Likelihood Based Inference for Dynamic Panel Data Models

Seung C. Ahn
Department of Economics
Arizona State University

Gareth M. Thomas
Department of Economics
Arizona State University

This Version: February, 2004

Abstract

This paper considers maximum likelihood (ML) based inferences for dynamic panel data models. We focus on the analysis of the panel data with a large number of cross-sectional units and a small number of repeated time-series observations for each cross-sectional unit. We examine several different ML estimators and their asymptotic and finite-sample properties. Our major finding is that when data follow unit-root processes, the ML estimators have singular information matrices. This is not a non-identification problem because the ML estimators are still consistent. Nonetheless, the estimators have nonstandard asymptotic distributions and their convergence rates are lower than $N^{1/2}$. For this reason, the sizes of the Wald unit-root tests are severely distorted even asymptotically, and they reject the unit-root hypothesis too often. However, following Rotnitzky, Cox, Bottai and Robins (2000), we show that likelihood ratio (LR) tests for unit root follow mixtures of chi-square distributions. Our Monte Carlo experiments show that the LR tests are much better sized than the Wald tests, although they tend to slightly over-reject the unit root hypothesis in small samples. It is also shown that the LR tests have good finite-sample power properties.

Acknowledgment

The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. We would like to thank participants in the econometrics seminar at Rice University, the eighth annual Texas Econometrics Camp, and the University of California at San Diego.

1. Introduction

Panel data models assume that individual cross-section units have different intercept terms due to unobservable heterogeneity. Two different models are used to control the unobservable heterogeneity. One is the random effects (RE) model in which the individual-specific intercept terms, or individual effects, are treated as random variables. The other is the fixed effects (FE) model in which the effects are treated as parameters. The FE model is more general than the RE model in that it requires weaker distributional assumption about the effects. One difficulty in estimating the fixed effects models however is that the number of parameters increases with N. A traditional treatment for this so-called "incidental parameters problem" (Neyman and Scott, 1948) is the within estimator, i.e., least squares on data transformed into deviations from individual means, which is also a ML estimator. For models with strictly exogenous regressors, the within estimator is consistent. Unfortunately, however, when T is small, the within estimator is inconsistent for the dynamic models that use lagged dependent variables as regressors (Nickell, 1981). One way to avoid this problem is to use the random effects ML estimator that treats the effects as time invariant random variables (Hsiao, 1986). Instead, generalized method of moments (GMM) estimators have been widely used to analyze the FE dynamic panel models (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995, 1997; and Blundell and Bond, 1998). An important reason for the popularity of GMM is that it provides consistent estimators under quite general FE assumptions.

Recently, research interests in the ML estimation of dynamic panel data models have been revived. For example, Lancaster (2002), and Hsiao, Pesaran and Tahmiscoglu (2002, HPT) have developed alternative ML estimators for FE dynamic models.[1] Kruiniger (2002a) provides the general conditions under which the ML estimators of Lancaster and HPT are consistent. One reason for the recent revival of the ML approach may be that the panel data GMM estimators often have poor finite-sample properties (e.g., Bond and Windmeijer, 2002). Dynamic panel data models imply a large number of moment conditions. The GMM estimators imposing all of the available moment conditions appear to generate biased statistic inferences, especially when T is large. Thus, an important research agenda would be to develop the

[1] Hahn and Kuersteiner (2002) also consider the ML estimator for the FE dynamic model with large N and T. They find the ML estimator, which is the within estimator, is consistent, but it is asymptotically biased. Hahn and Kuersteiner provide a biased-corrected ML estimator.

alternative GMM estimators that use a smaller number of moment conditions, but without substantial loss of asymptotic efficiency. Ahn and Schmidt (1995) show that when data are normally distributed, an efficient GMM estimator is asymptotically identical to the ML estimator constructed on the same first and second order moment conditions. Given that ML estimators are the GMM estimators based on exact identifying moment conditions, the ML-based approach may be a viable alternative to the popular GMM approach.

This paper considers the asymptotic and finite-sample properties of the RE ML estimator and the two FE ML estimators of Lancaster (2002) and HPT. The relative asymptotic efficiency of these three estimators has been studied by Kruiniger (2002a). This paper readdresses this relative efficiency issue in a more systematic method. We also compare the finite-sample properties of the estimators.

Our major finding in this paper is that when data follow a unit root process, the information matrix of the RE ML estimator becomes singular. Other ML estimators suffer from the same problem. This, however, is not a non-identification problem. The RE and HPT ML methods can identify the parameters and provide consistent estimators, while the Lancaster ML cannot identify parameters. Rotnitzky, Cox, Bottai and Robins (2000, RCBR) analyze the general asymptotic properties of ML estimators when their information matrices are singular. Following their approach, we derive the asymptotic distribution and convergence rate of the RE and HPT ML estimators. Their asymptotic distributions are not normal and their convergence rates are lower than $\sqrt{N}$. Thus, Wald-type tests for unit root tests generate biased inferences. In contrast, likelihood ratio (LR) test statistics for unit root follow mixed χ^2 distributions that can be easily simulated by Monte Carlo experiments. We find that the LR tests are much better sized than other Wald-type tests.

2. ML Estimation

The foundation of this paper is the simple dynamic model

$$y_{it} = \delta y_{i,t-1} + (\alpha_i + \varepsilon_{it}). \tag{1}$$

Here i = 1, 2, ..., N denotes cross-section unit (individual) and t = 1, 2, ..., T denotes time. Our parameter of interest is δ. We initially assume that δ is within a unit circle. We will relax this assumption later. The composite error $(\alpha_i + \varepsilon_{it})$ contains a time invariant individual effect α_i

and random noise ε_{it}. The initial observed value of y for individual i is y_{i0}. We assume that the random error vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$ is uncorrelated with y_{i0} and α_i, and that the ε_i, y_{i0} and α_i are cross-sectionally independent. We assume that the error terms ε_{it} are serially uncorrelated. For the maximum likelihood estimation of model (1), we need to make distributional assumptions about the y_{i0} and α_i. We assume that all of the ε_i, y_{i0} and α_i are normally distributed:

$$\begin{pmatrix} y_{i0} \\ \alpha_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0_{T\times 1} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{0\alpha} & 0_{1\times T} \\ \sigma_{0\alpha} & \sigma_\alpha^2 & 0_{1\times T} \\ 0_{T\times 1} & 0_{T\times 1} & \nu I_T \end{pmatrix} \right), \tag{2}$$

where $\nu = \mathrm{var}(\varepsilon_{it})$. Here, we assume that the y_{i0} and α_i have zero means. This assumption is just for convenience. We can allow nonzero means without altering our main results. Since we make an explicit normality assumption about the individual effects α_i, we shall refer to (2) as the random-effect (RE) assumption. In this paper, we do not consider the more general models that contain exogenous regressors. However, our results can be easily extended to such models. The ML estimation of model (1) has been considered by Anderson and Hsiao (1981), Bhargava and Sargan (1983), and Hsiao (1986). We can easily derive the log-likelihood function, viewing (1) as a recursive simultaneous equations model (treating y_{i0}, y_{i1}, ... , y_{iT} as endogenous variables).

The log-likelihood function for model (1) depends on the five parameters, δ, $\sigma_{0\alpha}$, σ_0^2, σ_α^2 and ν. However, we find a convenient reparameterization by which the parameter σ_0^2 can be orthogonalized to other parameters. Using this method, the parameter of our interest δ can be estimated independently from σ_0^2. In addition, this reparameterization will facilitate our comparisons of the RE ML estimator the two alternative fixed-effects (FE) ML estimators developed by Hsiao, Pesaran and Tahmiscoglu (2002, HPT), and Lancaster (2002).

We define

$$p_i \equiv (\delta - 1) y_{i0} + \alpha_i \,; E(p_i \mid y_{i0}) = \psi y_{i0}\,;\;\; u_i = \left(p_i - \psi y_{i0}\right) + \varepsilon_{i1},$$

where $\psi = (\delta - 1) + \sigma_{0\alpha} / \sigma_0^2$. With this notation, a one-to-one transformation of model (1) leads to

$$\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i \end{pmatrix} = \begin{pmatrix} y_{i0} \\ 0_{(T-1)\times 1} \end{pmatrix} \psi + \begin{pmatrix} 0 \\ \Delta y_{i,-1} \end{pmatrix} \delta + \begin{pmatrix} u_i \\ \Delta \varepsilon_i \end{pmatrix}, \tag{3}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$, and

$$\Delta y_i = (\Delta y_{i2}, ..., \Delta y_{iT})'; \;\; \Delta y_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{i,T-1})'; \;\; \Delta \varepsilon_i = (\Delta \varepsilon_{i2}, ..., \Delta \varepsilon_{iT})'.$$

Under the RE assumption, the variance-covariance matrix of the error vector $(u_i, \Delta\varepsilon_i')'$ is given:

$$Var\begin{pmatrix} u_i \\ \Delta \varepsilon_i \end{pmatrix} \equiv v\Omega_T \equiv v \begin{pmatrix} \omega & -c_{T-1}' \\ -c_{T-1} & B_{T-1} \end{pmatrix}, \tag{4}$$

where $\omega = \mathrm{var}(u_i)/v = (\alpha_\alpha^2 - \sigma_{0\alpha}^2/\sigma_0^2)/v + 1$, c_{T-1} is a (T-1)×1 vector whose first entry equals one while other entries equal zero, $B_{T-1} = [B_{T,jh}]$ is a (T-1)×(T-1) matrix such that $B_{T-1,jj} = 2$, $B_{T-1,j,j+1} = B_{T,j,j-1} = -1$, and all other entries equal zero. We can show that

$$\Omega_T^{-1} = \begin{pmatrix} 0 & 0_{1\times(T-1)} \\ 0_{(T-1)\times 1} & B_{T-1}^{-1} \end{pmatrix} + \frac{T}{\xi_T} \begin{pmatrix} 1 \\ k_{T-1} \end{pmatrix} \begin{pmatrix} 1 & k_{T-1}' \end{pmatrix}, \tag{5}$$

where $k_{T-1} = ((T-1)/T, (T-2)/T, ..., 1/T)'$.

In the reparameterized model (3), the parameter vector to be estimated jointly is given by $\theta = (\delta, v, \omega, \psi)'$. The initial observation y_{i0} is uncorrelated with the error vector $(u_i, \Delta\varepsilon_i')'$. So, we can construct a likelihood function treating y_{i0} as predetermined. This is so because under our normality assumption,

$$f_i(\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT}, y_{i0} \mid \theta, \sigma_0^2) = f_i(\Delta y_{i1}, ..., \Delta y_{iT} \mid y_{i0}, \theta) f_i(y_{i0} \mid \sigma_0^2).$$

From now on, we use notation $f_i(\bullet)$ to denote a density function for the i'th observation; and $\ell_i(\theta)$ to denote the log of $f_i(\bullet)$. Under the RE assumption, and using (4) and (5), we can easily derive the log-density function of $(\Delta y_{i1}, \Delta y_{i2}, ... \Delta y_{iT})'$ conditional on y_{i0}: [2]

[2] For the cases in which the y_{i0} and α_i do have non-zero expectations, the term $(\Delta y_{i1} - \psi y_{i0})$ will be replaced by $(\Delta y_{i1} - \psi \Delta y_{i0} - a)$, where a is a constant.

$$
\begin{aligned}
\ell_{RE,i}(\theta) &\equiv \ell_i(\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT} \mid \theta, y_{i0}) \\
&= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln(\xi_T) - \frac{1}{2\nu}\left(\Delta y_i - \delta\Delta y_{i,-1}\right)' B_{T-1}^{-1}\left(\Delta y_i - \delta\Delta y_{i,-1}\right) \\
&\quad - \frac{T}{2\nu\xi_T}\left((\Delta y_{i1} - \psi y_{i0}) - k_{T-1}{}'(\Delta y_i - \delta\Delta y_{i,-1})\right)^2,
\end{aligned} \tag{6}
$$

where $\xi_T = \det(\Omega_T) = T\omega - (T-1)$.[3]

The direct log-density function for model (1) can be easily derived under the RE assumption. However, our simulation experiments reveal that the ML procedure based on this the direct log-likelihood often fails to locate the maximum point especially when the true value of δ is close to one. In contrast, we find that the ML procedure based on the log-density function (6) never fails to find the maximum point. Another advantage of the ML estimation based on (6) instead of the direct likelihood function is that it does not require the joint normality of the initial values y_{i0} and the effects α_i. It only requires that the effects α_i are normal conditional on the y_{i0} and their conditional means are linear in the y_{i0}.

The function (6) also provides a foundation by which the ML estimator of δ can be compared to the HPT and Lancaster estimators. HPT propose to estimate δ based on the following differenced model: [4]

$$
\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i \end{pmatrix} = \begin{pmatrix} 0 \\ \Delta y_{i,-1} \end{pmatrix}\delta + \begin{pmatrix} u_{HPT,i} \\ \Delta\varepsilon_i \end{pmatrix}, \tag{7}
$$

where,

$$
Var\begin{pmatrix} u_{HPT,i} \\ \Delta\varepsilon_i \end{pmatrix} = \nu\Omega_{HPT,T} = \nu\begin{pmatrix} \omega_{HPT} & c_{T-1}{}' \\ c_{T-1} & B_{T-1} \end{pmatrix}. \tag{8}
$$

If we assume the normality of the error vector $(u_{HPT,i}, \Delta\varepsilon_i{}')'$, the model (7) leads to the following log-density function:

$$
\begin{aligned}
\ell_{HPT,i}(\delta,\nu,\omega) &\equiv -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln(\xi_{HPT,T}) - \frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) \\
&\quad - \frac{1}{2\nu\xi_{HPT,T}}\left(\Delta y_{i1} - k_{T-1}{}'(\Delta y_i - \delta\Delta y_{i,-1})\right)^2,
\end{aligned} \tag{9}
$$

---

[3] For the derivation of $\det(\Omega_T)$, see HPT.

[4] HPT in fact include an intercept term a for the Δy_{i1} equation. But the interceptor term equals zero under our zero-mean assumptions.

where $\xi_{HPT,T} = \det(\Omega_{HPT,T}) = T\omega_{HTP} - (T-1)$. Observe that this log-density function and the corresponding log-likelihood function depends on only three parameters, while the RE log-likelihood function depends on four parameters.

While the function (9) is almost identical to (6), a critical difference between the two functions is that the former does not depend on the initial value y_{i0}. That is, (9) is the unconditional log-density of $\Delta y_{i1}, ..., \Delta y_{iT}$. Note that under the RE assumption,

$$u_{HPT,i} = \psi y_{i0} + u_i\,,\;\; \omega_{HPT} = \mathrm{var}(u_{HPT,i})/\nu = \left(\psi^2\sigma_0^2\right)/\nu + \omega\,. \tag{10}$$

The HPT ML method treats the y_{i0} as unobservables. This treatment does not lead to inconsistent estimators. But it will lead to inefficiency under the RE assumption. The HPT estimation does not exploit the information about δ that is contained in the level data y_{i0}. It only exploits the information contained in differenced data. In contrast, the RE estimator utilizes the information about δ contained in level data (through the non-zero correlation with Δy_{i1} and y_{i0}). Of course, both the estimators become asymptotically equivalent if $\psi = 0$.

An intriguing question is whether the HPT estimator is a FE treatment in the sense that it does not require any distributional assumption about the α_i. Observe that the HPT estimator requires the normality of $u_{HPT,i} = (\delta\text{-}1)y_{i0} + \alpha_i + \varepsilon_{i1}$. This essentially requires the normality of y_{i0}, but not necessarily the normality of α_i under some circumstances as we see below.

Suppose that the y_{it} follow a stationary process. Specifically, assume that

$$y_{i0} = \eta_i + q_{i0};\quad y_{it} = \eta_i + q_{it};\quad q_{it} = \delta q_{i,t-1} + \varepsilon_{it},\quad t = 1,...,T. \tag{11}$$

This process implies that $y_{it} = \delta y_{i,t-1} + (1-\delta)\eta_i + \varepsilon_{it}$, where $(1-\delta)\eta_i$ equals α_i in our notation. Under this assumption, $p_i = (\delta-1) + \alpha_i = (\delta-1)q_{i0}$. Thus, as long as y_{i0} (or q_{i0}) is normal conditional on α_i, $u_{HPT,i}$ remains normal. When the stationary condition does not hold, the normality of $u_{HPT,i}$ requires the normality of α_i. This means that the HPT estimator is indeed a FE treatment when data follow the stationary process (11). However, when data are not stationary, the HPT estimator may not be viewed as a real FE treatment because it requires a distributional assumption about the effects α_i.

As we mentioned earlier, the RE estimator does not require the normality of y_{i0}, but it requires the normality of α_i conditional on y_{i0} to secure the normality of the error term in (3), $u_i = \left(\alpha_i - E(\alpha_i \mid y_{i0})\right) + \varepsilon_{i1}$. Thus, even if the data are stationary, the RE ML estimator cannot be viewed as a FE effect treatment. However, it is important to note that the consistency of the RE ML estimator does not require the normality of the effects.

We now turn to the Lancaster estimator which has a Bayesian flavor. We derived model (3) by first-differencing model (1). Instead, if we difference out y_{i0} from y_{it}, model (1) reduces to

$$\Delta_0 y_i = \Delta_0 y_{i,-1}\delta + 1_T\, p_i + \varepsilon_i \,,$$

where $p_i = (\delta-1)y_{i0} + \alpha_i$, $\mathbf{1}_T$ is a T×1 vector of ones, $\Delta_0 y_i = (y_{i1}-y_{i0}, y_{i2}-y_{i0},..., y_{iT}-y_{i0})'$ and $\Delta_0 y_{i,-1} = (0, y_{i1}-y_{i0}, y_{i2}-y_{i0},..., y_{i,T-1}-y_{i0})'$. Lancaster treats the p_i as the unobservable effects instead of the α_i. This treatment is similar to that of HPT in that both do not exploit the information contained in the level of y_{i0}.

