{(p)},u\right)' z_{it}\widetilde{\varepsilon}_{it}K_h(u_{it}-u)\right)^{\frac{1}{\gamma}-1} x_{it}z_{it}'\widetilde{\lambda}\left(\beta^{(p)}\right)^{\otimes 2} z_{it}'x_{it}K_h(u_{it}-u)^2\right\| = o_p(1),$$

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda} \left(\beta^{(p)}, u \right)' \widehat{z}_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)^{\frac{1}{\gamma} - 1} x_{it} \widehat{z}_{it}' \widetilde{\lambda} \left(\beta^{(p)} \right) \times \right.$$

$$\left. \widehat{z}_{it} \widetilde{\varepsilon}_{it}' \frac{\partial \widetilde{\lambda} \left(\beta^{(p)} \right)}{\partial \beta'} z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right)^2 \right\| \leq \left\| \widehat{z}_{it} - z_{it} \right\|^2 \times$$

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda} \left(\beta^{(p)}, u \right)' \widehat{z}_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)^{\frac{1}{\gamma} - 1} x_{it} \widetilde{\lambda} \left(\beta^{(p)} \right) \times \right.$$

$$\left. \widetilde{\varepsilon}_{it}' \frac{\partial \widetilde{\lambda} \left(\beta^{(p)} \right)}{\partial \beta'} z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right)^2 \right\| + \left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda} \left(\beta^{(p)}, u \right)' z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)^{\frac{1}{\gamma} - 1} \times \right.$$

$$\left. x_{it} z_{it}' \widetilde{\lambda} \left(\beta^{(p)} \right) z_{it} \widetilde{\varepsilon}_{it}' \frac{\partial \widetilde{\lambda} \left(\beta^{(p)} \right)}{\partial \beta'} z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right)^2 \right\| = o_p(1),$$

and similarly for

$$\left\| \frac{1}{n} \sum \left(1 - \gamma \widetilde{\lambda} \left(\beta^{(p)}, u \right)' \widehat{z}_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)^{\frac{1}{\gamma} - 1} x_{it} z_{it}' \frac{\partial \widetilde{\lambda} \left(\beta^{(p)} \right)}{\partial \beta'} K_h \left(u_{it} - u \right) \right.$$

$$\left. - \Sigma_0(u) f(u) \frac{\partial \widetilde{\lambda} \left(\beta^{(p)} \right)}{\partial \beta'} \right\| = o_p(1).$$

Similarly to (63) we have that

$$\begin{aligned} D_{(p)}^{CR}(u) &= \frac{h^p}{n} \sum \left(\widehat{z}_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) \right)' \Lambda_0(u)' f(u) \Sigma_0(u)' \left(\Omega_0(u) \nu_0 \right)^{-1} \times \\ &\quad \Sigma_0(u) \Lambda_0(u) \sum \widehat{z}_{it} \varepsilon_{it} K_h \left(u_{it} - u \right) + o_p(1), \end{aligned} \tag{65}$$

and the result follows as in the proof of Theorem 4. ■

Proof of Theorem 8. By the same arguments as those used in the proof of Theorem 5 it follows that

$$n \widetilde{\pi}_{it} \left(\widetilde{\beta}, \widetilde{\lambda}, u \right) - 1 = \widetilde{\lambda}(u)' \left(z_{it} \widetilde{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right) + o_p(1)$$

and thus

$$\begin{aligned} \sum \left(n \widehat{\pi}_{it}^{CR} \left(\widehat{\beta}, \widehat{\lambda}, u \right) - n \widetilde{\pi}_{it}^{CR} \left(\widetilde{\beta}, \widetilde{\lambda}, u \right) \right)^2 &= \left(\widehat{\lambda}(u) - \widetilde{\lambda}(u) \right)' \sum \left(z_{it} \widehat{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)^{\otimes 2} \times \\ &\quad \left(\widehat{\lambda}(u) - \widetilde{\lambda}(u) \right) + o_p(1) \\ &= \frac{h^p}{n} \sum \left(z_{it} \widehat{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right)' \left(f(u) \Omega(u) \nu_0 \right)^{-1} \times \\ &\quad \sum \left(z_{it} \widehat{\varepsilon}_{it} K_h \left(u_{it} - u \right) \right), \end{aligned}$$

where the second equality follows by (21) with $\widehat{\varepsilon}_{it}$ replacing $\widetilde{\varepsilon}_{it}$. By (33)

$$\begin{aligned}\widehat{\lambda}(u) &= (\Omega_0(u) f(u) \nu_0)^{-1} - (\Omega_0(u) f(u) \nu_0)^{-1} P_0(u) (\Omega_0(u) f(u) \nu_0)^{-1} \times \\ &\quad \frac{1}{n} \sum (z_{it}\varepsilon_{it} K_h(u_{it} - u)) + o_p(1),\end{aligned}$$

while some algebra shows that for $\widetilde{\lambda}(u)$

$$\begin{aligned}\widetilde{\lambda}(u) &= (\Omega_0(u) f(u) \nu_0)^{-1} - (\Omega_0(u) f(u) \nu_0)^{-1} \Sigma_0(u) K_0(u) \Sigma_0(u)' \Omega_0(u)^{-1} \times \\ &\quad \frac{1}{n} \sum (z_{it}\varepsilon_{it} K_h(u_{it} - u)) + o_p(1),\end{aligned}$$