The density of $\Delta_0 y_i$ conditional on δ, ν and p_i equals

$$\begin{aligned} f_{FE,i}(\delta,\nu,p_i) &= \frac{1}{(2\pi)^{T/2}\nu^{T/2}}\exp\left(-\frac{1}{2\nu}(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)'(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)\right) \\ &= \frac{1}{(2\pi)^{T/2}\nu^{T/2}}\exp\begin{pmatrix} -\frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) \\ -\frac{T}{2\nu}(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta) - p_i)^2 \end{pmatrix}, \end{aligned} \tag{12}$$

where the second equality is shown in the appendix. Observe that $y_{it} - y_{i0} = \Sigma_{s=1}^{t}\Delta y_{is}$. Thus, the second equality of (12) implies that the density $f_{FE,i}(\delta,\nu,p_i)$ can be also viewed as a density of the first-differenced y_{it}'s conditional on δ, ν and p_i.

Using the method of Cox and Raid (1987), Lancaster reparameterizes the effects p_i defining $p_i = \tilde{p}_i \exp(-b(\delta))$, where

$$b(\delta) = \frac{1}{T}\Sigma_{t=1}^{T-1}\frac{T-t}{t}\delta^t \,. \tag{13}$$

This reparameterization is chosen so that the new fixed effects $\tilde{p}_i$ are *information orthogonal* to (δ,ν)′, in the sense that

$$E\left(\frac{\partial^2 \ln(f_{FE,i})}{\partial(\delta,\nu)'\partial\tilde{p}_i}\right) = 0_{2\times1} \,.$$

With this reparameterization and assuming uninformative uniform priors to the $\tilde{p}_i$, we can integrate them out from $f_{FE,i}$. If we do so, we can show that the conditional density of the differenced y_{it}'s, $(\Delta y_{i1},\Delta y_{i2,\ldots,}\Delta y_{iT})'$, conditional on $(\delta,\nu)'$, is given

$$f_{Lan,i}(\delta,\nu) \propto \frac{1}{\nu^{(T-1)/2}}\exp(b(\delta))\exp\left(-\frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta)\right),$$

which leads to the log-density function

$$\ell_{Lan,i}(\delta,\nu) = b(\delta) - \frac{(T-1)}{2}\ln(\nu) - \frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta). \tag{14}$$

Observe that Lancaster's ML depends on only two parameters. It does not depend on the variance ω of the composite error term u_i in model (3). Notice that the error term u_i contains the projection error component of the effect α_i (the error in the population regression of α_i on y_{i0}). Thus, the fact that (14) does not depend on ω seems to imply that the Lancaster estimator is indeed a fixed-effect treatment. However, it is not without costs. That is, while the Lancaster estimator does not require any distributional assumption about the effect α_i, it loses the usual ML properties as we see below.

An interesting property of the Lancaster estimator, $(\hat{\delta}_{Lan},\hat{\nu}_{Lan})'$ is that the asymptotic covariance matrix of the Lancaster estimator is not of the inverted Hessian form. Define

$$H_{Lan,i}(\delta,\nu) = \frac{\partial^2 \ell_{Lan,i}}{\partial(\delta,\nu)'\partial(\delta,\nu)}; B_{Lan,i}(\theta_{Lan}) = \left(\frac{\partial \ell_{Lan,i}}{\partial(\delta,\nu)'}\right)\left(\frac{\partial \ell_{Lan,i}}{\partial(\delta,\nu)}\right)'.$$

Then, it can be shown[5]:

$$\sqrt{N}\begin{pmatrix}\hat{\delta}_{Lan}-\delta_o \\ \hat{\nu}_{Lan}-\nu_o\end{pmatrix} \rightarrow_d N\left(0_{2\times 1},\left(-E(H_{Lan,i}(\delta_o,\nu_o)\right)^{-1}E(B_{Lan,i}(\delta_o,\nu_o))\left(-E(H_{Lan,i}(\delta_o,\nu_o)\right)^{-1}\right),$$

where "$\rightarrow_d$" means "converges in distribution," and $(\delta_o,\nu_o)'$ means the true value of (δ, ν)′. The reason why the asymptotic covariance matrix of the Lancaster estimator is not of the inverted Hessian form is that $\int f_{Lan,i}(\theta_{Lan})d(\Delta y_{i1},\ldots\Delta y_{iT}) \neq 1$. That is, $f_{Lan,i}(\theta_{Lan})$ is not a proper density. Thus, the Lancaster estimator is not really a ML estimator.[6] This problem arises from the orthogonal reparameterization used by Lancaster.

Kruiniger (2002a) shows that the Lancaster estimator of δ is inefficient compared to the

---

[5] See Kruiniger (2002a).

[6] Lancaster also acknowledges this point.

HPT estimator of δ when T > 2.[7] We can obtain the same result in a more intuitive way.

Proposition 1: Suppose that the prior density of p_i, $f_i(p_i|\omega_{HPT},\nu)$, is given $N(0,(\omega_{HPT}-1)\nu)$. Then,

$$\int_{-\infty}^{\infty} f_{FE,i}(\delta,\nu,p_i)\, f_i(p_i \mid \omega_{HPT},\nu)dp_i = f_{HPT,i}(\theta_{HPT}) .$$

All of the proofs are in the appendix. Proposition 1 implies that if $f_{FE,i}(\delta,\nu,p_i)$ were integrated with the normal prior $f_i(p_i|\omega_{HPT},\nu)$, the Lancaster estimator becomes equivalent to the HPT estimator. This explains why the HPT estimator should be more efficient than the Lancaster estimator under our RE assumption or the assumptions justifying the HPT ML method. Under the RE assumption, the prior $f_i(p_i|\omega_{HPT},\nu)$ is informative.

3. Maximum Likelihood When Data Are Random Walks

3.1. Random Walks without Drifts

In this section, we consider the asymptotic distribution of the random effects ML estimator when data follow unit root processes. We begin by considering the cases in which the y_{it} are random walks:

$$y_{it} = y_{i,t-1} + \varepsilon_{it}, \quad t = 1,...,T. \tag{15}$$

Let θ_o be the true value of θ. Then, we can easily see that the data generating process (15) is equivalent to the parametric restriction $\theta = (\delta,\nu,\omega,\psi)' = (1,\nu_*,1,0)'$ on model (3), where ν_* remains unrestricted.

We define:

$$H_{RE,i}(\theta) = \ell_{RE,i,\theta\theta}; B_{RE,i} = \left(\ell_{RE,i,\theta}\right)\left(\ell_{RE,i,\theta}\right)',$$

where we use the subscripted 'θ' to denote the derivative with respect to θ. We will use the

[7] When T = 2, both estimators are asymptotically equivalent, although they are numerically different. This however does not mean that $f_{Lan,i}(\theta_{Lan})$ becomes a proper density function when T = 2. As shown in the appendix (Lemma A.7), the variance-covariance matrix of the Lancaster estimator depends on both $b' = db/d\delta$ and $b'' = d^2b/d\delta^2$, while the variance-covariance matrix of the HPT depends only on b'. However, when T = 2, $b'' = 0$. This provides some intuition as to why the two estimators are asymptotically identical when T = 2. However, as shown in the next section, this equivalence no longer holds when data follow random walks.

same rule to denote the derivatives of the log-density function with respect to individual parameters: for example, $\ell_{RE,i,\delta} = \partial \ell_{RE,i} / \partial \delta$ and $\ell_{RE,i,\delta\delta} = \partial^2 \ell_{RE,i} / \partial \delta^2$.

In usual maximum likelihood theory, the information matrices of log-density (in our case, ($E(-H_{RE,i}(\theta))$) are assumed to be nonsingular. However, when data are generated following the process (15), this is no longer the case. We state this finding formally.

Proposition 2: Let $\theta_* = (1, \nu_*, 1, 0)'$. When $\theta_o = \theta_*$, $E(-H_{RE,i}(\theta_o))$ and $E(B_{RE,i}(\theta_o))$ are singular.

The proposition implies that the information matrix of the RE log-likelihood function (6) is singular. This proposition also applies to the HPT and Lancaster ML estimation. It means that the usual asymptotic theory of ML estimation does not apply to the RE, HPT and Lancaster ML estimation when data follow random walks without drift. Singular information matrices often imply model non-identification. However, Proposition 2 does not imply that the RE model (3) is not identifiable when $\theta_o = \theta_*$. If the RE model were not identifiable, it should be the case that θ_* is not a unique maximizer of $E(\ell_{RE,i}(\theta))$. But this is not the case. To see why, consider a simple case of T = 2. If we concentrate $E(\ell_{RE,i}(\theta))$ by maximizing it with respect to the nuisance parameter vector $(\nu, \omega, \psi)'$ given δ, we can obtain:

$$E\left(\ell^c_{RE,i}(\delta)\right) = -\frac{1}{2}\ln\left(\frac{1}{2} + \frac{\delta^2}{2}\right) - \frac{1}{2}\ln\left(\frac{5}{2} - 2\delta + \frac{1}{2}\delta^2\right) + g(\nu_*, T),$$

where the superscript "c" means "concentrated", and g(ν*,N,T) is some function of the variance parameter ν* and T. Then, it can be show that at δ = 1,

$$E(\ell^c_{RE,i,\delta}(\delta)) = E(\ell^c_{RE,i,\delta\delta}(\delta)) = E(\ell^c_{RE,i,\delta\delta\delta}(\delta)) = 0; E(\ell^c_{RE,i,\delta\delta\delta\delta}(\delta)) = -3,$$

where the subscripts denote the derivatives with respect to the corresponding parameters. Thus, although the second derivative of $\ell^c_{RE,i}(\delta)$ equals zero, δ = 1 is still a local maximum point. Moreover, it can be shown that

$$E\left(\ell^c_{RE,i,\delta}(\delta)\right) = -\frac{2(\delta-1)^3}{\left[(\delta-2)^2+1\right](\delta^2+1)}.$$

Thus, $E\left(\ell^{c}_{RE,i}(\delta)\right)$ is always increasing when δ < 1 and decreasing when δ > 1. This indicates that δ = 1 is the global maximum point of $E\left(\ell^{c}_{RE,i}(\delta)\right)$. This result is confirmed in Figure 1. This result also implies that θ* is the global maximum point of $E(\ell_{RE,i}(\theta))$ when $\theta_o = \theta_*$. We can generalize this result to the cases with general T. Stated formally:

Proposition 3: Suppose $\theta_o = \theta_*$. Then, θ* is a unique global maximizer of $E(\ell_{RE,i}(\theta))$. Thus, the RE ML estimator, $\hat{\theta}_{RE} = (\hat{\delta}_{RE}, \hat{\nu}_{RE}, \hat{\omega}_{RE}, \hat{\psi}_{RE})'$, is a consistent estimator when $\theta_o = \theta_*$.

This proposition applies to the HPT estimator. However, somewhat surprisingly, it does not apply to the Lancaster ML estimation. When data are random walks, the point δ = 1 is a inflexion point of the expectation of Lancaster log-likelihood function.

Proposition 4: Suppose $\theta_o = \theta_*$. Then, θ* does not maximize $E\left(\ell_{Lan,i}(\delta,\nu)\right)$. In fact, θ* is an inflexion point.

This proposition is shown by investigating the concentrated expected log-likelihood function $E\left(\ell^{c}_{Lan,i}(\delta)\right)$ constructed similarly to $E\left(\ell^{c}_{RE,i}(\delta)\right)$. For example, when T = 2, we obtain:

$$E\left(\frac{\partial \ell^{c}_{Lan,i}(\delta)}{\partial \delta}\right) = \frac{(\delta-1)^2}{2(\delta^2+1)} \geq 0 .$$

Thus, $E\left(\ell^{c}_{Lan,i}(\delta)\right)$ is always increasing except when δ = 1. The actual shape of $E\left(\ell^{c}_{Lan,i}(\delta)\right)$ is given in Figure 2. This result indicates that the Lancaster estimator of δ is inconsistent. This is a somewhat counterintuitive result in that the HPT estimator is still consistent when T = 2. Recall that when $\delta_o < 1$ and T = 2, that the Lancaster and HPT estimators are asymptotically equivalent. Proposition 4 implies that this equality breaks down when data are random walks ($\theta_o = \theta_*$).

Although the RE and HPT estimators are consistent, Proposition 2 indicates that their asymptotic distributions may not be normal when $\theta_o = \theta_*$. We now investigate the asymptotic distributions of these estimators using the method developed by Rotnitzky, Cox, Bottai and Robins (2000, RCBR). Here, we will focus on the RE estimator only. All of the results we obtain below also apply to the HPT estimator.

RCBR study the cases in which the derivatives of log-density functions are linearly dependent (so that information matrices become singular). For such cases, they derive the asymptotic distributions of ML estimators and likelihood-ratio (LR) test statistics. To use their method, we need to show that the derivatives of $\ell_{RE,i}$ are linearly independent. Stated formally:

Proposition 5: $\ell_{RE,i,\delta}(\theta_*) - \ell_{RE,i,\omega}(\theta_*) + \nu_* \ell_{RE,i,\nu}(\theta_*) = 0$.

To get tractable asymptotic results, we now need to reparameterize the model. Following RCBR, we define the following new parameter vector:

$$\theta_n \equiv \begin{pmatrix} \delta_n \\ \nu_n \\ \omega_n \\ \psi_n \end{pmatrix} = \begin{pmatrix} \delta \\ \nu \\ \omega \\ \psi \end{pmatrix} + \begin{pmatrix} 0_{1\times 3} \\ \left[E\left(\ell_{RE,i,\varphi}(\theta_*)\ell_{RE,i\varphi}(\theta_*)'\right)\right]^{-1} E\left(\ell_{RE,i,\varphi}(\theta_*)\ell_{RE,i,\delta}(\theta_*)\right) \end{pmatrix}(\delta-1)$$

$$= \begin{pmatrix} \delta \\ \nu \\ \omega \\ \psi \end{pmatrix} + \begin{pmatrix} 0 \\ -\nu_* \\ 1 \\ 0 \end{pmatrix}(\delta-1) = \begin{pmatrix} \delta \\ \nu - \nu_*(\delta-1) \\ \omega + (\delta-1) \\ \psi \end{pmatrix},$$

where $\varphi = (\nu,\omega,\psi)'$. This parameterization is chosen to secure that when $\theta = \theta_*$, $\phi = \theta_*$. The reparameterization requires the knowledge of the true variance of ε_{it}, ν_*. However, this problem can be resolved by replacing ν_* by ν_n.

The above reparameterization means that we retain δ and ψ in $\ell_{RE,i}(\theta)$, but treat ω and ν as the functions of δ and the new parameters ω_n and ν_n:

$$\omega = \omega_n - (\delta-1); \nu = \nu_n + \nu_n(\delta-1) = \nu_n\delta .$$

Let $\theta_n = (\delta,\nu_n,\omega_n,\psi)'$;

$$\tilde{\ell}_{RE,i}(\theta_n) = \ell_{RE,i}(\delta,\nu(\delta,\nu_n),\omega(\delta,\omega_n),\psi);$$

$$s_{i,\delta\delta}(\theta_n) = \tilde{\ell}_{RE,i,\delta\delta}(\theta_n); s_{i,\nu_n}(\theta_n) = \tilde{\ell}_{RE,i,\nu_n}(\theta_n); s_{i,\omega_n}(\theta_n) = \tilde{\ell}_{RE,i,\omega_n}(\theta_n); s_{i,\psi}(\theta_n) = \tilde{\ell}_{RE,i,\psi}(\theta_n);$$

$$s_i(\theta_n) = (s_{i,\delta\delta}(\theta_n)/2, s_{i,\nu_n}(\theta_n), s_{i,\omega_n}(\theta_n), s_{i,\psi}(\theta_n))'$$

$$\Upsilon = Var(s_i(\theta_*) \mid \theta_{n,o} = \theta_*) = [\Upsilon_{ij}], \quad i, j = 1,2,3,4;$$

$$\Upsilon^{-1} = [\Upsilon^{ij}].$$

Under our reparameterization, $\theta_n = \theta_*$ whenever $\theta = \theta_*$. In addition, the reparameterization is designed to have

$$\tilde{\ell}_{RE,i,\delta}(\theta_*) = \ell_{RE,i,\delta}(\theta_*) - \ell_{RE,i,\omega}(\theta_*) + \nu_* \ell_{RE,i,\nu}(\theta_*) = 0.$$

This reparameterization is necessary because the RCBR approach is basically for the cases where a first derivative of a log-likelihood function equals (not asymptotically, but exactly) zero. The asymptotic distributions of the ML estimators of ω and ν are different from those of the ML estimators of ω_n and ν_n, although the former can be derived from the latter (see RCBR). However, the distribution of the ML estimator of δ can be directly obtained from that of the ML estimator from the reparameterized model.

Using the RCBR method, we need to check (i) whether or not $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ equals zero or is linearly related to $\tilde{\ell}_{RE,i,\delta}(\theta_*)$, $s_{i,\omega_n}(\theta_*)$, $s_{i,\nu_n}(\theta_*)$, and $s_{i,\psi_n}(\theta_*)$, and (ii) whether or not $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is a linear combination of $s_i(\theta_*)$. We find that only $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ equals zero, not $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$, and that $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ is not a linear combination of $\tilde{\ell}_{RE,i,\delta}(\theta_*)$, $s_{i,\omega_n}(\theta_*)$, $s_{i,\nu_n}(\theta_*)$, and $s_{i,\psi_n}(\theta_*)$. We also find that $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is not a linear combination of $s_i(\theta_*)$. Based on this finding, we obtain the following result:

Proposition 6: Let $Z = (Z_1, Z_2, Z_3, Z_4)'$ denote a mean-zero random vector with $Var(Z) = \Upsilon$. Let $\hat{\theta}_n = (\hat{\delta}, \hat{\nu}_n, \hat{\omega}_n, \hat{\psi})'$ be the ML estimator of the reparameterized model. Then, when $\theta_{n,0} = \theta_*$,

$$\begin{pmatrix} N^{1/4}(\hat{\delta}-1) \\ N^{1/2}(\hat{\nu}_n - \nu_*) \\ N^{1/2}(\hat{\omega}_n - 1) \\ N^{1/2}(\hat{\psi} - 0) \end{pmatrix} \to_d \begin{pmatrix} (-1)^B Z_1^{1/2} \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} 1(Z_1 > 0) + \begin{pmatrix} 0 \\ Z_2 - (\Upsilon^{21}/\Upsilon^{11}) Z_1 \\ Z_3 - (\Upsilon^{31}/\Upsilon^{11}) Z_1 \\ Z_4 - (\Upsilon^{41}/\Upsilon^{11}) Z_1 \end{pmatrix} 1(Z_1 < 0),$$

where B is a Bernoulli random variable with success probability equal to half and independent of Z. In addition, the LR test statistic for $H_o : (\delta_o, \omega_o, \psi_o)' = (1,1,0)'$ is a mixture of $\chi^2(2)$ and $\chi^2(3)$ with mixing probabilities equal to half.

Several comments follow on Proposition 6. First, the ML estimator of $\hat{\theta}_n$ is not normal even asymptotically. Since $\hat{\delta}_{RE} = \hat{\delta}$, the above theorem indicates that the convergence rate of $\hat{\delta}_{RE}$ is $N^{1/4}$. The result also implies that the ML-based t-test for H_o: $\delta_o = 1$ would not be properly sized. Second, the probability that the ML estimator $\hat{\delta}_{RE}$ is equal to the true value of δ converges to half as $N \to \infty$. This implies that the asymptotic distribution of the ML estimator $\hat{\delta}_{RE}$ will be humped at $\delta = 1$ when $\theta_o = \theta_*$. Third, we can think of the LR statistic for testing H_o': $\delta_o = 1$. It can be shown that when $\theta_o = \theta_*$, this test statistic follows a mixture of $\chi^2(1)$ and zero. Similarly, the LR statistic for testing H_o'': $(\delta_o,\psi_o) = (1,0)$ follows, when $\theta_o = \theta_*$, a mixture of $\chi^2(2)$ and $\chi^2(1)$.

We can obtain the similar results for the HPT estimator. The HPT-based LR test for the hypothesis that $(\delta,\omega)' = (1,1)'$ follows a mixture of $\chi^2(2)$ and $\chi^2(1)$.

3.2. Random Walks with Drifts

In this section, we consider the ML estimation of the dynamic panel data model with trend, and the LR test for the hypothesis of random walk with drift. Specifically, we consider the following model:

$$y_{it} = \delta y_{i,t-1} + \beta t + (\alpha_i + \varepsilon_{it}), \tag{16}$$

We assume that $(y_{i0}, \alpha_i, \varepsilon_i')'$ satisfies the RE assumption (2), but we now allow the initial values y_{i0} and the effect α_i to have nonzero means. Assume that $E(\alpha_i \mid y_{i0}) = \gamma_1 + \gamma_2 y_{i0}$. Then, similarly to what we have done from (1) to (3), we can transform (16) into

$$\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i \end{pmatrix} = \begin{pmatrix} y_0 & 1 \\ 0_{(T-1)\times 1} & 0_{(T-1)\times 1} \end{pmatrix} \begin{pmatrix} \psi \\ a \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \Delta y_{i,-1} & 1_{T-1} \end{pmatrix} \begin{pmatrix} \delta \\ \beta \end{pmatrix} + \begin{pmatrix} u_i^m \\ \Delta\varepsilon_i \end{pmatrix}, \tag{17}$$

where $a = \gamma_1 + \beta$, $\psi = (1-\delta) + \gamma_2$ and $u_i^m = (\alpha_i - E(\alpha_i \mid y_{i0})) + \varepsilon_{i1}$. The log-density function for this model is given:

$$
\begin{aligned}
\ell_{RE,i}^m(\phi) &\equiv \ell_i(\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT} \mid \phi, y_{i0}) \\
&= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln(\xi_T^m) \\
&\quad -\frac{1}{2\nu}\left(\Delta y_i - \delta \Delta y_{i,-1} - 1_{T-1}\beta\right)' B_{T-1}^{-1}\left(\Delta y_i - \delta \Delta y_{i,-1} - 1_{T-1}\beta\right) \\
&\quad -\frac{T}{2\nu\xi_T}\Big((\Delta y_{i1} - \psi y_{i0} - a) - k_{T-1}{}'(\Delta y_i - \delta \Delta y_{i,-1} - 1_{T-1}\beta)\Big)^2,
\end{aligned} \tag{18}
$$

where $\xi_T^m = T\omega^m - (T-1)$, $\omega^m = \mathrm{var}(u_i^m)/\nu = \left(\sigma_\alpha^2 - \sigma_{0\alpha}^2 / \sigma_0^2\right)/\nu + 1$, and $\phi = (\delta, \nu, \omega^m, \psi, a, \beta)'$.

Suppose that data are generated by the following trend-stationary process:

$$
y_{it} = \eta_i + gt + q_{it}; \quad q_{it} = \delta q_{i,t-1} + \varepsilon_{it}, \quad t = 0, ..., T, \tag{19}
$$

where $\varepsilon_i \sim N(0_{T\times 1}, \nu I_T)$, η_i and q_{i0} are also normally distributed, and all of ε_i, η_i and q_{i0} are mutually independent. Assume that $E(\eta_i \mid y_{io}) = \tau_1 + \tau_2 y_{i0}$.

The data generation process (19) is the same as the process (16), if we set:

$$
\alpha_i = (1-\delta)\eta_i;
$$

$$
\psi = (1-\delta)(-1+\tau_2); \quad a = (1-\delta)\tau_1 + \delta g; \quad \beta = (1-\delta)g; \quad w^m = (\sigma_\alpha^2 - \sigma_{0\alpha}^2/\sigma_0^2)/\nu + 1.
$$

However, there exists an important difference between (16) and (19). In the former, we do not assume that the initial values y_{i0} are distributed around the effect α_i. Thus, we do not impose any restriction on $\sigma_{0\alpha} = \mathrm{cov}(\alpha_i, y_{i0})$. Without any restriction on $\sigma_{0\alpha}$, the parameter vector ϕ is orthogonal to the variance σ_0^2 of the initial values y_{i0}. Thus, the estimation of ϕ based on the conditional log-density (18) is equivalent to the joint estimation of ϕ and α_0^2 based on the unconditional log-density of $(y_{i0}, \Delta y_{i1}, ..., \Delta y_{iT})'$. However this is not the case for (19). The process (19) implies that the y_{i0} are distributed around $\eta_i = \alpha_i/(1-\delta)$, and that

$$
\sigma_{0\alpha} = (1-\delta)\sigma_{0\eta} = (1-\delta)\sigma_\eta^2. \tag{20}
$$

But this restriction implies that

$$
\psi = -(1-\delta) + (1-\delta)^2\sigma_\eta^2/\sigma_0^2; \quad \omega^m = (1-\delta)^2\sigma_\eta^2(1 - 1/\sigma_0^2)/\nu + 1.
$$

Observe that under (19), the two parameters ψ and ω^m depend on only one free parameter σ_η^2

given δ, ν and σ_0^2. This means that any knowledge of σ_0^2 would help to obtain a more efficient estimator of ϕ. This implies that the estimation of ϕ based on the conditional log-density (18) is not equivalent to the joint estimation of ϕ and σ_0^2 based on the unconditional log-density of $(y_{i0}, \Delta y_{i1}, ..., \Delta y_{iT})'$. The ML estimators based on the unconditional density will be more efficient. Of course, these unconditional ML estimators will be inconsistent if the condition (20) is violated.

We now turn to the cases where data are random walks with drifts; that is δ = 1 in (19). We can show that the unconditional ML estimators computed with the restriction (20) are consistent and asymptotically normal. Thus, the usual ML theory applies. However, the conditional ML estimators have a different story. The hypothesis of random walk with a drift implies the following restrictions on (18):

$$H_o^m : \phi_o = \phi_* \equiv (1, \nu_*, 1, 0, a_*)',$$

where ν_* and a_* are unrestricted. When this hypothesis holds, we obtain essentially the same results as Propositions 5 and 6. Stated formally:

Proposition 7: $\ell^m_{RE,i,\delta}(\phi_*) - a_*\ell^m_{RE,i,\beta}(\phi_*) - \ell^m_{RE,i,\omega}(\phi_*) + \nu_*\ell^m_{RE,i,\nu}(\theta_*) = 0$.

Proposition 8: Let $\phi_n = (\delta, \nu_n^m, \omega_n^m, \psi, a, \beta_n^m)'$ such that,

$$\nu = \nu_n^m \delta\,;\quad \omega = \omega_n^m + (\delta - 1)\,;\quad \beta = \beta_n^m - a(\delta - 1)\,.$$

Let $\tilde{\ell}^m_{RE,i}(\phi_n) = \ell^m_{RE,i}(\delta, \nu(\nu_n, \delta), \omega(\delta, \omega_n), \psi, a, \beta(\beta_n, a, \delta))$; and

$$s^m_{i,\delta}(\phi_n) = \tilde{\ell}^m_{RE,i,\delta\delta}(\phi_n); s^m_{i,\nu_n}(\phi_n) = \tilde{\ell}^m_{RE,i,\nu_n}(\phi_n); s^m_{i,\omega_n}(\phi_n) = \tilde{\ell}^m_{RE,i,\omega_n}(\phi_n); s^m_{i,\psi}(\phi_n) = \tilde{\ell}^m_{RE,i,\psi}(\phi_n)\,;$$

$$s^m_{i,a}(\phi_n) = \tilde{\ell}^m_{RE,i,a}(\phi_n)\,;\quad s^m_{i,\beta_n}(\phi_n) = \tilde{\ell}^m_{RE,i,\beta_n}(\phi_n)\,;$$

$$s^m_i(\phi_n) = (s^m_{i,\delta\delta}(\phi_n)/2, s^m_{i,\nu_n}(\phi_n), s^m_{i,\omega_n}(\phi_n), s^m_{i,\psi}(\phi_n), s^m_{i,a}(\phi_n), s^m_{i,\beta_n}(\phi_n))'$$

$$\Upsilon = Var(s^m_i(\phi_*) \mid \phi_n = \phi_*) = [\Upsilon_{ij}],\quad i, j = 1,2,3,4,5,6\,;$$

$$\Upsilon^{-1} = [\Upsilon^{ij}]\,.$$

Let $Z = (Z_1, Z_2, Z_3, Z_4, Z_5, Z_6)'$ denote a mean-zero random vector with $Var(Z) = \Upsilon$. Let $\hat{\phi}_n = (\hat{\delta}, \hat{\nu}_n, \hat{\omega}_n, \hat{\psi}, \hat{a}, \hat{\beta}_n)'$ be the ML estimator. Then, when $\phi_{n,0} = \phi_*$,

$$\begin{pmatrix} N^{1/4}(\hat{\delta}-1) \\ N^{1/2}(\hat{\nu}_n-\nu_*) \\ N^{1/2}(\hat{\omega}_n-1) \\ N^{1/2}(\hat{\psi}_n-1) \\ N^{1/2}(\hat{a}-1) \\ N^{1/2}(\hat{\beta}_n-1) \end{pmatrix} \to_d \begin{pmatrix} (-1)^B Z_1^{1/2} \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \end{pmatrix} 1(Z_1>0) + \begin{pmatrix} 0 \\ Z_2-(\Upsilon^{21}/\Upsilon^{11})Z_1 \\ Z_3-(\Upsilon^{31}/\Upsilon^{11})Z_1 \\ Z_5-(\Upsilon^{41}/\Upsilon^{11})Z_1 \\ Z_6-(\Upsilon^{51}/\Upsilon^{11})Z_1 \\ Z_7-(\Upsilon^{61}/\Upsilon^{11})Z_1 \end{pmatrix} 1(Z_1<0),$$

where B is a Bernoulli random variable with success probability equal to half and independent of Z. In addition, the LR test statistic for $H_o:(\delta_o,\omega_o,\psi_o,\beta_o)'=(1,1,0,0)'$ is a mixture of $\chi^2(3)$ and $\chi^2(4)$ with mixing probabilities equal to half.

4. Monte Carlo Experiments

The foundation of our Monte Carlo experiments is the stationary data generating process (DGP) we discussed in section 3.1. For convenience, we write the DGP again here:

$$y_{i0}=\eta_i+q_{i0}; \quad y_{it}=\delta y_{i,t-1}+(1-\delta)\eta_i+\varepsilon_{it}, \quad t=1,2,...T,$$

where $\eta_i \sim N(0,\sigma_\eta^2)$, $q_{i0} \sim N(0,\sigma_q^2)$, and $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$. We set $var(\eta_i) = 2$, $var(q_{i0}) = 2$, $var(\varepsilon_i) = \nu = 1$, and $T = 5$. We try four different values, 0.5, 0.8, 0.9 and 1, for δ. Observe that when $\delta = 1$, the y_{it} are random walks without drifts. For each trial, we use 1000 iterations. We estimate the RE ($\hat{\delta}_{RE}$), HPT ($\hat{\delta}_{HPT}$), and Lancaster ML estimators ($\hat{\delta}_{Lan}$). We consider two cases, N = 500 and 100, to examine both the large-sample and small-sample properties of the estimators.

Table 1 reports the bias and MSE (mean square error) of the RE ML estimator of δ. We also report the finite-sample size and power properties of the t-tests based on the RE ML estimation. For all true values of δ, the biases of the ML estimator are small even when N = 100. Even when $\delta_o = 1$, the bias remains small, although MSE generally increases with δ_o. When δ_o is small, the t-test is properly sized. However, as δ_o increases, the test tends to over-reject correct hypotheses. Not surprisingly, the degree of the over-rejections of the t test is inversely related to the sample size. The power of the t test increases with N. But, when N = 100, the power of the t test to reject the unit-root hypothesis is generally low when $\delta_o \geq 0.8$. When $\delta_o = 1$, the t test rejects the correct hypothesis too much. This trend is not dependent on the sample size. This result is consistent with Proposition 6. The HPT estimators have the similar properties as the RE estimator. Table 1 report the results for the Lancaster estimator only for the cases with δ_o = 0.5 and δ_o = 0.8. For the cases with δ_o = 0.9, 1, we often fail to locate the maximum points.

Figures 3-4 show the finite-sample distributions of $(\hat{\delta}_{RE} - \delta_o)$. Figure 3 is for the case with N = 500 and Figure 4 is for the case with N = 100. The distribution of the RE estimator becomes wider as δ_o increases. Up to δ_o = 0.9, the estimator is roughly normally distributed. Then, when δ_o = 1, the estimator is no longer normally distributed. Its distribution has a hump near δ = 1 as Proposition 6 suggests.

Figures 5-12 compare the distributions of the three different ML estimators. Somewhat surprisingly, all of the estimators are similarly distributed. It appears that the efficiency gains of the RE estimator over the HPT estimator are not substantial.

We now examine the finite-sample property of the LR test for the hypothesis of unit root. The parametric restriction we test is given $(\delta, \omega, \psi)' = (1,1,0)'$. Figure 13-14 show the distributions of the LR test statistic when the unit root hypothesis is correct. The distribution of the LR statistic is compared to the distribution of $\chi^2(3)$ and the mixed distribution of $\chi^2(2)$ and $\chi^2(3)$. For both cases with N = 100 and N = 500, the distribution of the LR statistic is generally similar to the mixed distribution of $\chi^2(2)$ and $\chi^2(3)$. But when N = 100, the distribution of the LR statistic is more skewed to the left than the mixed distribution. This suggests that the LR test may under-reject the unit root hypothesis when N is small.

Table 4 supports this conjecture. Table 4 reports the rejection rates of the LR test based on three different distributional assumptions, $\chi^2(2)$, $\chi^2(3)$ and their mixture. For sensitivity analysis, we also report the results we obtain with many different combinations of the values of σ_η^2, $\sigma_{q_0}^2$, and ν. For our base choice of $\sigma_\eta^2 = 2$, $\sigma_{q_0}^2 = 2$, and $\nu = 1$, when N = 500 and the normal size is 5%, the rejection rates of the LR test with the $\chi^2(3)$, $\chi^2(2)$, and their mixed distributions are 3.66%, 8.02%, 5.10%. Thus, the LR test with the mixed distribution is reasonably well sized. In addition, the test has a power to reject unit root even when δ_o is close to one. Even when N = 100, the LR test with the mixed χ^2 distribution performs well. It tends to over-reject the unit root hypothesis, but the size of this distortion is small. The test also retains a good power property even when N is small.

Table 5 reports the finite-sample performances of the LR test for unit root based on the HPT estimator. The test performs well. The LR test based on the HPT estimator is generally better sized than the test based on the RE estimator, but only marginally. But in terms of power, the test based on the RE estimator is superior.

The general findings from our Monte Carlo experiments can be summarized as follows. First, the finite-sample performances of the RE, HPT and Lancaster estimators are quite similar when the true value of δ is small. But, the Lancaster estimation often fails to locate the maximum points when the true δ is close to one. Second, when the true value of δ is near one, the distributions of the RE and HPT estimators do not follow usual normal distributions even if the sample size is large. When the data follow the random walks without drifts, the Wald unit-root tests based on the ML estimators reject the correct hypothesis too often. In contrast, both the LR tests based on the RE and HPT estimators perform well. Even when N is small, they are sized properly and have power to reject the unit root hypothesis. Thus, use of the LR tests would be a useful alternative way to test unit roots.

5. Concluding Remark

This paper has considered the asymptotic and finite-sample properties of a random effects ML estimator and the two fixed-effects ML estimators of Lancaster (2002) and Hsiao, Pesaran and Tahmiscoglu (2002). The random effects ML estimator is asymptotically more efficient than the other two FE ML estimators when both the individual effects and the initial observations are normal. Nonetheless, the finite-sample performances are generally similar when data are not too much persistent. We also have considered the Wald-type tests for unit root based on the ML estimators. These tests reject the unit root hypothesis too often. In contrast, the alternative Likelihood-Ratio tests constructed following Rotnitzky, Cox, Bottai and Robins (2000) perform much better than the Wald-type tests. They have good power properties too, especially when the number of observations is large.