where

$$K_0(u) = \frac{\nu_0}{f(u)} \Xi_0(u)^{-1} (I - \Lambda_0(u)),$$

hence

$$\sum \left(n\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right) - n\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right)\right)^2 = D_{(p)}^{CR}(u) + o_p(1), \tag{66}$$

whereas as in (53)

$$\begin{aligned}P_4^{CR}(u) &= \sum_{i=1}^{N} \sum_{t=1}^{N} \frac{\left(NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right) - NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)\right)^2}{NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)} \quad \text{or} \qquad (67) \\ &= \sum_{i=1}^{N} \sum_{t=1}^{N} \frac{\left(NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right) - NT\widehat{\pi}_{it}^{CR}\left(\widehat{\beta}, \widehat{\lambda}, u\right)\right)^2}{NT\widetilde{\pi}_{it}^{CR}\left(\widetilde{\beta}, \widetilde{\lambda}, u\right)} = D_{(p)}^{CR}(u) + o_p(1).\end{aligned}$$

The result follows using the same arguments as those used in the proof of Theorem 5. For the case of estimated instruments $\widehat{z}_{it}$, the result follows using the same arguments as those used in the proof of Theorem 7. ■

Proof of Theorem 9 . By the same arguments as those used in the proof of Theorems 6 and 7, (64) and (65) it follows that

$$nh^p Cov\left(\widehat{\beta}(u_j) - \widetilde{\beta}(u_j), \widehat{\beta}(u_k) - \widetilde{\beta}(u_k)\right) = o(1)$$

and

$$\begin{aligned}&(nh^p)^{1/2} \begin{bmatrix} \widehat{\beta}(u_1) - \widetilde{\beta}(u_1) \\ \vdots \\ \widehat{\beta}(u_m) - \widetilde{\beta}(u_m) \end{bmatrix} \xrightarrow{d} \\ &N\left(\begin{array}{c} 0, diag\left[\frac{v_0}{f(u_1)} \Xi_0(u_1)^{-1} R' \left(R\Xi_0(u_1)^{-1} R'\right)^{-1} R\Xi_0(u_1)^{-1}, \ldots \right. \\ \left. \frac{v_0}{f(u_m)} \Xi_0(u_m)^{-1} R' \left(R\Xi_0(u_m)^{-1} R'\right)^{-1} R\Xi_0(u_m)^{-1}\right] \end{array} \right),\end{aligned}$$

hence by (54), (62) and CMT we have that

$$\max_j D_{(p)}^{CR}(u_j) \xrightarrow{d} \max_j \chi_j^2(p).$$

The result for $\max_j P_k^{CR}(u_j)$ $(k = 3, 4)$ follows similarly using (66) and (67). ■

Proof of Corollary 9.1. Under the local Pitman alternative and $(nh^p)^{1/2}\gamma_n^{(p)}(u) \to \gamma^{(p)}(u) > 0$, the same Lagrange multiplier argument used in the proof of Theorem 7 shows that

$$(nh^p)^{1/2}\left(\widehat{\beta}(u) - \widetilde{\beta}(u)\right) \xrightarrow{d} N\left(\gamma_\Xi^p(u), \frac{v_0}{f(u)}\Xi_0(u)^{-1} R'\left(R\Xi_0(u)^{-1} R'\right)^{-1} R\Xi_0(u)^{-1}\right), \quad (68)$$

where

$$\gamma_\Xi^{(p)}(u) = \Xi(u)^{-1} R'\left[I_p, \Xi_0^{(pp)}(u)\,\Xi_0^{(pk-p)}(u)\right]'\left(\Xi_0^{(pp)}(u)\right)^{-1}\gamma^{(p)}(u)\, f(u),$$

and $\Xi_0^{(pp)}(u)$ and $\Xi_0^{(pk-p)}(u)$ are, respectively, the upper $p \times p$ and lower $p \times (k-p)$ left blocks of $\Xi_0(u)^{-1}$. Then by (63) the result follows by CMT noting that

$$\gamma_\Xi^{(p)}(u)'\,\Xi(u)\,\gamma_\Xi^{(p)}(u)\, f(u)/v_0 = \gamma_\Xi^{(p)}(u)'\left(\Xi_0^{(pp)}(u)\right)^{-1}\gamma_\Xi^{(p)}(u)\, f(u)/v_0.$$

The consistency under the condition $(nh^p)^{1/2}\gamma_n^{(p)}(u) \to \infty$ follows immediately as that for $P_k^{CR}(u)$ $(k = 3, 4)$ using (66) and (67). The result for the case $\{u_j\}_{j=1}^m$ follows using (68) and the same arguments used in the proof of Theorem 6. The consistency under the condition $(nh^p)^{1/2}\gamma_n(u_j) \to \infty$ follows similarly. ■

6 Tables and figures

Table 1. MSE $\times 10^{-3}$ of $\widehat{\beta}_j(\cdot)$ with observed instruments

	$\sigma_\varepsilon^2 = 0.5, \sigma_\eta^2 = 0.5$		$\sigma_\varepsilon^2 = 0.2, \sigma_\eta^2 = 0.8$	
$N = 100$ $T = 5$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	1.204	2.131	0.635	0.776
ET	1.205	2.145	0.645	0.782
GMM	1.312	2.290	0.765	0.866
$N = 100$ $T = 50$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.924	1.843	0.512	0.623
ET	0.926	1.852	0.514	0.632
GMM	1.152	1.892	0.648	0.757
$N = 400$ $T = 5$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.324	0.504	0.124	0.154
ET	0.376	0.512	0.131	0.169
GMM	0.521	0.623	0.167	0.202
$N = 400$ $T = 50$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.274	0.399	0.104	0.120
ET	0.294	0.405	0.121	0.127
GMM	0.434	0.543	0.127	0.142

Table 2. MSE $\times 10^{-3}$ of $\widehat{\beta}_j(\cdot)$ with estimated instruments

	$\sigma_\varepsilon^2 = 0.5, \sigma_\eta^2 = 0.5$		$\sigma_\varepsilon^2 = 0.2, \sigma_\eta^2 = 0.8$	
$N = 100$ $T = 5$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	1.099	1.877	0.543	0.687
ET	1.105	1.763	0.588	0.703
GMM	1.221	1.998	0.623	0.832
$N = 100$ $T = 50$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.910	1.675	0.490	0.547
ET	0.915	1.594	0.502	0.563
GMM	1.054	1.766	0.572	0.641
$N = 400$ $T = 5$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.243	0.376	0.106	0.254
ET	0.287	0.399	0.132	0.221
GMM	0.432	0.576	0.209	0.297
$N = 400$ $T = 50$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
EL	0.193	0.324	0.097	0.190
ET	0.185	0.342	0.105	0.197
GMM	0.365	0.502	0.164	0.212

Table 3. Finite sample sizes for $D_{(p)}^{EL}(u)$, $D_{(p)}^{ET}(u)$, $D_{(p)}^{GMM}(u)$, $P_3^{EL}(u)$, $P_3^{ET}(u)$

	$u = 2.5$		$u = 3.5$	
$N = 100, T = 5$				
$D_{(p)}^{EL}(u)$	0.015^a	0.054^b	0.014^a	0.057^b
$D_{(p)}^{ET}(u)$	0.016^a	0.056^b	0.015^a	0.055^b
$D_{(p)}^{GMM}(u)$	$0.026^a\dagger$	$0.064^b\dagger$	$0.023^a\dagger$	$0.061^b\dagger$
$P_3^{EL}(u)$	0.013^a	0.053^b	0.012^a	0.054^b
$P_3^{ET}(u)$	0.012^a	0.054^b	0.024^a	0.055^b
$N = 100, T = 50$				
$D_{(p)}^{EL}(u)$	0.013^a	$0.057^b\dagger$	0.015^a	0.055^b
$D_{(p)}^{ET}(u)$	0.015^a	0.055^b	0.014^a	$0.057^b\dagger$
$D_{(p)}^{GMM}(u)$	$0.028^a\dagger$	0.059^b	0.030^a	$0.058^b\dagger$
$P_3^{EL}(u)$	0.023^a	0.055^b	0.023^a	0.053^b
$P_3^{ET}(u)$	0.021^a	0.053^b	0.026^a	0.054^b

a 0.01 nominal level, *b* 0.05 nominal level, † statistically different from nominal level

Table 4. Finite sample sizes for $\max_j D_{(p)}^{EL}(u_j)$, $\max_j D_{(p)}^{ET}(u_j)$, $\max_j D_{(p)}^{GMM}(u_j)$, $\max_j P_3^{EL}(u_j)$, $\max_j P_3^{ET}(u_j)$

	$N = 100, T = 5$		$N = 100, T = 50$	
$\max_j D_{(p)}^{EL}(u_j)$	$0.017^a\dagger$	$0.056^b\dagger$	0.015^a	0.054^b
$\max_j D_{(p)}^{ET}(u_j)$	$0.018^a\dagger$	$0.059^b\dagger$	$0.016^a\dagger$	$0.057^b\dagger$
$\max_j D_{(p)}^{GMM}(u_j)$	$0.026^a\dagger$	$0.063^b\dagger$	$0.025^a\dagger$	$0.060^b\dagger$
$\max_j P_3^{EL}(u_j)$	0.015^a	0.048^b	0.014^a	0.047^b
$\max_j P_3^{ET}(u_j)$	0.015^a	0.053^b	0.015^a	0.052^b

a 0.01 nominal level, *b* 0.05 nominal level, † statistically different from nominal level