Appendix

Proof of Equation (12): Define $P_T = T^{-1}1_T 1_T'$, $Q_T = I_T - P_T$; and

$$D' = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}_{(T-1)\times T} .$$

Observe that $D'(y_{i1},...,y_{iT}) = \Delta y_i$. Since $D'\mathbf{1}_T = 0$, by Rao (1973, p. 77), $Q_T = DB_{T-1}^{-1}D'$. Also, observe that

$$\overline{\Delta_0 y_i} \equiv T^{-1}\Sigma_{t=1}^{T}(y_{it} - y_{i0}) = \Delta y_{i1} + k_T{}'\Delta y_i .$$

Similary, $\overline{\Delta_0 y_{i,-1}} = k_T{}'\Delta y_{i,-1}$. Using these results, we obtain

$$\begin{aligned}
&(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)'(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\quad = (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)' Q_T (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\qquad + (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)' P_T (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\quad = (y_i - y_{i,-1}\delta)' Q_T (y_i - y_{i,-1}\delta) + (\overline{\Delta_0 y_i} - \overline{\Delta_0 y_{i,-1}}\delta - p_i)^2 \\
&\quad = (\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1} (\Delta y_i - \Delta y_{i,-1}\delta) + (\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta) - p_i)^2 .
\end{aligned}$$

Proof of Proposition 1: A tedious but straight algebra shows

$$\begin{aligned}
&-\frac{T}{2\nu}\left(\left(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right) - p_i\right)^2 \\
&\quad = -\frac{\xi_{HPT,T}}{2\nu(\omega_{HPT}-1)}\left(p_i - \frac{T(\omega_{HPT}-1)}{\xi_{HPT,T}}\left(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)\right)^2 \\
&\qquad + \left(\frac{T^2(\omega_{HPT}-1)}{2\nu\xi_{HPT,T}} - \frac{T}{2\nu}\right)\left(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2 .
\end{aligned}$$

Using this result, we can have

$$\frac{1}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT}-1)^{1/2}}\exp\left[-\frac{T}{2v}\left(\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta\right)-p_i\right)^2-\frac{1}{2v(\omega_{HPT}-1)}p_i^2\right]$$

$$=(\xi_{HPT,T})^{-1/2}\frac{(\xi_{HPT,T})^{1/2}}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT}-1)^{1/2}}\exp\left[\begin{array}{l}-\dfrac{\xi_{HPT,T}}{2v(\omega_{HPT}-1)}\\ \times\left(p_i-\dfrac{T(\omega_{HPT}-1)}{\xi_{HPT,T}}\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta\right)\right)^2\end{array}\right]$$

$$\times\exp\left[-\frac{T}{2v\xi_{HPT,T}}\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta)\right)^2\right].$$

Thus,

$$\int_{-\infty}^{\infty}f_{FE,i}(\delta,v,p_i)f_i(p_i\,|\,\omega_{HPT},v)dp_i$$

$$=\frac{1}{(2\pi)^{T/2}v^{T/2}(\xi_{HPT,T})^{1/2}}\exp\left[\begin{array}{l}-\dfrac{1}{2v}(\Delta y_i-\Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i-\Delta y_{i,t-1}\delta)\\ -\dfrac{T}{2v\xi_{HPT,T}}\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta)\right)^2\end{array}\right]$$

$$\times\int_{-\infty}^{\infty}\frac{(\xi_{HPT,T})^{1/2}}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT}-1)^{1/2}}\exp\left[\begin{array}{l}-\dfrac{\xi_{HPT,T}}{2v(\omega_{HPT}-1)}\\ \times\left(\begin{array}{l}p_i\\ -\dfrac{T(\omega_{HPT}-1)}{\xi_{HPT,T}}\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta)\right)\end{array}\right)^2\end{array}\right]dr_i$$

$$=\frac{1}{(2\pi)^{T/2}v^{T/2}(\xi_{HPT,T})^{1/2}}\exp\left[\begin{array}{l}-\dfrac{1}{2v}(\Delta y_i-\Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i-\Delta y_{i,t-1}\delta)\\ -\dfrac{T}{2v\xi_{HPT,T}}\left(\Delta y_{i1}+k_{T-1}{}'(\Delta y_i-\Delta y_{i,-1}\delta)\right)^2\end{array}\right]$$

$$=f_{HPT,i}(r_i\,|\,\theta_{HPT}),$$

where $r_i=(\Delta y_{i1},\Delta y_{i2},...\Delta y_{iT})'$.

The following lemmas are useful to prove the propositions in section 3.

Lemma A.1: Define $m_{T-1}=(1/T,2/T,...,(T-1)/T)'$, and

$$
\tilde{D}'_{T-1} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ : & : & : & & : \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}_{(T-1)\times(T-1)} ,
$$

which equals D′ with its last column excluded. Then,

$$
B_{T-1}^{-1} = \left(\tilde{D}_{T-1}\right)^{-1}\left(\tilde{D}'_{T-1}\right)^{-1} - Tm_{T-1}m'_{T-1} .
$$

Lemma A.2: $c'_{T-1}B_{T-1}^{-1}c_{T-1} = (T-1)/T$.

Lemma A.3: $\Omega_{T-1} = (\omega-1)c_{T-1}c_{T-1}{}' + \tilde{D}_{T-1}\tilde{D}'_{T-1}$.

Lemma A.4: The first-order and second-order derivatives of $\ell_{RE,i}(\theta)$ are given:

$$
\ell_{RE,i,\delta} = \frac{1}{\nu}\Delta y_{i,-1}{}'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) + \frac{T}{\nu\xi_T}\Delta y_{i,-1}{}'k_{T-1}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right);
$$

$$
\ell_{RE,i,\omega} = -\frac{T}{2\xi_T} + \frac{T^2}{2\nu\xi_T{}^2}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2;
$$

$$
\begin{aligned}
\ell_{RE,i,\nu} = {} & -\frac{T}{2\nu} + \frac{1}{2\nu^2}\left(\Delta y_i - \Delta y_{i,-1}\delta\right)B_{T-1}^{-1}\left(\Delta y_i - \Delta y_{i,-1}\delta\right) \\
& + \frac{T}{2\nu^2\xi_T}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \delta\Delta y_{i,-1})\right)^2;
\end{aligned}
$$

$$
\ell_{RE,i,\psi} = \frac{T}{\nu\xi_T}y_{i0}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \delta\Delta y_{i,-1})\right);
$$

$$
\ell_{RE,i,\delta\delta} = -\frac{1}{\nu}\Delta y_{i,-1}{}'B_{T-1}^{-1}\Delta y_{i,-1} - \frac{T}{\nu\xi_T}\Delta y_{i,-1}{}'k_{T-1}k_{T-1}{}'\Delta y_{i,-1};
$$

$$
\ell_{RE,i,\delta\nu} = -\frac{1}{\nu^2}\Delta y_{i,-1}{}'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) - \frac{T}{\nu^2\xi_T}\Delta y_{i,-1}{}'k_{T-1}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right);
$$

$$
\ell_{RE,i,\delta\omega} = -\frac{T^2}{\nu\xi_T{}^2}\Delta y_{i,-1}{}'k_{T-1}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right);
$$

$$\ell_{RE,i,\delta\psi} = -\frac{T}{\nu\xi_T}\Delta y_{i,-1}{}'k_{T-1}y_{i0};$$

$$\ell_{RE,i,\nu\nu} = \frac{T}{2\nu^2} - \frac{1}{\nu^3}(\Delta y_i - \Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta)$$
$$-\frac{T}{\nu^3\xi_T}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2;$$

$$\ell_{RE,i,\nu\omega} = -\frac{T^2}{2\nu^2\xi_T{}^2}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2;$$

$$\ell_{RE,i,\nu\psi} = -\frac{T}{\nu^2\xi_T}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)y_{i0};$$

$$\ell_{RE,i,\omega\omega} = \frac{T^2}{2\xi_T{}^2} - \frac{T^3}{\nu\xi_T{}^3}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2;$$

$$\ell_{RE,i,\omega\psi} = -\frac{T^2}{\nu\xi_T{}^2}(\Delta y_{i1} - \psi y_{i0})\,y_{i0};$$

$$\ell_{RE,i,\psi\psi} = -\frac{T}{\nu\xi_T}(y_{i0})^2.$$

Lemma A.5: Under the RE assumption,

$$E\left(\Delta y_{i,-1}B_{T-1}^{-1}\Delta y_{i,-1}\right) = \frac{T-1}{T}\psi^2\sigma_0^2 + \nu \times tr\left(B_{T-1}^{-1}C_{T-1}{}'\Omega_{T-1}C_{T-1}\right); \tag{A.1}$$

$$E\left(\Delta y_{i,-1}{}'k_T k_T{}'\Delta y_{i,-1}\right) = \left(\frac{T-1}{T}\right)^2\psi^2\sigma_0^2 + \nu \times tr\left(k_T{}'C_{T-1}{}'\Omega_{T-1}C_{T-1}k_T\right); \tag{A.2}$$

$$E\left(\Delta y_{i,-1}B_{T-1}^{-1}\Delta\varepsilon_i\right) = -\nu b(\delta)'; \tag{A.3}$$

$$E\left(\Delta y_{i,-1}{}'k_{T-1}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta)\right)\right) = \frac{\nu\xi_T}{T}b(\delta)', \tag{A.4}$$

where b(δ) is defined in (13), $b(\delta)' = db/d\delta$, and

$$C_{T-1}' = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \delta & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \delta^{T-3} & \delta^{T-4} & \delta^{T-5} & \dots & 0 \\ \delta^{T-2} & \delta^{T-3} & \delta^{T-4} & \dots & 1 \end{pmatrix}_{(T-1)\times(T-1)}.$$

Proof: Define $\tilde{h}_i = (u_i, \Delta\varepsilon_{i1}, ..., \Delta\varepsilon_{i,T-1})$, so that $Var(\tilde{h}_i) = \Omega_{T-1}$, and

$$\Delta y_{i,-1} = c_{T-1}\psi y_{i0} + C'_{T-1}\tilde{h}_i .$$

Then,

$$\begin{aligned} E\left(\Delta y_{i,-1} B_{T-1}^{-1} \Delta y_{i,-1}\right) &= \psi^2\sigma_0^2 c_{T-1}{}' B_{T-1}^{-1} c_{T-1} + E\left(\tilde{h}_i' C_{T-1} B_{T-1}^{-1} C'_{T-1}\tilde{h}_i\right) \\ &= \frac{T-1}{T}\psi^2\sigma_0^2 + tr\left(E\left(C_{T-1}B_{T-1}^{-1}C'_{T-1}\tilde{h}_i\tilde{h}_i{}'\right)\right) \\ &= \frac{T-1}{T}\psi^2\sigma_0^2 + \nu \times tr\left(B_{T-1}^{-1}C_{T-1}{}'\Omega_{T-1}C_{T-1}\right). \end{aligned}$$

(A.2)-(A.4) can be obtained by the similar method.

Lemma A.6: Under the RE assumption,

$$E\left(-H_{RE,i}(\theta_o)\right) = \begin{pmatrix} \frac{(T-1)\omega_o}{\xi_{T,o}}\frac{\psi_o{}^2\sigma_0^2}{\nu} + A & 0 & \frac{Tb(\delta_o)'}{\xi_{T,o}} & (T-1)\frac{\psi_o\sigma_0^2}{\nu_o\xi_{T,o}} \\ 0 & \frac{T}{2\nu_o{}^2} & \frac{T}{2\nu\xi_{T,o}} & 0 \\ \frac{Tb(\delta_o)'}{\xi_{T,o}} & \frac{T}{2\nu_o\xi_{T,o}} & \frac{T^2}{2\xi_{T,o}^2} & 0 \\ (T-1)\frac{\psi\sigma_0^2}{\nu_o\xi_{T,o}} & 0 & 0 & \frac{T\sigma_0^2}{\nu_o\xi_{T,o}} \end{pmatrix}, \qquad (A.5)$$

where $A = tr\left(B_{T-1}^{-1}C_{T-1}{}'\Omega_{T-1}C_{T-1}\right) + \frac{T}{\xi_T} tr\left(k_T{}'C_{T-1}{}'\Omega_{T-1}C_{T-1}k_T\right)$.

Proof of Proposition 1: At $\theta = \theta_* = (1, \nu_*, 1, 0)'$, Lemma A.3 and the definition of C_{T-1} in Lemma A.5 imply:

$$\Omega_{T-1} = \tilde{D}_{T-1}\tilde{D}'_{T-1}; \;\; C_{T-1} = \left(\tilde{D}'_{T-1}\right)^{-1}.$$

Thus, $C'_{T-1}\Omega_{T-1}C_{T-1} = T_{T-1}$. In addition, by Lemma 1,

$$tr\left(B_{T-1}^{-1}\right) = tr\left(\left(\tilde{D}_{T-1}\right)^{-1}\left(\tilde{D}'_{T-1}\right)^{-1}\right) - T \times tr\left(m_{T-1}m'_{T-1}\right) = \frac{T(T-1)}{2} - \frac{(T-1)(2T-1)}{6};$$

$$tr\left(k_{T-1}k_{T-1}{}'\right) = tr\left(k_{T-1}{}'k_{T-1}\right) = \frac{(T-1)(2T-1)}{6T}.$$

At $\theta = (1,\nu,1,0)'$, $\xi_T = 1$. Thus, A = T(T-1)/2 = Tb(1)′. Substituting this result, $\psi = 0$, and $\xi_T = 1$, into (A.5), we have:

$$E\left(-H_{RE,i}(\theta_*)\right) = \begin{pmatrix} Tb(1)' & 0 & Tb(1)' & 0 \\ 0 & \dfrac{T}{2\nu^2} & \dfrac{T}{2\nu} & 0 \\ Tb(1)' & \dfrac{T}{2\nu} & \dfrac{T^2}{2} & 0 \\ 0 & 0 & 0 & \dfrac{T\sigma_0^2}{\nu} \end{pmatrix},$$

which has a zero determinant. The likelihood theory indicates $E\left(H_{RE,i}(\theta_*) + B_{RE,i}(\theta_*)\right) = 0$. Thus, $E\left(B_{RE,i}(\theta_*)\right)$ must be also singular.

Lemma A.7: Under the RE assumption,

$$E\left(-H_{HPT,i}(\delta_o,\nu_o,\omega_o)\right) \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} A & 0 & \dfrac{Tb(\delta_o)'}{\xi_{T,o}} \\ 0 & \dfrac{T}{2\nu_o{}^2} & \dfrac{T}{2\nu_o\xi_{T,o}} \\ \dfrac{Tb(\delta_o)'}{\xi_{T,o}} & \dfrac{T}{2\nu\xi_{T,o}} & \dfrac{T^2}{2\xi_{T,o}^2} \end{pmatrix};$$

$$E\left(-H_{Lan,i}(\delta_o,\nu_o)\right) = \begin{pmatrix} -b(\delta_o)'' + A_2 & \dfrac{-b(\delta_o)'}{\nu_o} \\ \dfrac{-b(\delta_o)'}{\nu_o} & \dfrac{(T-1)}{2\nu_o{}^2} \end{pmatrix};$$

$$E\left(B_{Lan,i}(\delta_o,\nu_o)\right) = \begin{pmatrix} (b(\delta_o)')^2 - 2b(\delta_o)'' + A_2 & \dfrac{-b(\delta_o)'}{\nu_o} \\ \dfrac{-b(\delta_o)'}{\nu_o} & \dfrac{(T-1)}{2\nu_o{}^2} \end{pmatrix}$$

where $M_{11} = \mathrm{diag}(A, T/2\nu_o^2)$ and

$$A = A_1 + A_2 = tr\left(B_{T-1}^{-1}C_{T-1}{}'\Omega_{T-1}C_{T-1}\right) + (T/\xi_T)tr\left(k_T{}'C_{T-1}{}'\Omega_{T-1}C_{T-1}k_T\right).$$

When T = 2,

$$M_{11} - M_{12}(M_{22})^{-1}M_{21} = E\left(-H_{Lan,i}(\theta_{Lan,o})\right)\left[E\left(B_{Lan,i}(\theta_{Lan,o})\right)\right]^{-1}E\left(-H_{Lan,i}(\theta_{Lan,o})\right).$$

This implies that the HPT and Lancaster estimators of $(\delta,\nu)'$ are asymptotically equally efficient.