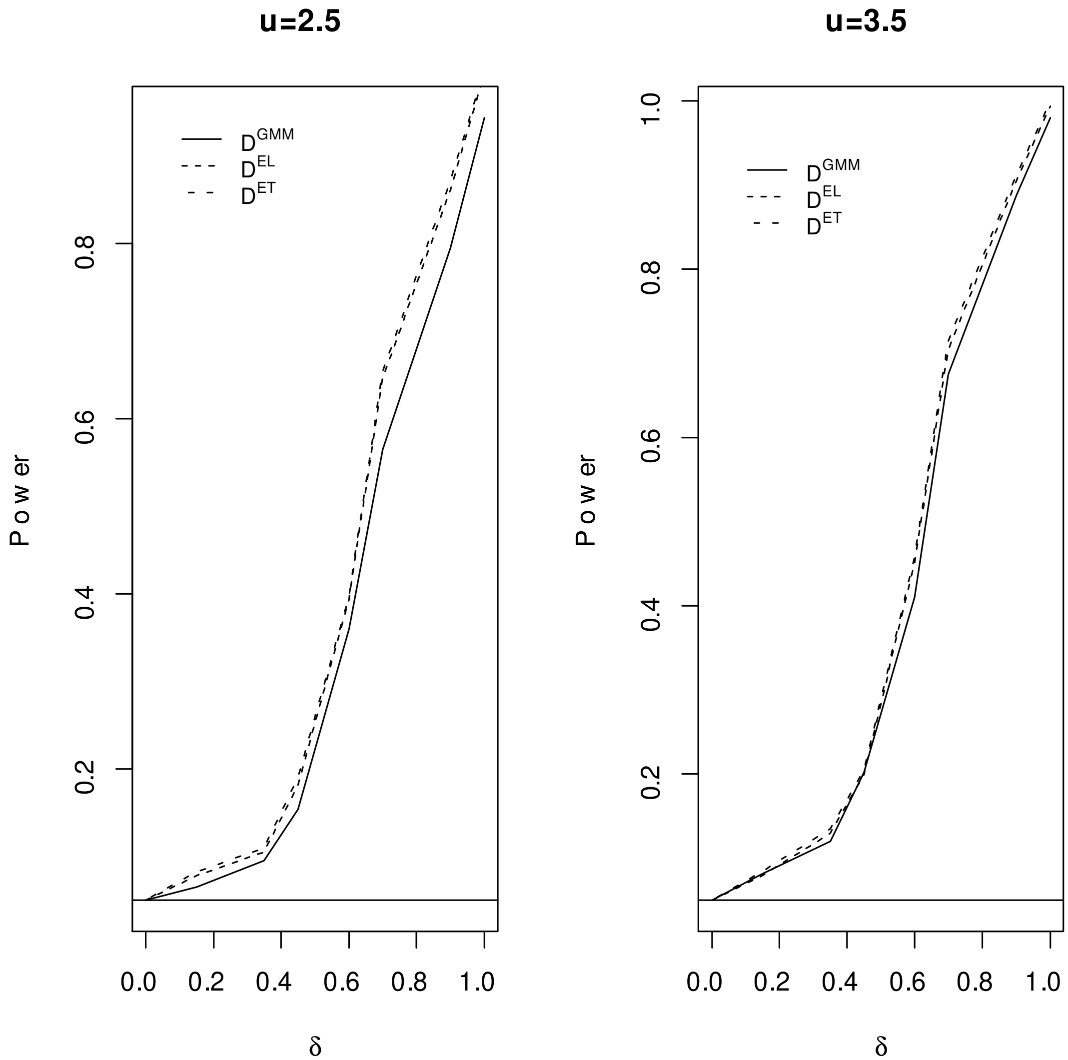

Figure 1. Finte sample power for $D^{EL}_{(p)}(u)$, $D^{ET}_{(p)}(u)$ and $D^{GMM}_{(p)}(u)$ for $N = 100$, $T = 5$.

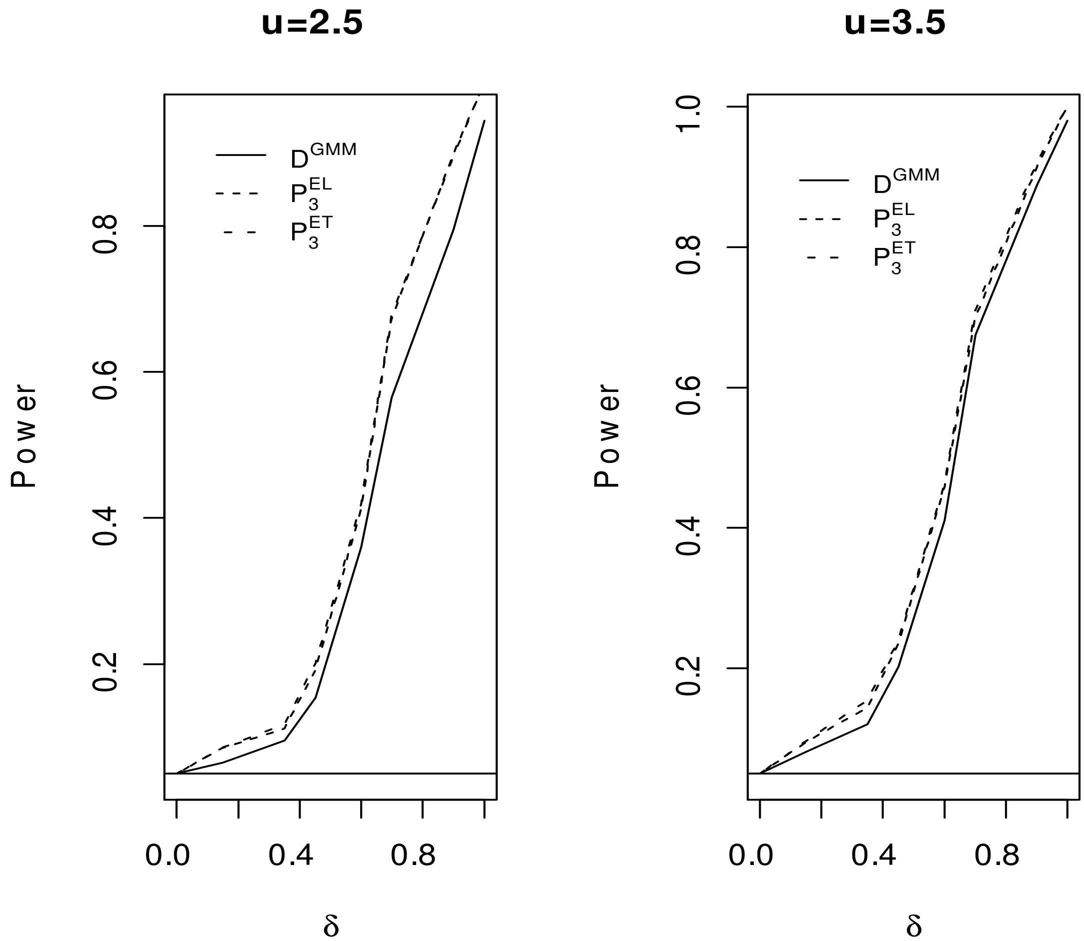

Figure 2. Finite sample power for $P_3^{EL}(u)$, $P_3^{ET}(u)$ and $D_{(p)}^{GMM}(u)$ for $N = 100, T = 5$.

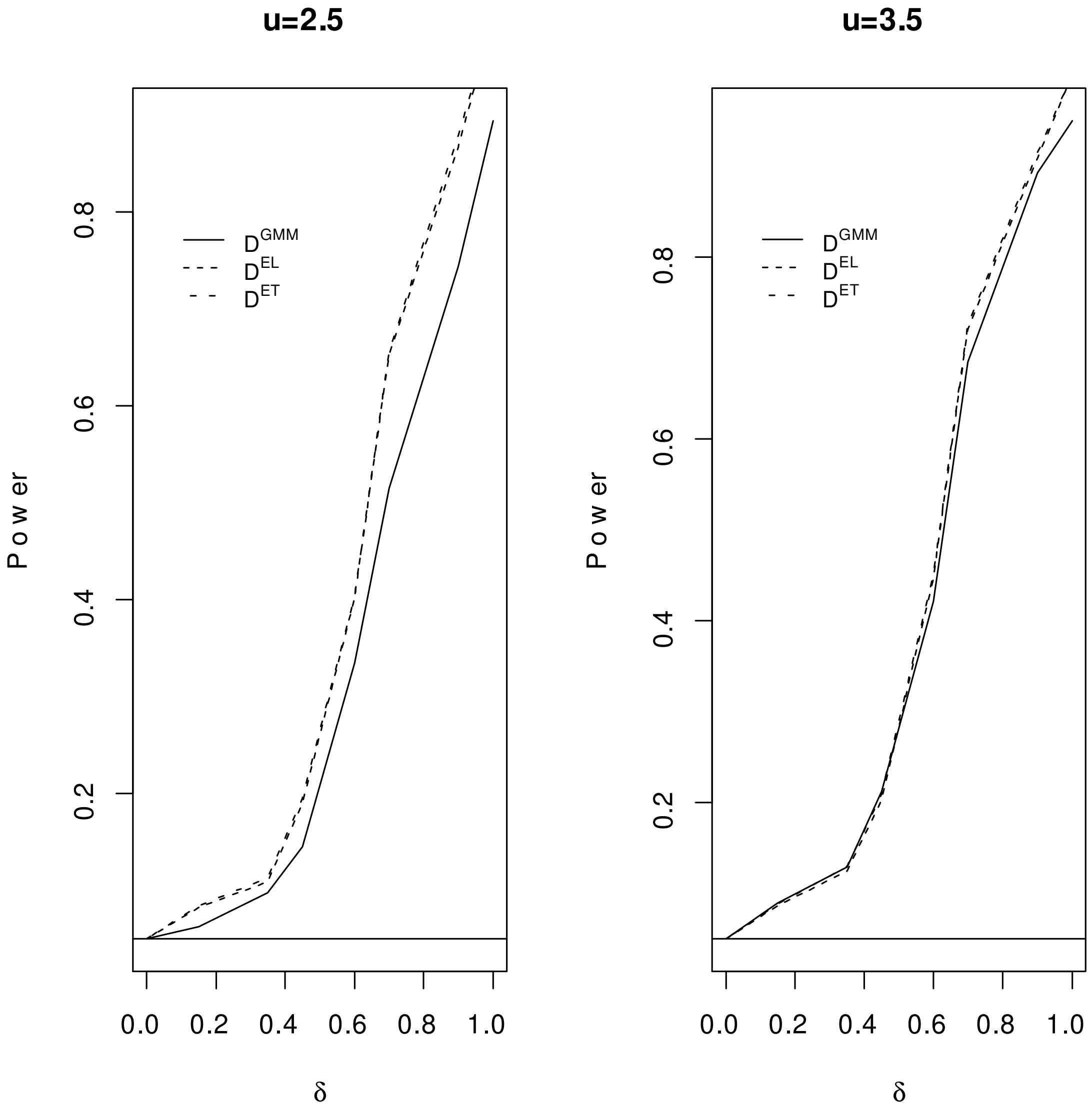

Figure 3. Finte sample power for $D_{(p)}^{EL}(u)$, $D_{(p)}^{ET}(u)$ and $D_{(p)}^{GMM}(u)$ for $N = 100$, $T = 50$.