Proof of Proposition 3: We first derive $E(\ell_{RE,i}(\theta)\,|\,\theta_o = \theta_*)$. We can show that at $\theta_o = \theta_*$,

$$\begin{aligned} &E\left(\left(\Delta y_{i1} - \psi y_{io} + k'_{T-1}(\Delta y_i - \Delta y_{i,-1}\delta)\right)^2\right) \\ &\qquad = \nu_* \frac{(T+1)(2T+1)}{6T} - 2\nu_* \frac{(T-1)(T+1)}{3T}\delta + \sigma_o^2\pi^2 + \nu_* \frac{(T-1)(2T-1)}{6T}\delta^2; \\ &E\left((\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta)\right) \\ &\qquad = \frac{\nu_*}{6}\left((T-1)(T+1) - 2(T-1)(T-2)\delta + (T-1)(T+1)\delta^2\right). \end{aligned}$$

Thus, we have

$$\begin{aligned} m(\delta,\nu,\omega,\psi) &\equiv E\left(\ell_{RE,i}(\theta)\,|\,\theta_o = \theta_*\right) \\ &= -\frac{T}{2}\ln(\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln\xi_T \\ &\quad - \frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1) - 2(T-1)(T-2)\delta + (T-1)(T+1)\delta^2\right) \\ &\quad - \frac{1}{12}\frac{\nu_*}{\nu\xi_T}\begin{pmatrix}(T+1)(2T+1) - 4\nu_*(T-1)(T+1)\delta \\ +\sigma_o^2\psi^2 + (T-1)(2T-1)\delta^2\end{pmatrix}. \end{aligned}$$

We now concentrate out ψ, ω, and ν from m(δ,ν,ω,ψ). Clearly, ψ = 0 maximizes m(δ,ν,ω,ψ). Thus the concentrated value of m becomes:

$$\begin{aligned} m^c(\delta,\nu,\omega) &= -\frac{T}{2}\ln(\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln\xi_T \\ &\quad - \frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1) - 2(T-1)(T-2)\delta + (T-1)(T+1)\delta^2\right) \\ &\quad - \frac{1}{12}\frac{\nu_*}{\nu\xi_T}E\left((T+1)(2T+1) - 4\nu_*(T-1)(T+1)\delta + (T-1)(2T-1)\delta^2\right). \end{aligned}$$

Now, solve the first-order condition with respect to ξ_T (instead of ω):

$$\frac{\partial m^c\left(\delta,\nu,\omega\right)}{\partial \xi_T} = -\frac{1}{2\xi} + \frac{1}{12}\frac{\nu_*}{\nu {\xi_T}^2}\begin{pmatrix}(T+1)(2T+1)-4\nu_*(T-1)(T+1)\delta \\ +(T-1)(2T-1)\delta^2\end{pmatrix} = 0\,.$$

Then we obtain:

$$\xi_T = \frac{1}{6}\frac{\nu_*}{\nu}\left((T+1)(2T+1)-4\nu_*(T-1)(T+1)\delta+(T-1)(2T-1)\delta^2\right).$$

Thus,

$$\begin{aligned} m^c\left(\delta,\nu\right) = &-\frac{T}{2}\ln(\pi)-\frac{T}{2}\ln(\nu)-\frac{1}{2}\ln\left(\frac{\nu_*}{\nu}\right)-\frac{1}{2} \\ &-\frac{1}{2}\ln\left(\frac{(T+1)(2T+1)}{6}-2\frac{(T-1)(T+1)}{3}\delta+\frac{(T-1)(2T-1)}{6}\delta^2\right) \\ &-\frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right). \end{aligned}$$

Now, solving

$$\frac{\partial m^c\left(\delta,\nu\right)}{\partial \nu} = -\frac{T-1}{2\nu}+\frac{1}{12}\frac{\nu_*}{\nu^2}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right)=0\,,$$

we have:

$$\nu = \nu_*\frac{1}{T-1}\left(\frac{(T-1)(T+1)}{6}-2\frac{(T-1)(T-2)}{6}\delta+\frac{(T-1)(T+1)}{6}\delta^2\right).$$

Substituting this solution to $m^c(\delta,\nu)$ yields

$$\begin{aligned} m^c\left(\delta\right) = &-\frac{T}{2}\ln(\pi)-\frac{T}{2}\ln(\nu_*)-\frac{T-1}{2}\ln\left(\frac{1}{T-1}\right)-\frac{T}{2} \\ &-\frac{T-1}{2}\ln\left(\frac{(T-1)(T+1)}{6}-2\frac{(T-1)(T-2)}{6}\delta+\frac{(T-1)(T+1)}{6}\delta^2\right) \\ &-\frac{1}{2}\ln\left(\frac{(T+1)(2T+1)}{6}-2\frac{(T-1)(T+1)}{3}\delta+\frac{(T-1)(2T-1)}{6}\delta^2\right). \end{aligned}$$

Then, a little algebra shows:

$$\frac{\partial m^c(\delta)}{\partial\delta} = -\frac{(T-1)\left(2+\delta+(T-1)\delta\right)}{1-2\delta(T-2)+T+\delta^2(T+1)}$$
$$-\frac{(T-1)\left(\delta(2T-1)-2(T+1)\right)}{(1+T)(1+2T)-4\delta\left(T^2-1\right)+\delta^2(1-T)(1-2T)}$$
$$= -\frac{(\delta-1)^3(T-1)T(T+1)(2T-1)}{\left(1+4\delta+\delta^2+(\delta-1)T\left(2T(\delta-1)-3(\delta+1)\right)\right)\left(1+4\delta+\delta^2+(\delta-1)^2 T\right)}$$
$$= -\frac{(\delta-1)^3(T-1)T(T+1)(2T-1)}{(T-1)\left[(2T-1)\left\{\delta-\frac{2(T+1)}{2T-1}\right\}^2+\frac{3}{2T-1}\right]\left[(T+1)\left\{\delta-\frac{T-2}{T+1}\right\}^2+\frac{3(T-1)}{T+2}\right]}.$$

Since the denominator of $\partial m^c(\delta)/\partial\delta$ is positive for any $T \geq 2$, it is positive for $\delta < 1$ and negative for $\delta > 1$. The derivative equals zero only at $\delta = 1$. This indicates that $\delta = 1$ is the global maximum point.

Proof of Proposition 4: At $\theta_o = \theta_*$, it can be shown that

$$E\left(\ell_{Lan,i}(\delta,\nu)\right) = b(\delta) - \frac{T-1}{2}\ln(\nu)$$
$$-\frac{\nu_*}{2\nu}\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

Without loss of generality we set $\nu_* = 1$. Then, the first-order maximization condition with respect to ν yields

$$\nu = \frac{1}{T-1}\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

Substituting this into the expected value of $E(\ell_{Lan,i}(\delta,\nu))$, we can get

$$E\left(\ell^c_{Lan,i}(\delta)\right) = b(\delta) - \frac{T-1}{2}\ln\left(\frac{1}{T-1}\right) - \frac{T-1}{2}$$
$$-\frac{T-1}{2}\ln\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

If we differentiate the concentrated likelihood function, we yield

$$\frac{\partial E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial\delta} = b(\delta)' - \frac{(T-1)\left((T+1)\delta - T(T-2)\right)}{(T+1)\delta^2 - 2T(T-2)\delta + (T+1)};$$

$$\frac{\partial^2 E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial\delta^2} = b(\delta)'' - \frac{2\left(T-1\right)\left(T\left(\delta-1\right)+\delta+2\right)^2}{\left(\left(T\left(\delta-2\right)+4+\delta\right)\delta+\left(T+1\right)\right)^2} + \frac{\left(T-1\right)\left(T+1\right)}{\left(\left(T\left(\delta-2\right)+4+\delta\right)\delta+\left(T+1\right)\right)}.$$

$$\frac{\partial^3 E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial\delta^3} = b(\delta)''' + \frac{8\left(T-1\right)\left(T\left(\delta-1\right)+\delta+2\right)^3}{\left(\left(T\left(\delta-2\right)+4+\delta\right)\delta+\left(T+1\right)\right)^3} - \frac{6\left(T-1\right)\left(T+1\right)\left(T\left(\delta-1\right)+\delta+2\right)}{\left(\left(T+1\right)\delta^2-2\left(T-2\right)\delta+\left(T+1\right)\right)^2}.$$

It can be shown that when δ = 1,

$$b(1)' = \frac{T-1}{2}; b(1)'' = \frac{(T-1)(T-2)}{6}; b(1)''' = \frac{(T-1)(T-2)(T-3)}{12}.$$

Using these results, we can easily show that at δ = 1,

$$\frac{\partial E\left(\ell_{Lan,i}(\delta)\right)}{\partial\delta} = \frac{\partial^2 E\left(\ell_{Lan,i}(\delta)\right)}{\partial\delta^2} = 0,$$

but,

$$\left.\frac{\partial^3 E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial\delta^3}\right|_{\delta=1} = \frac{\left(T-1\right)\left(T-2\right)\left(T-3\right)}{12} - \frac{\left(T-1\right)\left(T+1\right)}{2} \neq 0.$$

Thus, δ = 1 is an inflexion point of $E\left(\ell^c_{Lan,i}(\delta)\right)$.

Lemma A.7: Define:

$$L_T{}' = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}_{T\times T} ; \quad H_T = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{T\times T} ;$$

$$S_{1,T}{}' = \begin{pmatrix} I_{T-1} & 0_{(T-1)\times 1} \end{pmatrix}_{(T-1)\times T}; \quad S_{2,T}{}' = \begin{pmatrix} 0_{(T-1)\times 1} & 0 \\ I_{T-1} & 0_{(T-1)\times 1} \end{pmatrix}.$$

Then,

$$S_{1,T}{}'B_T^{-1}D_T{}' + TS_{2,T}{}'\bar{k}_T\bar{k}_T{}'L_T{}' = H_T\ ; L_T\bar{k}_T\bar{k}_T{}'L_T{}' = \frac{1}{T^2}1_T1_T{}'; D_TB_{T-1}^{-1}D_T' = I_T - \frac{1}{T}1_T1_T{}',$$

$$B_T^{-1} = \left(\tilde{D}_T{}'\tilde{D}_T\right)^{-1} - Tm_Tm_T{}'; S_{2,T}k_Tk_T{}'L_T{}' = \frac{1}{T}\begin{pmatrix} \bar{k}_T 1_{T+1}{}' \\ 0_{1\times T} \end{pmatrix},$$

where $\bar{k}_T = (1, k_T')'$.

Proof of Proposition 5: Define $r_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$. At $\theta = \theta_*$, using Lemma A.7, we can have

$$\begin{aligned}\ell_{RE,i,\omega} &= -\frac{T}{2} + \frac{T^2}{2\nu}\left(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1})\right)^2 = -\frac{T}{2} + \frac{T^2}{2\nu}\left(\bar{k}_T{}'L_T{}'r_i\right)^2 \\ &= -\frac{1}{2} + \frac{T^2}{2\nu}r_i'L_T\bar{k}_T\bar{k}_T{}'L_T{}'r_i = -\frac{NT}{2} + \frac{1}{2\nu}\Sigma_i r_i'1_T1_T{}'r_i.\end{aligned}$$

$$\begin{aligned}\ell_{NT,\nu} &= -\frac{T}{2\nu} + \frac{1}{2\nu^2}(\Delta y_i - \Delta y_{i,-1})'B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}) + \frac{T}{2\nu^2}(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}))^2 \\ &= -\frac{T}{2\nu} + \frac{1}{2\nu^2}r_i'D_TB_{T-1}^{-1}D_T{}'r_i + \frac{T}{2\nu^2}r_i'\frac{1}{T^2}1_T1_T{}'r_i \\ &= -\frac{T}{2\nu} + \frac{1}{2\nu^2}\Sigma_i r_i'\left(I_T - \frac{1}{T}1_T1_T{}'\right)r_i + \frac{1}{2T\nu^2}\Sigma_i r_i'1_T1_T{}'r_i \\ &= -\frac{T}{2\nu} + \frac{1}{2\nu^2}\Sigma_i r_i'r_i,\end{aligned}$$

where ν = ν*. Thus,

$$\ell_{NT,\omega} - \nu\ell_{NT,\nu} = -\frac{1}{2\nu}r_i'(I_T - 1_{T+1}1_{T+1}{}')r_i = \frac{1}{\nu}r_i'H_Tr_i.$$

Now,

$$\begin{aligned}\ell_{RE,i,\delta} &= \frac{1}{\nu}\Delta y_{i,-1}B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}) + \frac{T}{\nu}\Delta y_{i,-1}{}'k_{T-1}(\Delta y_{i1} - k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1})) \\ &= \frac{1}{\nu}r_i'S_{1,T}B_{T-1}^{-1}D_T{}'r_i + \frac{T}{\nu}r_i'S_{2,T}\bar{k}_T\bar{k}_T{}'L_T{}'r_i = \frac{1}{\nu}r_i'H_Tr_i.\end{aligned}$$

Thus, $\ell_{NT,\delta} - \ell_{NT,\omega} + \nu\ell_{NT,\nu} = 0$.

Proof of Proposition 6: We first check whether (i) whether $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ equals zero or is linearly related to $\tilde{\ell}_{RE,i,\delta}(\theta_*)$, $\tilde{\ell}_{RE,i,\nu_n}(\theta_*)$, $\tilde{\ell}_{RE,i,\omega_n}(\theta_*)$, and $\tilde{\ell}_{RE,i,\psi_n}(\theta_*)$. Note that

$$\tilde{\ell}_{RE,i,\nu_n}(\theta_*) = \ell_{RE,i,\nu}(\theta_*)\,;\;\; \tilde{\ell}_{RE,i,\omega_n}(\theta_*) = \ell_{RE,i,\omega}(\theta_*)\,;\;\; \tilde{\ell}_{RE,i,\psi}(\theta_*) = \ell_{RE,i,\psi}(\theta_*)\,.$$

Since $\tilde{\ell}_{RE,i}(\delta,\nu_n,\omega_n,\psi) = \ell_{RE,i}(\delta,\nu(\nu_n,\delta),\omega(\omega_n,\delta),\psi)$, we have:

$$\tilde{\ell}_{RE,i,\delta} = \ell_{RE,i,\delta} - \ell_{RE,i,\omega} + \nu_n \ell_{RE,i,\nu}\,;$$

$$\begin{aligned}
\tilde{\ell}_{RE,i,\delta\delta} &= \ell_{RE,i,\delta\delta} - \ell_{RE,i,\delta\omega} + \nu_n \ell_{RE,i,\delta\nu} - \ell_{RE,i,\omega\delta} + \ell_{RE,i,\omega\omega} - \nu_n \ell_{RE,i,\omega\nu} \\
&\quad + \nu_n \left(\ell_{RE,i,\nu\delta} - \ell_{RE,i,\nu\omega} + \nu_n \ell_{RE,i,\nu\nu} \right) \\
&= \left(\ell_{RE,i,\delta\delta} - 2\ell_{RE,i,\delta\omega} + \ell_{RE,i,\omega\omega} \right) + 2\nu_n \left(\ell_{RE,i,\delta\nu} - \ell_{RE,i,\omega\nu} \right) + \nu_n^2 \ell_{RE,i,\nu\nu}.
\end{aligned}$$

Then, using Lemma A.7 and a little algebra, we can show that at $\theta_n = \theta_*$,

$$\begin{aligned}
\tilde{\ell}_{RE,i,\delta\delta}(\theta_*) &= \left(\ell_{RE,i,\delta\delta}(\theta_*) - 2\ell_{RE,i,\delta\omega}(\theta_*) + \ell_{RE,i,\omega\omega}(\theta_*) \right) \\
&\quad + 2\nu_* \left(\ell_{RE,i,\delta\nu}(\theta_*) - \ell_{RE,\omega\nu}(\theta_*) \right) + \nu_*^2 \ell_{RE,i,\nu\nu}(\theta_*) \\
&= \begin{pmatrix} -\dfrac{1}{\nu_*}\Delta y_{i,-1} B_{T-1}^{-1} \Delta y_{i,-1} - \dfrac{T}{\nu_*}\Delta y_{i,-1}' k_{T-1} k_{T-1}' \Delta y_{i,-1} \\ + \dfrac{2T^2}{\nu_*}\Delta y_{i,-1}' k_{T-1}(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1})) - \dfrac{T^3}{\nu_*}\Big(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}) \Big)^2 \end{pmatrix} \\
&\quad + \frac{T(T+1)}{2} \\
&= \left(-\frac{1}{\nu_*}\Delta y_{i,-1}' \left(\tilde{D}_T' \tilde{D}_T \right)^{-1} \Delta y_{i,-1} - \frac{2T}{\nu_*}\Delta y_{i,-1}' m_{T-1}(\Delta y_{iT}) - \frac{T}{\nu_*}(\Delta y_{iT})^2 \right) + \frac{T(T+1)}{2},
\end{aligned}$$

which is neither zero, nor a linear combination of $\tilde{\ell}_{RE,i,\nu_n}(\theta_*)$, $\tilde{\ell}_{RE,i,\omega_n}(\theta_*)$ and $\tilde{\ell}_{RE,i,\psi}(\theta_*)$.

We now check (ii) whether or not $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is a linear combination of $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$, $\tilde{\ell}_{RE,i,\nu_n}(\theta_*)$, $\tilde{\ell}_{RE,i,\omega_n}(\theta_*)$ and $\tilde{\ell}_{RE,i,\psi}(\theta_*)$. We can show

$$\begin{aligned}
\tilde{\ell}_{RE,i,\delta\delta\delta} &= \left(\ell_{RE,i,\delta\delta\delta} - 3\ell_{RE,,\delta\delta\omega} + 3\ell_{RE,i,\omega\omega\delta} - \ell_{RE,i,\omega\omega\omega} \right) \\
&\quad + 3\nu_n \left(\ell_{RE,.i,\delta\delta\nu} - 2\ell_{RE,i,\delta\omega\nu} + \ell_{RE,i,\omega\omega\nu} \right) + 3\nu_n{}^2 \left(\ell_{RE,i,\delta\nu\nu} - \ell_{RE,i,\omega\nu\nu} \right) + \nu_n^3 \left(\ell_{RE,i,\nu\nu\nu} \right).
\end{aligned}$$

But, at $\theta = \theta_*$,

$$\ell_{RE,i,\delta\delta\delta}(\theta_*) = 0\,;\;\; \ell_{RE,i,\delta\delta\omega} = \frac{T^2}{\nu_*}\Delta y_{i,-1}' k_{T-1} k_{T-1}' \Delta y_{i,-1}\,;$$

$$\ell_{RE,i,\omega\omega\delta}(\theta_*) = \frac{2T^3}{\nu_*}\Delta y_{i,-1}' k_{T-1}\left(\Delta y_{i1} + k_T'(\Delta y_i - \Delta y_{i,-1}) \right);$$

$$\ell_{RE,i,\omega\omega\omega}(\theta_*) = -T^3 + \frac{3T^4}{v_*}(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}))^2 ;$$

$$\ell_{RE,i,\delta\delta v}(\theta_*) = -\frac{1}{v_*}\ell_{RE,i,\delta\delta}(\theta_*) ;\ \ \ell_{RE,i,\delta\omega v}(\theta_*) = -\frac{1}{v_*}\ell_{RE,i,\delta\omega}(\theta_*) ;$$

$$\ell_{RE,i,\omega\omega,v}(\theta_*) = -\frac{1}{v_*}\ell_{RE,i,\omega\omega}(\theta_*) + \frac{T^2}{2v_*} ;\ \ \ell_{RE,i,\delta vv}(\theta_*) = -\frac{2}{v_*}\ell_{RE,i,\delta v}(\theta_*) ;$$

$$\ell_{RE,i,vv\omega}(\theta_*) = -\frac{2}{v_*}\ell_{RE,i,v\omega}(\theta_*) ;\ \ \ell_{RE,i,vvv}(\theta_*) = -\frac{3}{v_*}\ell_{RE,i,vv}(\theta_*) + \frac{T}{2{v_*}^3} .$$

Using these results, we can show:

$$\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*) = \frac{2T^3 + 3T^2 + T}{2} - \frac{3T^2}{v_*}\Delta y_{i,-1}' m_{T-1} m_{T-1}' \Delta y_{i,-1}$$
$$-\frac{6T^2}{v_*}(m_{T-1}'\Delta y_{i,-1})\Delta y_{iT} - \frac{3T^2}{v_*}(\Delta y_{iT})^2 - 3\tilde{\ell}_{RE,i,\delta\delta}(\theta_*),$$

which is neither zero, nor a linear combination of $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$, $\tilde{\ell}_{RE,i,v_n}(\theta_*)$, $\tilde{\ell}_{RE,i,\omega_n}(\theta_*)$ and $\tilde{\ell}_{RE,i,\psi}(\theta^*)$. For example, when T = 2:

$$\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*) = 15 - \frac{3}{v_*}\left(\Delta y_{i1}\right)^2 - \frac{12}{v_*}\Delta y_{i1}\Delta y_{i2} - \frac{12}{v_*}(\Delta y_{i2})^2 - 3\tilde{\ell}_{RE,i,\delta\delta}(\theta_*) ;$$

$$\tilde{\ell}_{RE,i,\delta\delta}(\theta_*) = \left(-\frac{1}{v_*}\left(\Delta y_{i1}\right)^2 - \frac{2}{v_*}\Delta y_{i1}\Delta y_{i2} - \frac{2}{v_*}(\Delta y_{i2})^2 \right) + 3 ;$$

$$\tilde{\ell}_{RE,i,v_n}(\theta_*) = \ell_{RE,i,v}(\theta_*) = -\frac{1}{v_*} + \frac{1}{2{v_*}^2}\left((\Delta y_{i1})^2 + (\Delta y_{i2})^2\right)$$

$$\tilde{\ell}_{RE,i,\omega_n}(\theta_*) = \ell_{RE,i,\omega}(\theta_*) = -1 + \frac{1}{2v_*}(\Delta y_{i1} + \Delta y_{i2})^2 ;$$

$$\tilde{\ell}_{RE,i,\psi_n}(\theta_*) = \ell_{RE,i,\psi}(\theta_*) = \frac{1}{v_*} y_{i0}\left(\Delta y_{i1} + \Delta y_{i2}\right) .$$

All of these terms are linearly independent. Thus, Theorem 3 of RCBR implies our result.