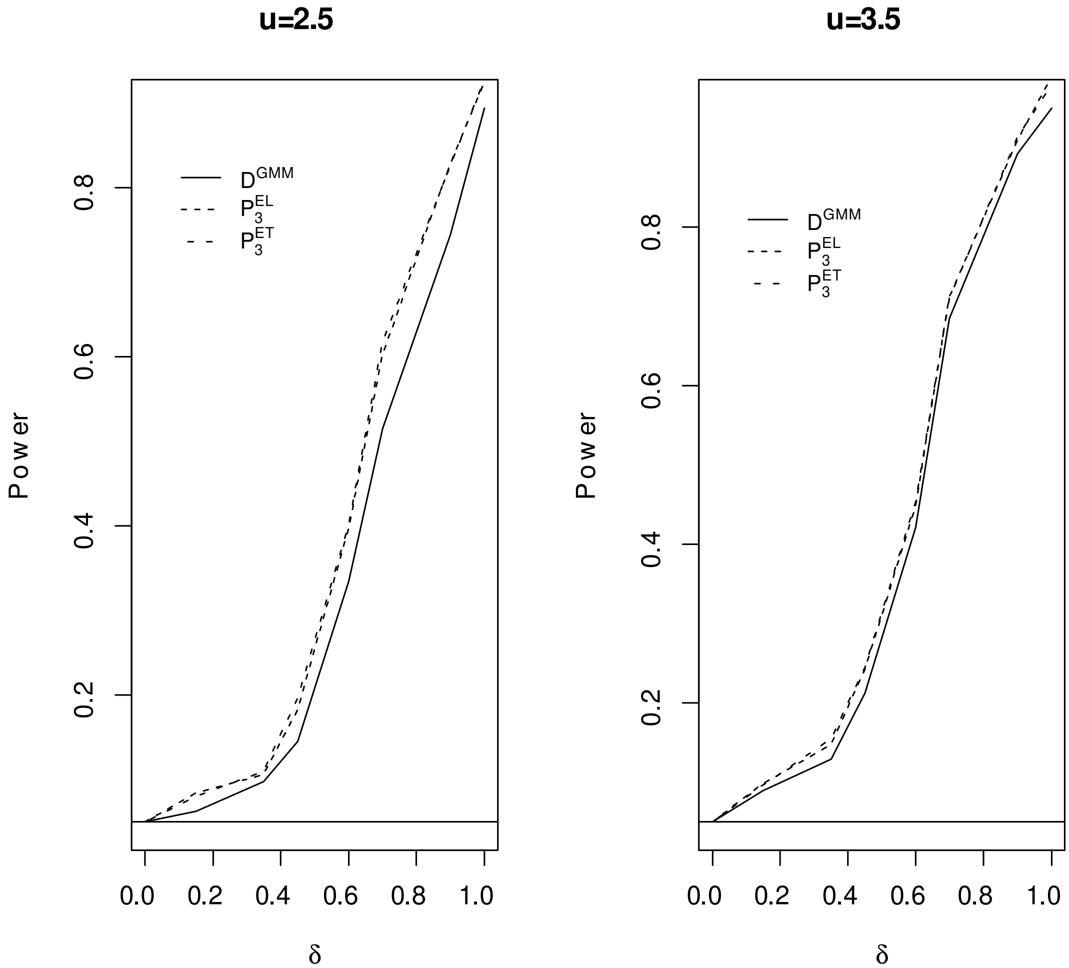

Figure 4. Finte sample power for $P_3^{EL}(u)$, $P_3^{ET}(u)$ and $D_{(p)}^{GMM}(u)$ for $N = 100$, $T = 50$.

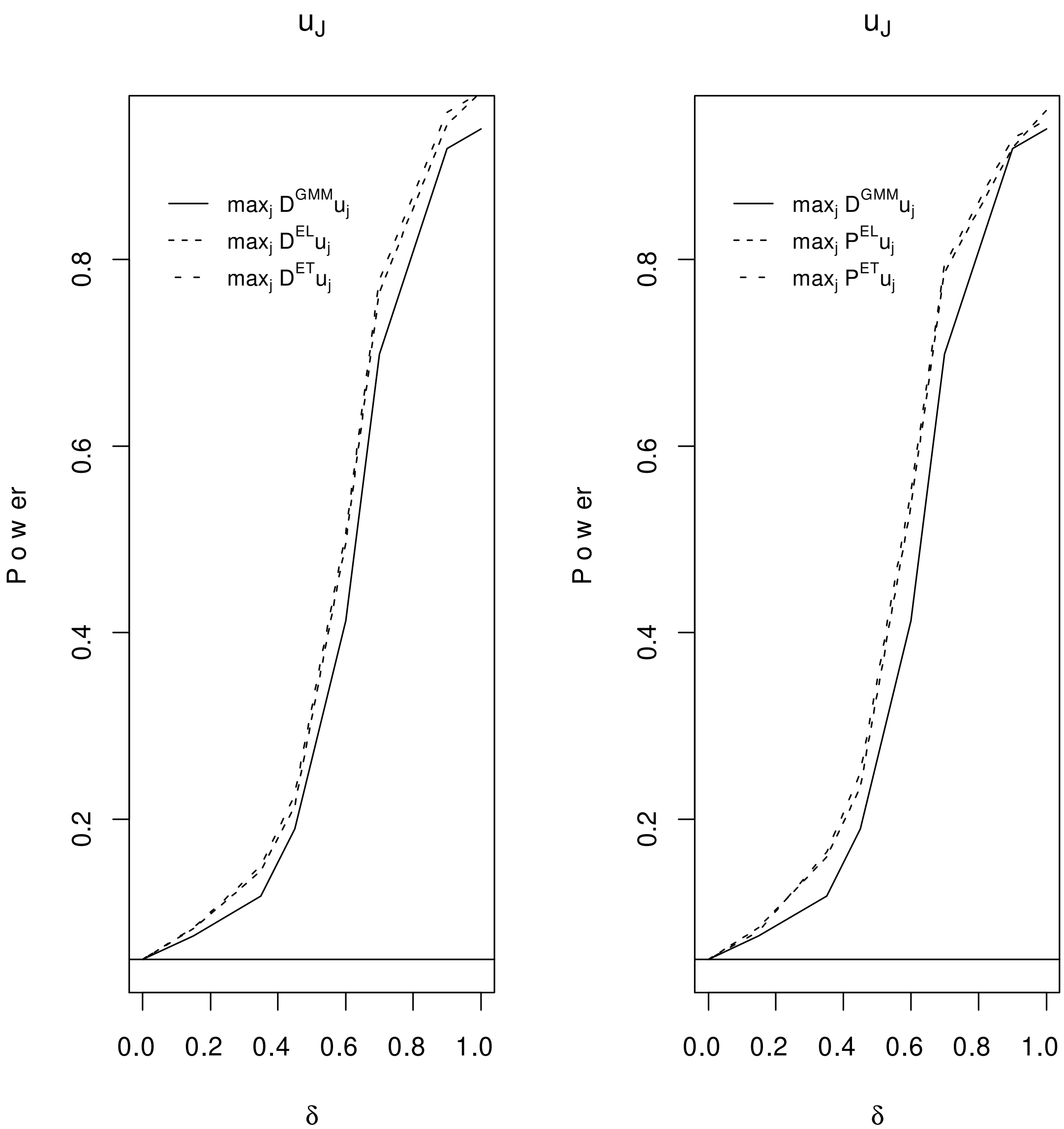

Figure 5. Finte sample power for $\max_j\, P_3^{EL}\,(u_j)$, $\max_j P_3^{ET}\,(u_j)$ and $\max_j D_{(p)}^{GMM}\,(u_j)$ for $N = 100$, $T = 5$.

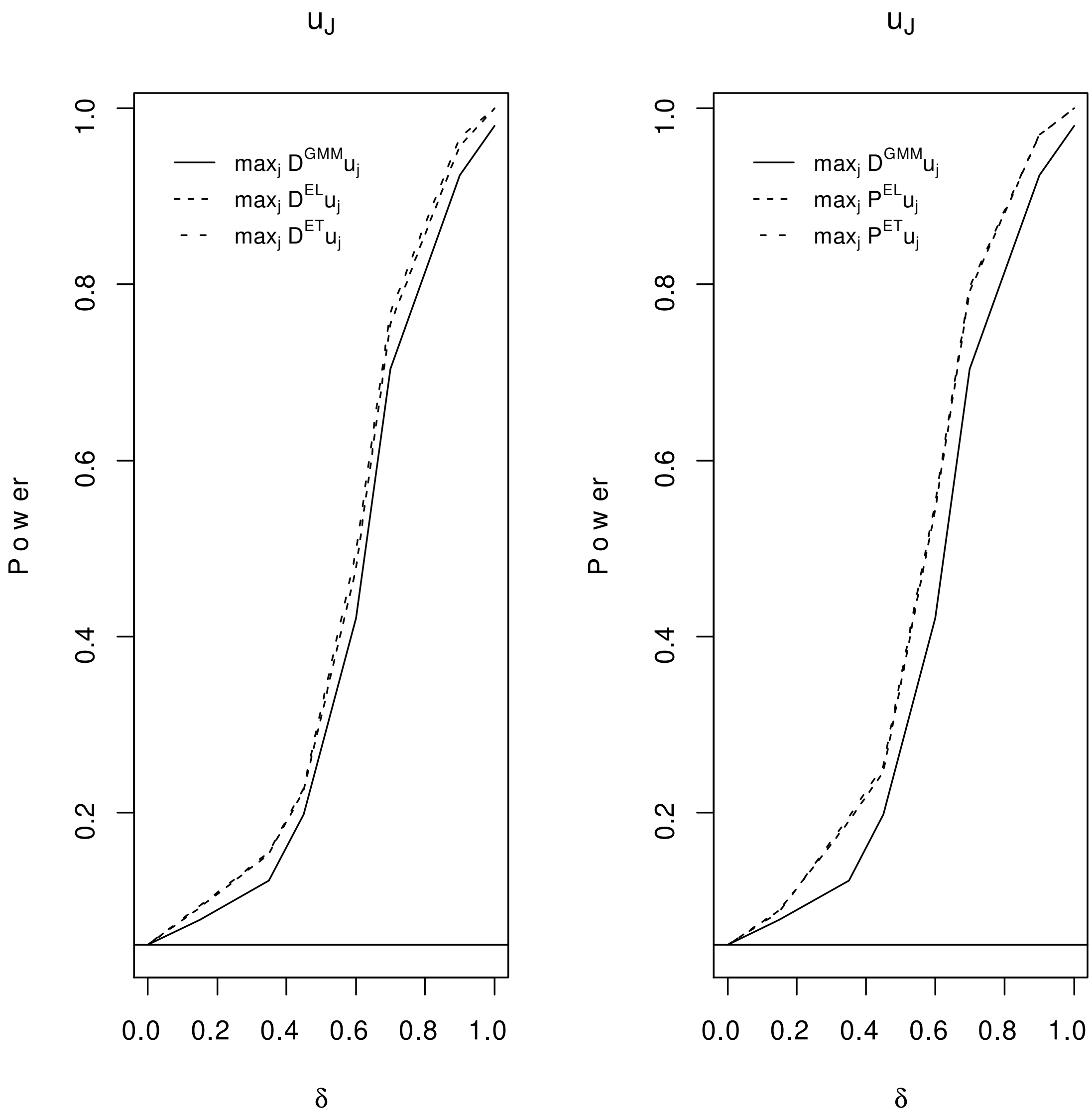

Figure 6. Finte sample power for $\max_j\ P_3^{EL}\left(u_j\right)$, $\max_j P_3^{ET}\left(u_j\right)$ and $\max_j D_{(p)}^{GMM}\left(u_j\right)$ for $N = 100$, $T = 50$.