Proofs of Propositions 7 and 8: It is easy to see that $(\ell^m_{RE,i,\delta}(\phi_*) - \ell^m_{RE,i,\beta}(\phi_*))$, $\ell^m_{RE,i,v}(\phi_*)$, and $\ell^m_{RE,i,\omega}(\phi_*)$ are the same as $\ell_{RE,i,\delta}(\theta_*)$, $\ell_{RE,i,v}(\theta_*)$, and $\ell_{RE,i,\omega}(\theta_*)$ with the Δy_{it} replaced by $\Delta y_{it} - a$. Thus, we can obtain the results by the same way that we used for Propositions 5 and 6.

REFERENCES

Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of Econometrics, 68, 5-27.

Ahn, S. C. and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: alternative assumptions and simplified assumptions, Journal of Econometrics 76, 309-321.

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

Arellano, M. and S. Bond, 1991, Tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58, 277-297.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variables estimation of error-component models, Journal of Econometrics 68, 29-51.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica 51, 1635-1660/

Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics 87, 115-143.

Bond, S. and F. Windmeijer, 2002, Finite sample inference for GMM estimators in linear panel data models, IFS, mimeo.

Cox, D. R., and N. Reid, 1987, Parameter orthogonally and approximate conditional inference (with discussion), Journal of the Royal Statistical Society, B, 49 (1), 1-39.

Hahn, J., 1999, How informative is the initial condition in the dynamic panel data model with fixed effects?, Journal of Econometrics 93, 309-326.

Hsiao, C.,. 1986, Analysis of Panel Data, Cambridge, University Press, Cambridge, UK.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscoglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time period, Journal of Econometrics 2002, 107-150.

Kruiniger, H., 2002a, On the estimation of panel regression models with fixed effects, Queen Mary, University of London, mimeo.

Kruiniger, H., 2002b, Maximum likelihood estimation of dynamic linear panel data models with fixed effects, Queen Mary, University of London, mimeo.

Neyman, J., and E. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

Nickell, S., 1981, Biases in dynamic models with fixed effects, Econometrica 49, 1399-1416.

Rao, C.R., Linear statistical inference and its applications, the 2nd eds., 1973, John Wiley & Sons Inc, New York, USA.

Rotnitzky, A., D. R. Cox, M. Bottai, and J. Robins, 2000, Likelihood-based inference with singular information matrix, Bernoulli 6 (2), 243-284.

Figure 1: The Value of the Expected Concentrated Likelihood for T=2

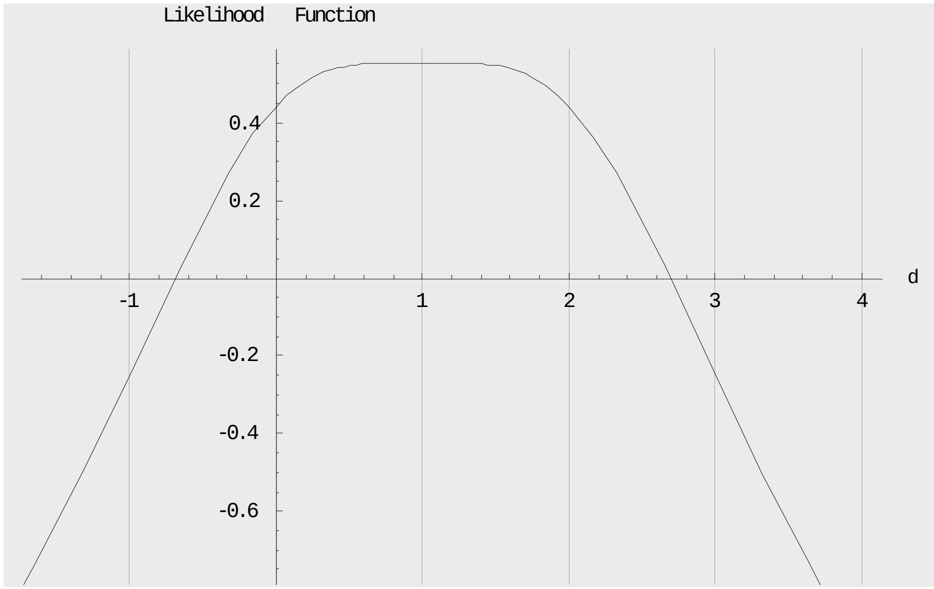


Figure 2: The Value of the expectation of Lancaster's Likelihood function for T=2

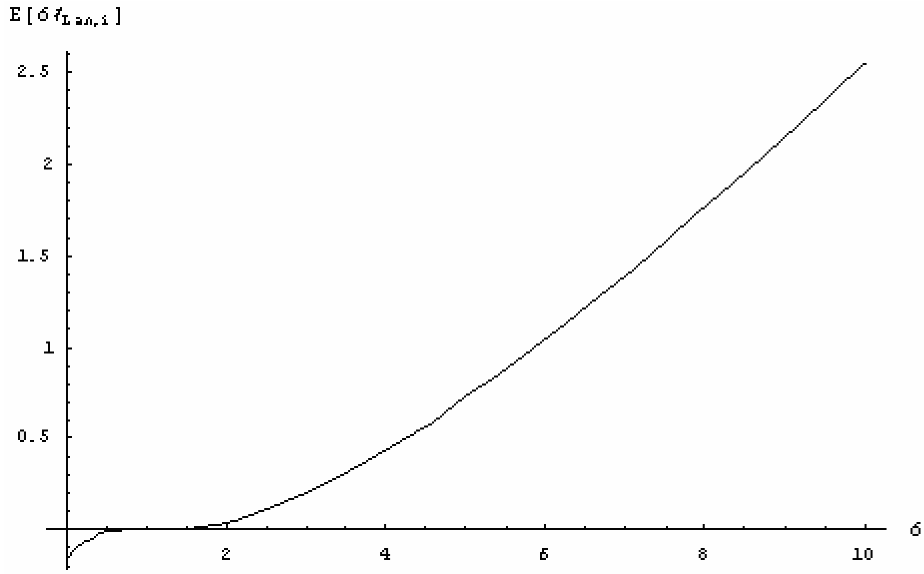


Table 1

Simulation results of the three estimation procedures

		Mean	Bias	MSE	Rejections of H_o: $\delta = \delta_0$ 1%	5%	10%	Rejections of H_o: $\delta = 1$ 1%	5%	10%
RE	N = 500									
	$\delta_0 = 0.5$	0.5001	0.0001	0.0007	1.22%	5.06%	9.98%	100%	100%	100%
	$\delta_0 = 0.8$	0.8061	0.0061	0.0032	4.00%	8.10%	12.52%	82.76%	90.72%	93.40%
	$\delta_0 = 0.9$	0.9053	0.0053	0.0057	9.12%	16.82%	22.20%	21.44%	32.48%	38.78%
	$\delta_0 = 1$	0.9982	-0.0018	0.0047	9.20%	16.06%	20.68%	9.20%	16.06%	20.68%
	N = 100									
	$\delta_0 = 0.5$	0.5011	0.0011	0.0033	0.94%	3.94%	8.32%	99.98%	99.98%	99.98%
	$\delta_0 = 0.8$	0.8165	0.0165	0.0134	8.68%	14.46%	18.64%	31.74%	36.36%	54.88%
	$\delta_0 = 0.9$	0.8977	-0.0023	0.0144	9.88%	17.00%	22.48%	13.50%	22.38%	28.14%
	$\delta_0 = 1$	0.9958	-0.0042	0.0112	9.24%	16.02%	21.50%	9.24%	16.02%	21.50%
HPT	N = 500									
	$\delta_0 = 0.5$	0.5001	0.0001	0.0007	1.14%	4.82%	9.92%	100%	100%	100%
	$\delta_0 = 0.8$	0.8220	0.0220	0.0076	9.56%	13.74%	17.40%	69.58%	77.52%	80.02%
	$\delta_0 = 0.9$	0.9015	0.0015	0.0056	5.80%	12.04%	17.56%	21.40%	30.58%	35.90%
	$\delta_0 = 1$	0.9986	-0.0014	0.0043	8.34%	14.76%	19.46%	8.34%	14.76%	19.46%
	N = 100									
	$\delta_0 = 0.5$	0.5026	0.0026	0.0046	1.12%	4.02%	8.46%	99.90%	99.92%	99.99%
	$\delta_0 = 0.8$	0.8229	0.0229	0.0156	5.74%	11.44%	15.84%	30.24%	42.38%	49.56%
	$\delta_0 = 0.9$	0.8846	-0.0054	0.0122	5.40%	10.08%	15.52%	12.96%	21.34%	26.58%
	$\delta_0 = 1$	0.9963	-0.0037	0.0099	8.10%	14.26%	18.94%	8.10%	14.26%	18.94%
Lancaster	N = 500									
	$\delta_0 = 0.5$	0.5012	0.0012	0.0016	1.00%	4.70%	9.92%	100%	100%	100%
	$\delta_0 = 0.8$	0.7941	-0.0059	0.0025	2.24%	6.04%	9.94%	72.56%	82.56%	86.74%
	N = 100									
	$\delta_0 = 0.5$	0.5012	0.0012	0.0034	1.42%	5.92%		99.94%	99.98%	99.98%

Notes: Table 1 reports the mean, bias and mean square error of 1000 simulations with the number of observations equal to N and the true value of delta equal to δ_0 along with the rejection rate of the likelihood ratio test with significance levels of 1%, 5% and 10% for the Random Effects (RE), Hsaio, Pesaran and Tahmiscoglu (HPT), and Lancaster's Estimation procedures. Lancaster's procedure fails to converge for large values of δ_0, thus results for $\delta_0 = 0.5$ and 0.8 only are given for N=500, and 0.5 only for N=100. With N=500 and δ_0=0.8, still 8% of the simulations did not converge, and the results are those for the simulations that did converge.

Figure 3: The distribution of the RE estimator, varying δ, N=500

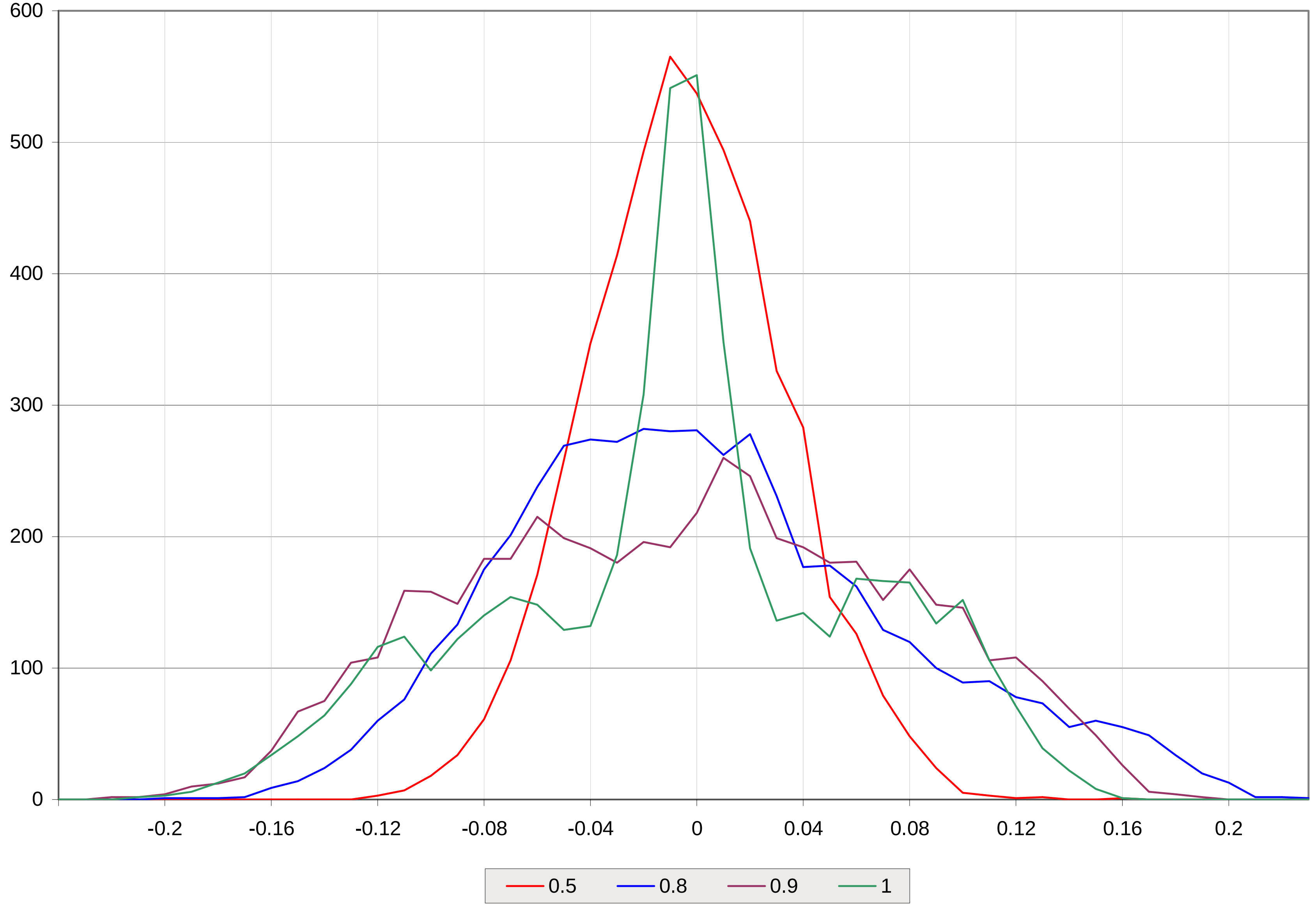


Figure 4: The distribution of the RE estimator, varying δ, N=100

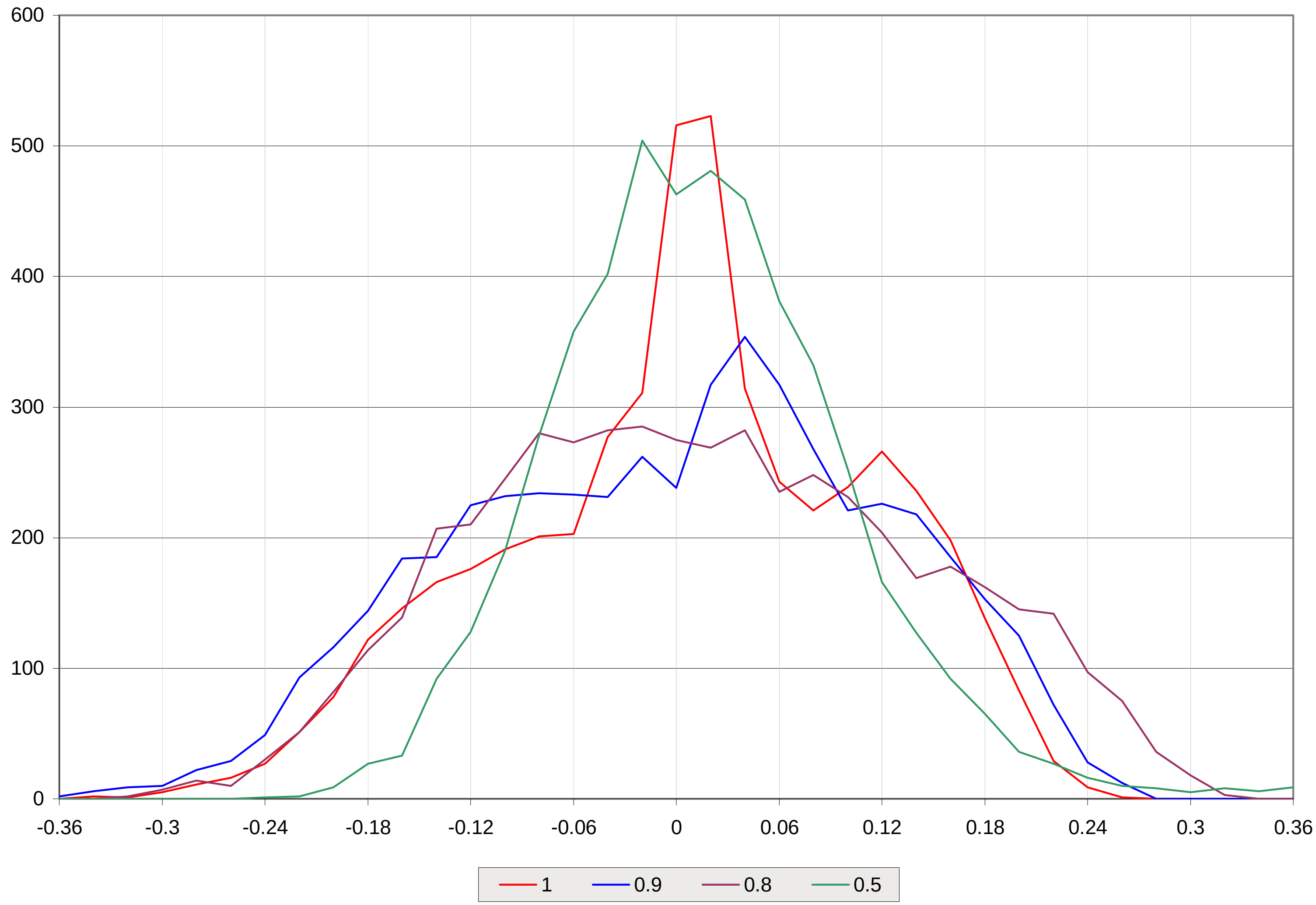


Figure 5: Procedure Comparison, N=500, δ=0.5

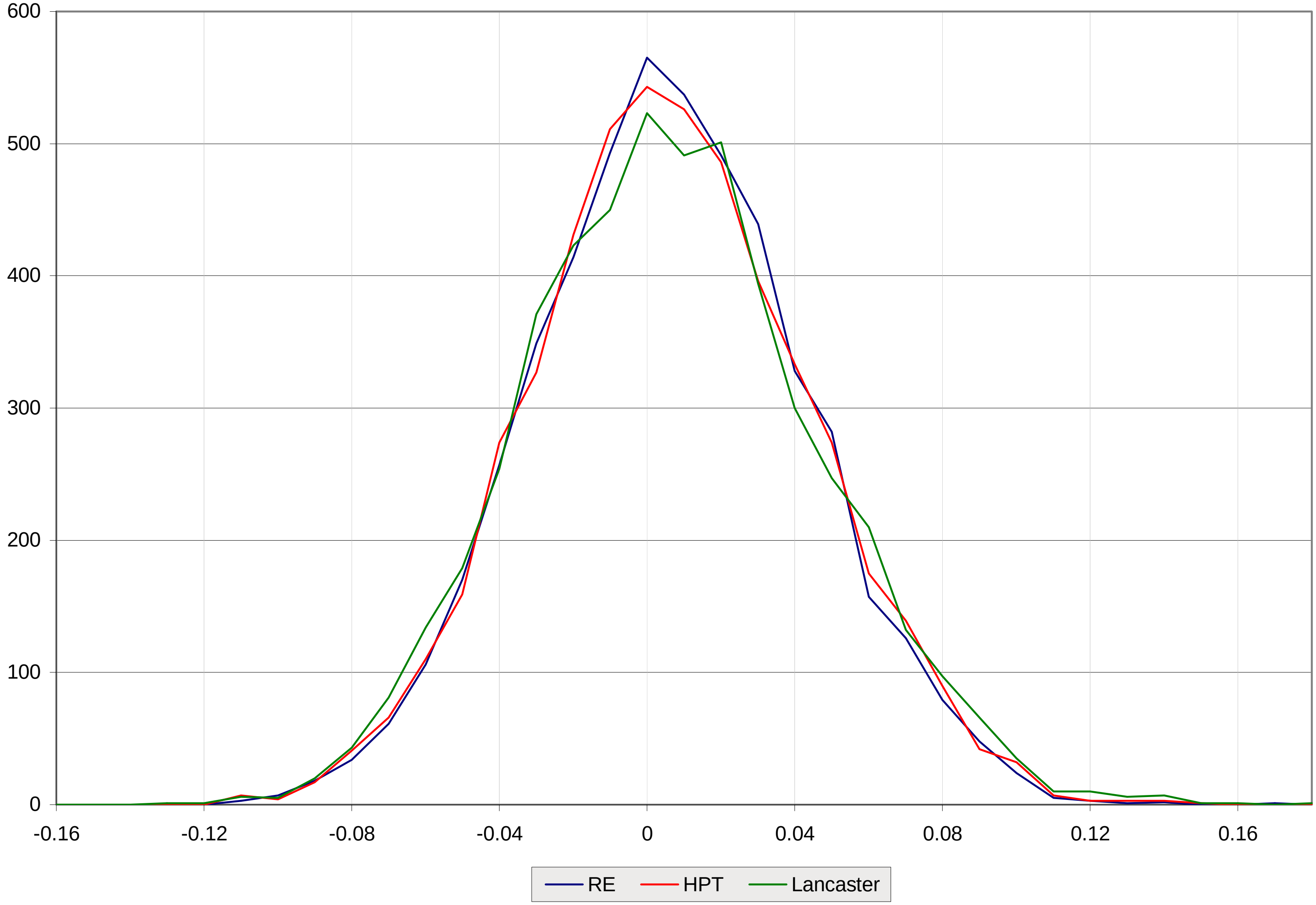


Figure 6: Procedure Comparison, N=500, δ=0.8

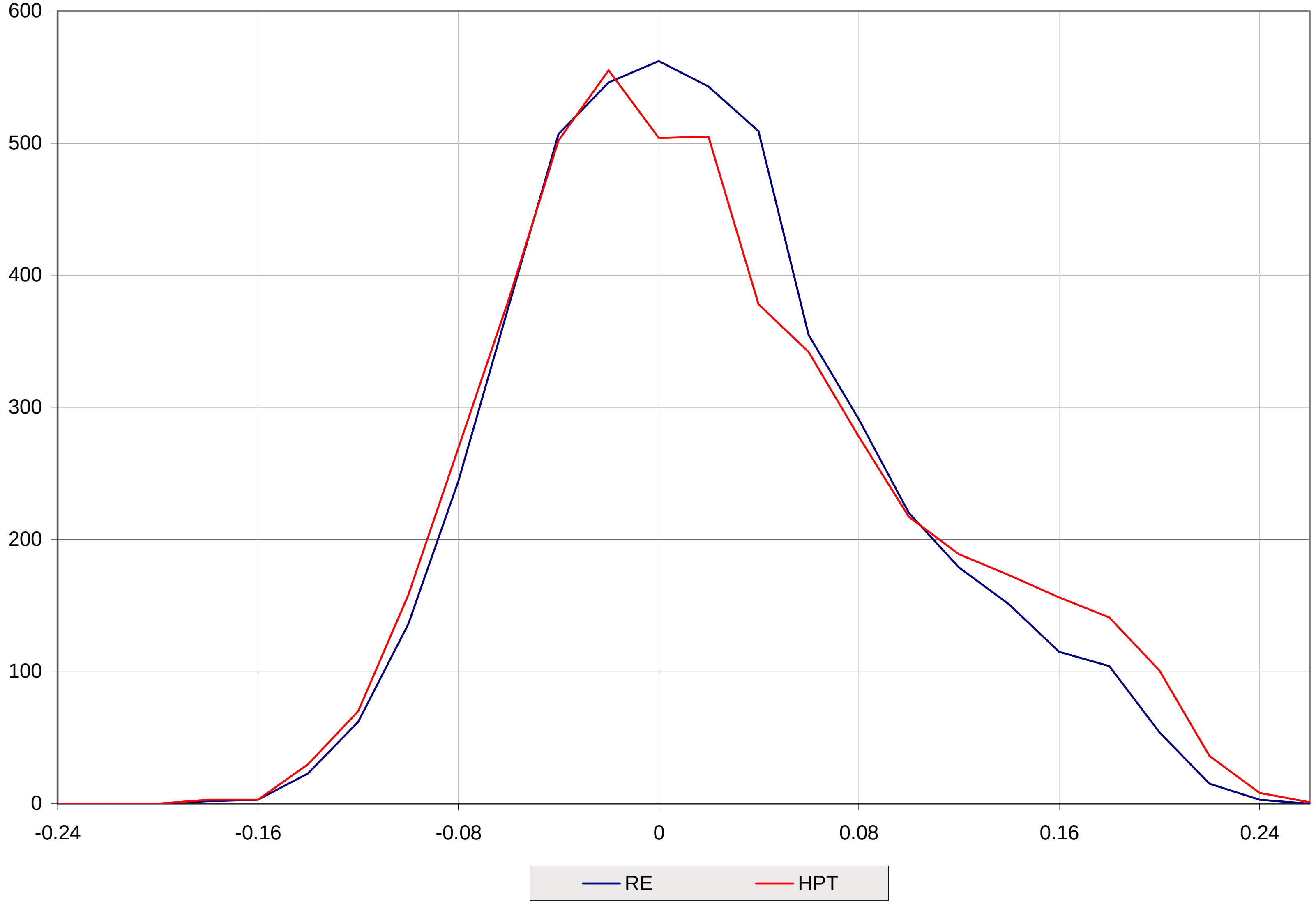


Figure 7: Procedure Comparison, N=500, δ=0.9

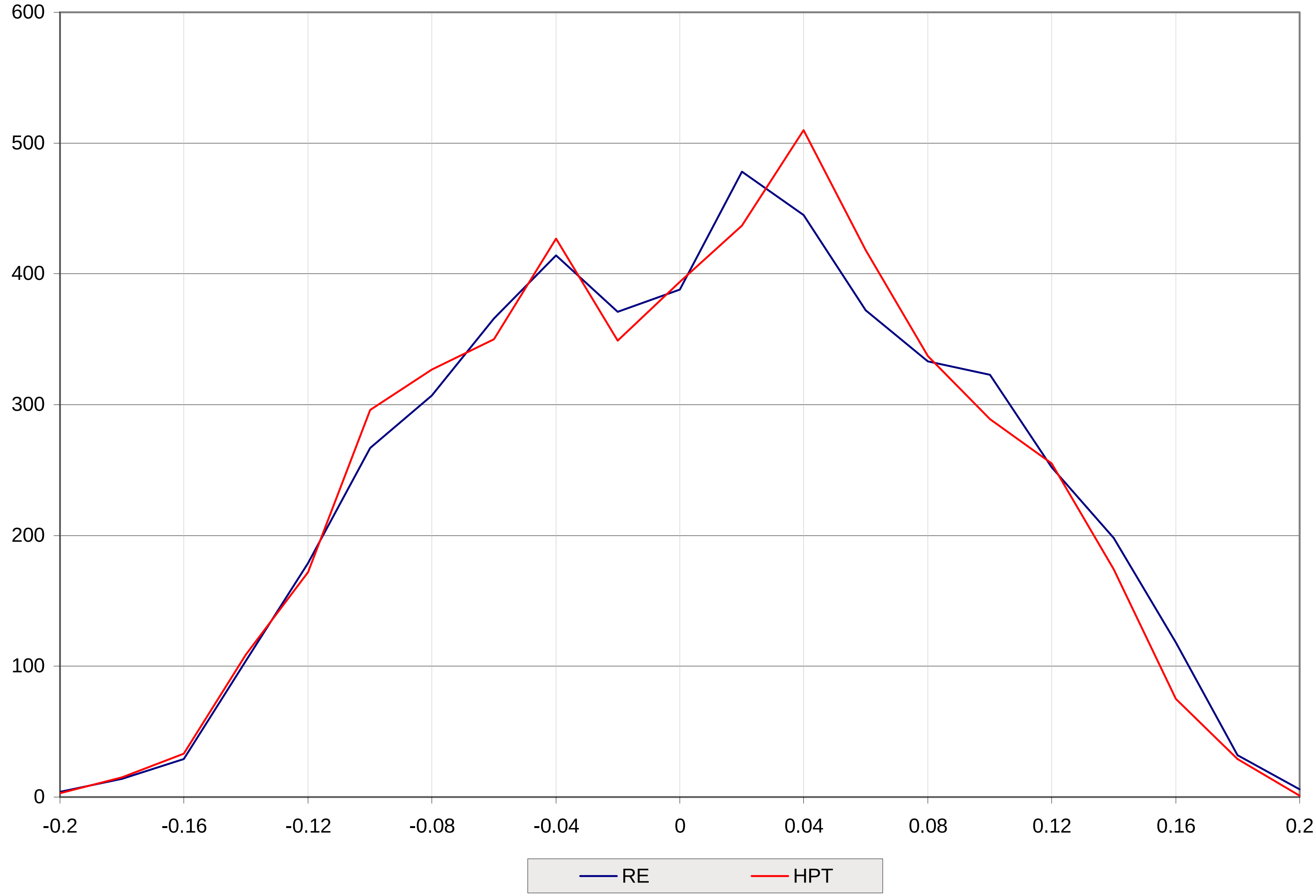