References

Baggerly, K. A.: 1998, Empirical likelihood as a goodness of fit measure, *Biometrika* **85**, 535–547.

Baltagi, B.: 2010, *Econometric analysis of panel data*, John Wiley and Sons, London.

Baltagi, B. and Li, Q.: 2002, On instrumental variable estimation of semiparametric dynamic panel data models, *Economics Letters* **76**, 1–9.

Bravo, F.: 2002, Blockwise empirical Cressie-Read test statistics for alpha-mixing processes, *Statistics and Probability Letters* **58**, 319–325.

Bravo, F.: 2005, Blockwise empirical entropy tests for time series regressions, *Journal of Time Series Analysis* **26**, 185–210.

Bravo, F.: 2014, Second-order comparisons for nonparametric conditional moment restrictions models, *Technical report*, Unpublished.

Cai, Z.: 2003, Nonparametric estimation equations for time series data, *Statistics and Probability* **62**, 379–390.

Cai, Z., Fan, J. and Li, R.: 2000, Efficient estimation and inference for varying-coefficient models, *Journal of the American Statistical Association* **95**, 888–902.

Cai, Z., Fan, J. and Yao, Q.: 2000, Functional-coefficient regression models for nonlinear time series, *Journal of the American Statistical Association* **95**, 941–956.

Cai, Z. and Li, Q.: 2008, Nonparametric estimation of varying coefficient dynamic panel data models, *Econometric Theory* **24**, 1321–1342.

Cleveland, W., Grosse, E. and Shyu, W.: 1991, *Local Regression Models*, Wadsworth/Brooks-Cole, pp. 309–376.

Davidson, J.: 1994, *Stochastic limit theory*, Oxford University Press.

DiCiccio, T. and Romano, J. P.: 1990, Nonparametric confidence limits by resampling methods and least favourable families, *International Statistical Review* **58**, 59–76.

Embrechts, P., Kluppelberg, C. and Mikosch, T.: 1997, *Modelling Extremal Events for Insurance and Finance*, Springer : Berlin, Heidelberg, New York.

Fan, J. and Gijbels, I.: 1996, *Local Polynomial Modeling and Applications*, Chapman and Hall, London.

Fan, J. and Wu, Y.: 2008, Semiparametric estimation of covariance matrices for longitudinal data, *Journal of the American Statistical Association* **103**, 1520–1533.

Fan, J. and Zhang, W.: 2008, Statistical methods with varying coefficients models, *Statistics and its interface* **1**, 179–195.

Guggenberger, P., Ramalho, J. and Smith, R.: 2012, GEL statistics under weak identification, *Journal of Econometrics* **170**, 331–349.

Hahn, J. and Kuersteiner, G.: 2002, Asymptotically unbiased inference for a dynamic panel data model with fixed effects when both n and T are large, *Econometrica* **70**, 1639–1657.

Hall, P. and Heyde, C.: 1980, *Martingale limit theory and its application*, Academic Press, New York.

Hastie, T. and Tibshirani, R.: 1993, Varying-coefficient models, *Journal of the Royal Statistical Society B* **55**, 757–796.

Henderson, D., Carroll, R. and Li, Q.: 2008, Nonparametric estimation and testing of fixed effects panel models, *Journal of Econometrics* **144**, 257–275.

Henderson, D. and Ullah, A.: 2005, A nonparametric random effect estimator, *Economics Letters* **88**, 403–407.

Hsiao, C.: 2003, *Analysis of Panel Data*, Cambridge University Press.

Kitamura, Y. and Stutzer, M.: 1997, An information theoretic alternative to generalized method of moments estimation, *Econometrica* **65**, 861–874.

Lee, Y.: 2014, Nonparametric estimation of dynamic panel models with fixed effects, *Econometric Theory* **30**, 1315–1347.

Li, Q. and Stengos, T.: 1996, Semiparametric estimation of partially linear panel data models, *Journal of Econometrics* **71**, 289–397.

Li, Q. and Ullah, A.: 1998, Estimating partially linear panel data modles with one-way error components, *Econometric Reviews* **17**, 145–166.

Lin, X. and Carroll, R.: 2006, Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society, B* **68**, 68–88.

Masry, E.: 1996, Multivariate local polynomial regression for time series: Uniform strong consistency and rates, *Journal of Time Series Analysis* **17**, 571–599.

Newey, W. and Smith, R.: 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**, 219–256.

Owen, A.: 1988, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika* **36**, 237–249.

Owen, A.: 1990, Empirical likelihood ratio confidence regions, *Annals of Statistics* **18**, 90–120.

Phillips, P. C. B. and Moon, H.: 1999, Linear regression limit theory for nonstationary panel data, *Econometrica* **67**, 1057–1111.

Rao, C.: 1973, *Linear Statistical Inference and Its Applications*, 2nd edn, New York: Wiley.

Su, L. and Ullah, A.: 2011, *Nonparametric and semiparametric panel econometric models: Estimation and testing*, CRC/Chapman & Hall, chapter 16, pp. 455–498.

Sun, Y., Carroll, R. and Li, D.: 2009, Semiparametric estimation of fixed effects panel data varying coefficient models, *Advances in Econometrics* **25**, 101–130.

Tran, K. and Tsionas, E.: 2010, Local GMM estimation of semiparametric panel data with smooth coefficient models, *Econometric Reviews* **29**, 39–61.

Volkonskii, V. and Rozanov, Y.: 1959, Some limit theorems for random functions: Part i, *Theory of Probability and its* **4**, 178–197.