Figure 8: Procedure Comparison, N=500, δ=1

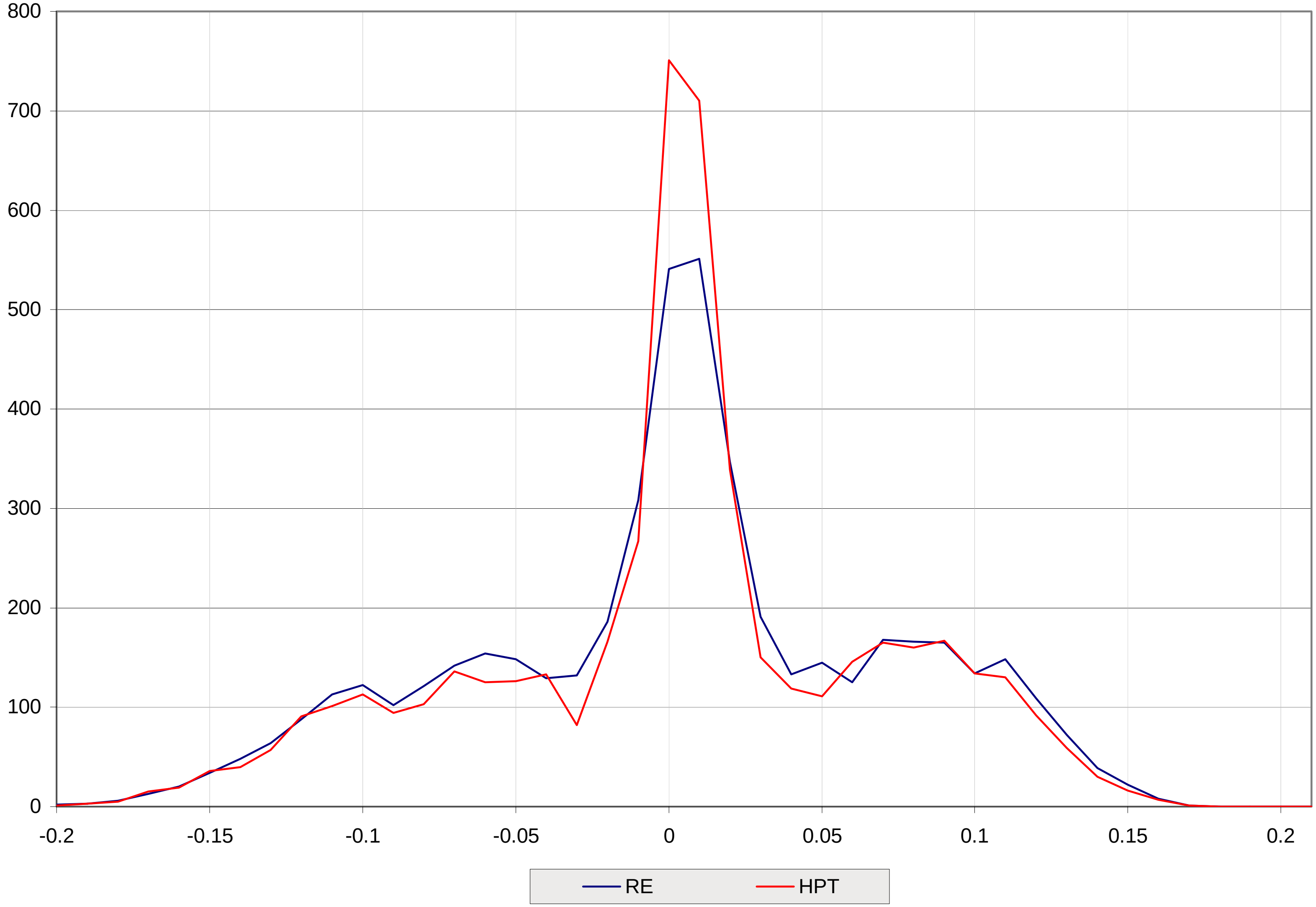


Figure 9: Procedure Comparison, N=100, δ=0.5

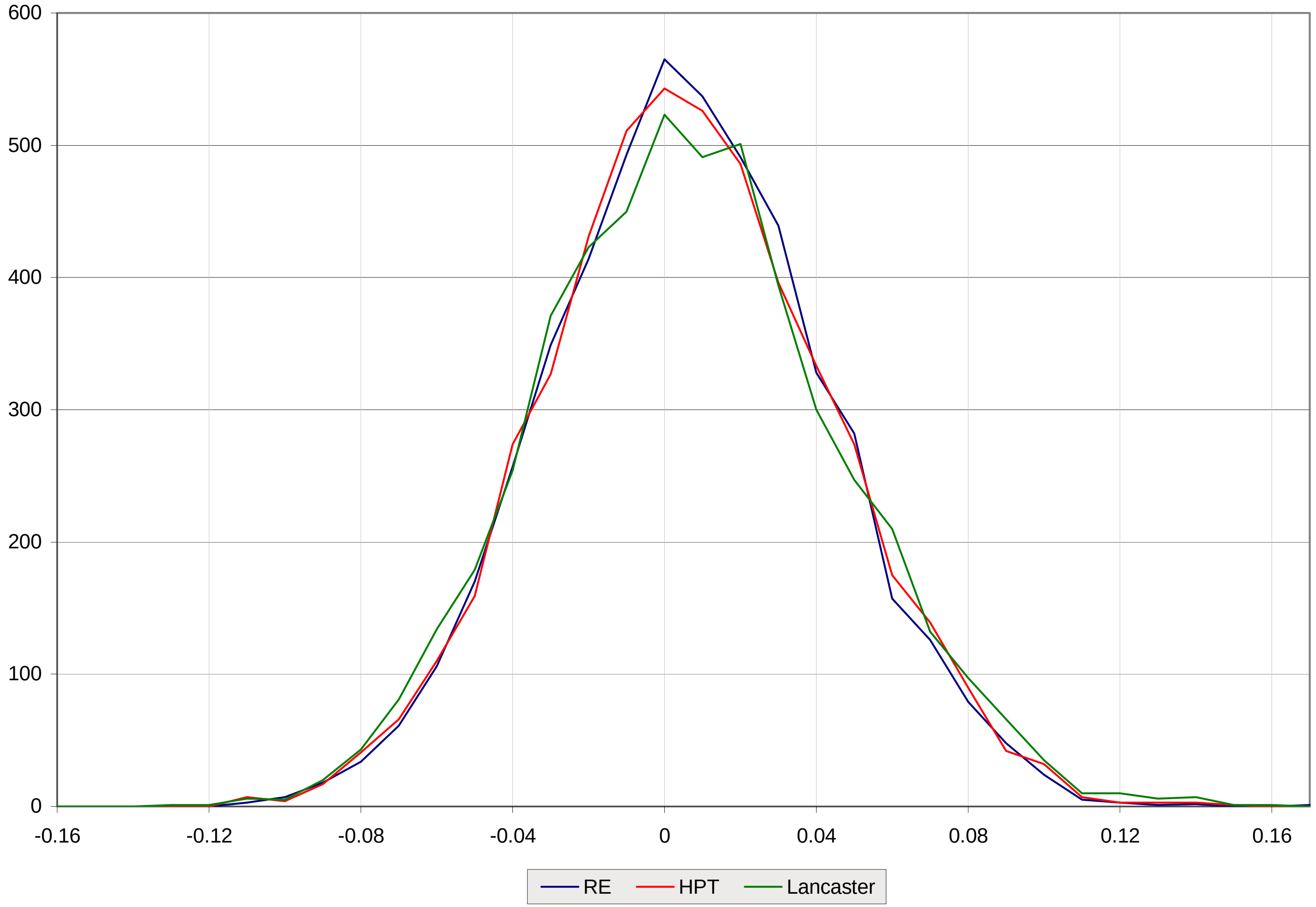


Figure 10: Procedure Comparison, N=100, δ=0.8

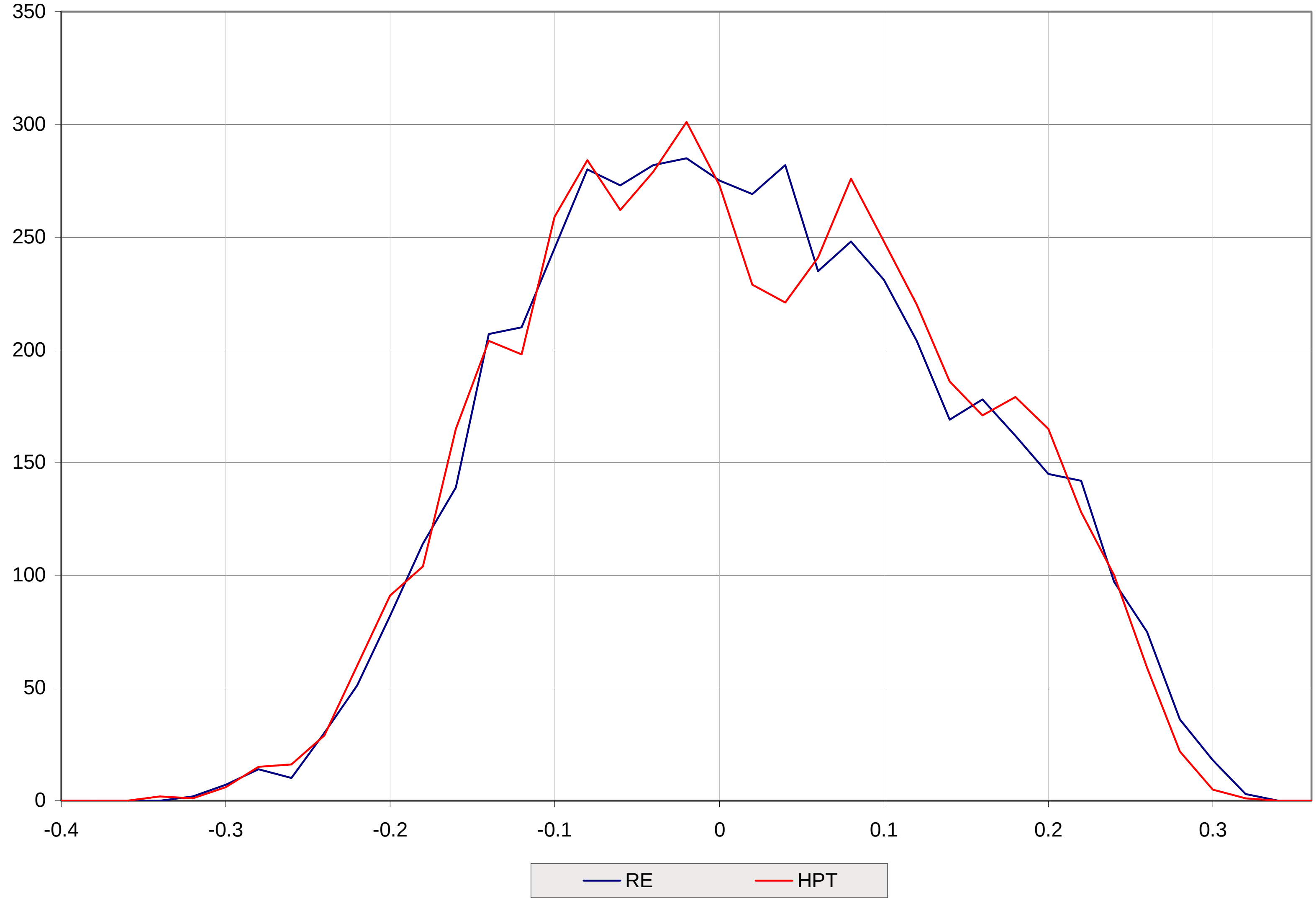


Figure 11: Procedure Comparison, N=100, δ=0.9

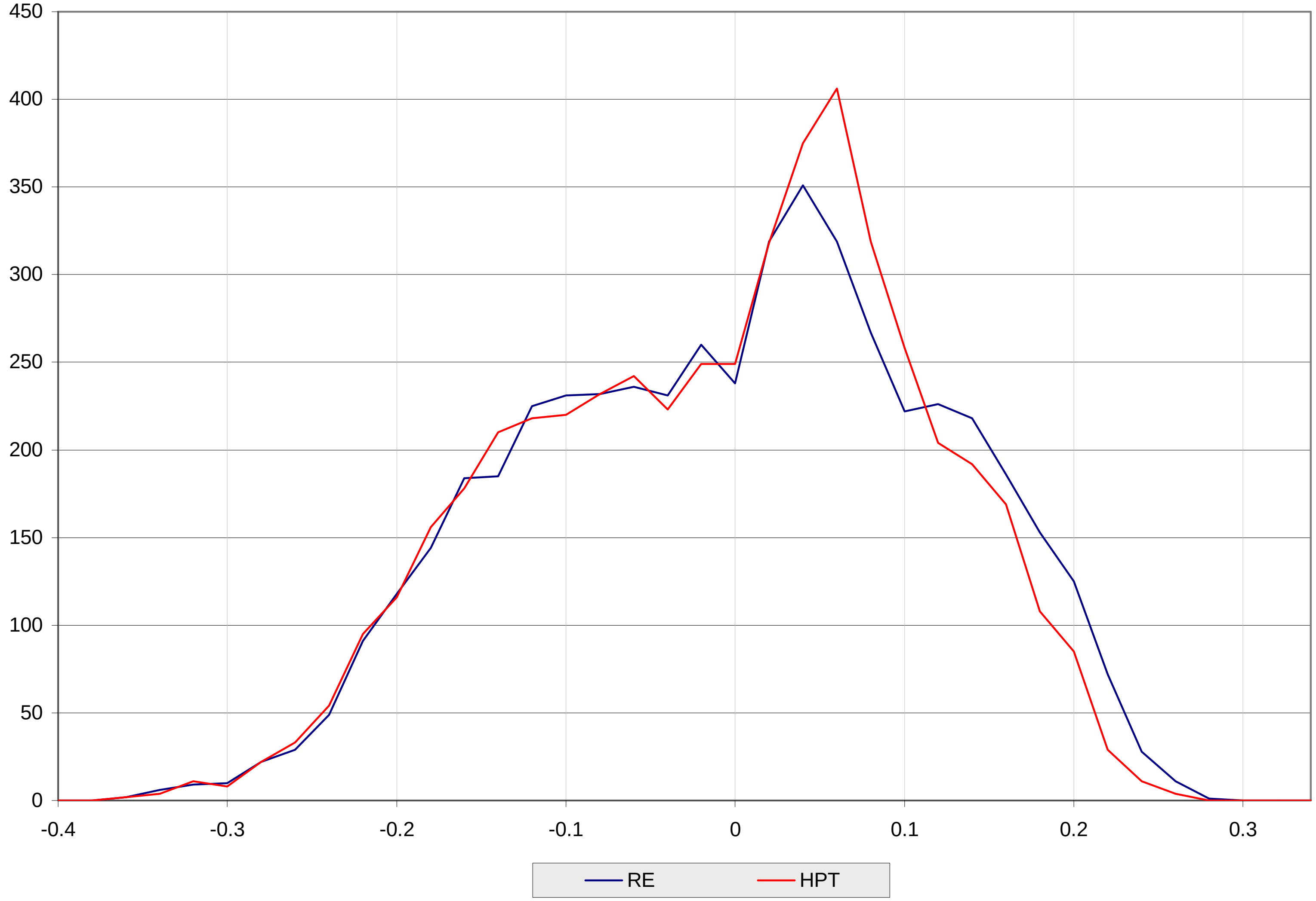


Figure 12: Procedure Comparison, N=100, δ=1

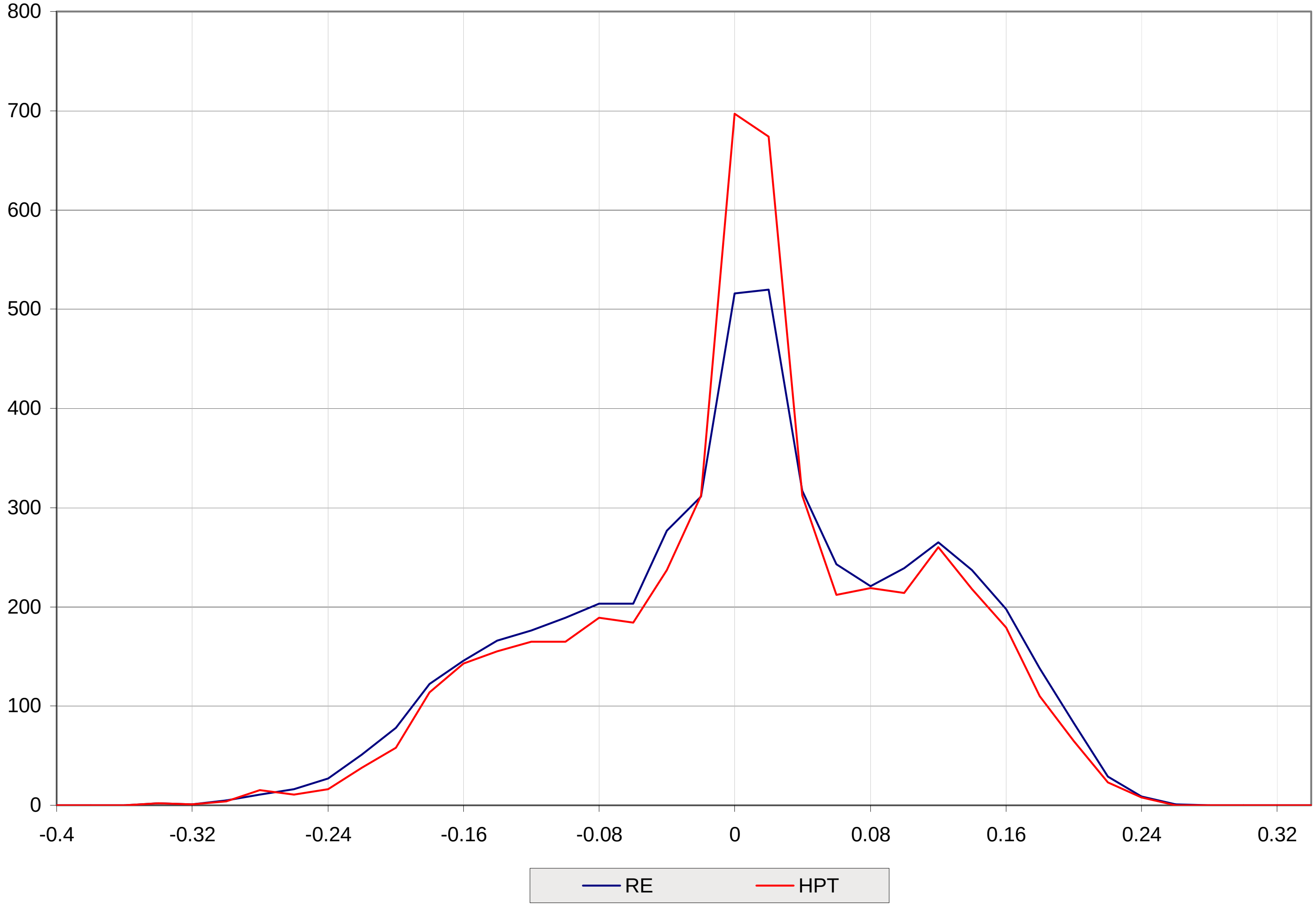
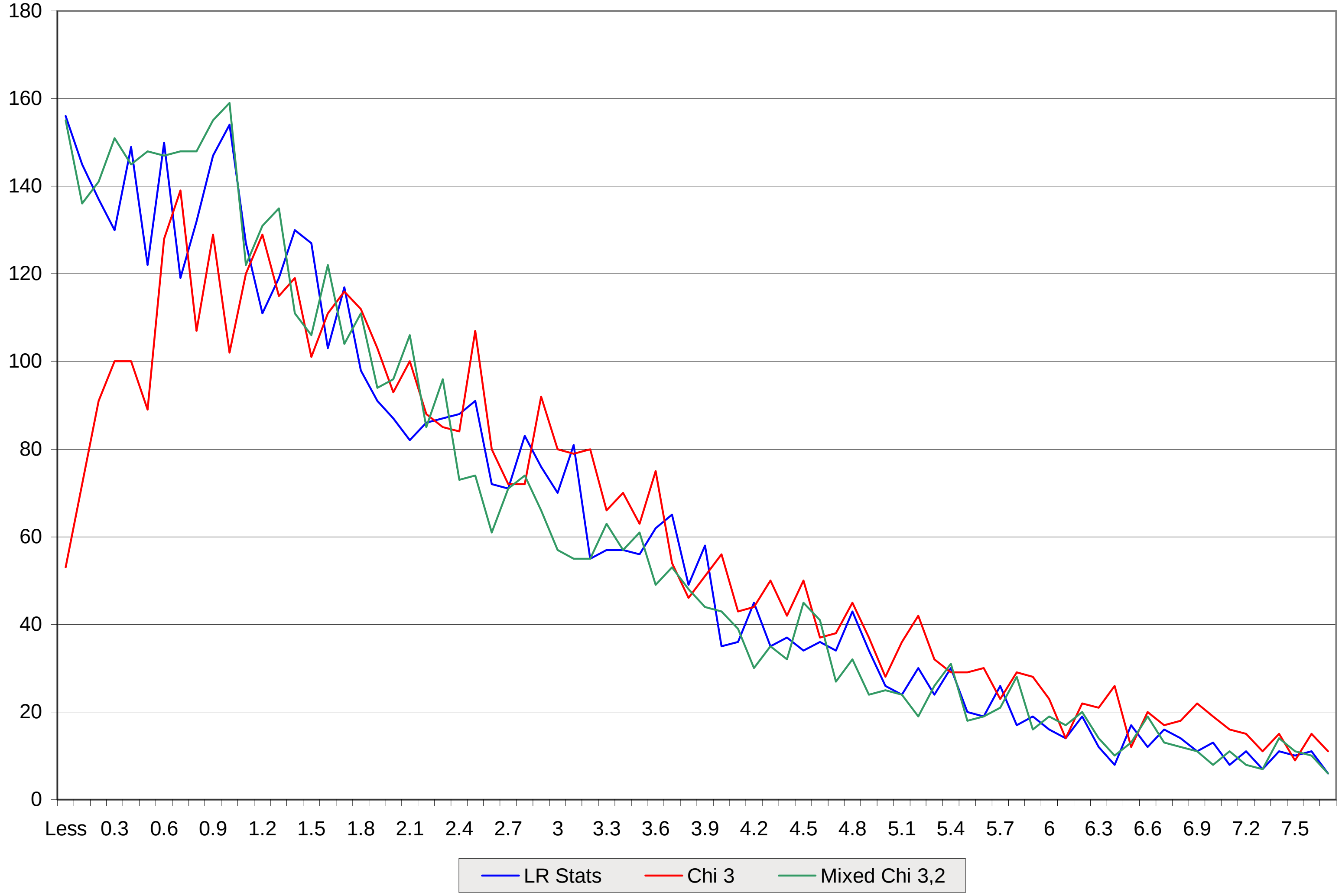
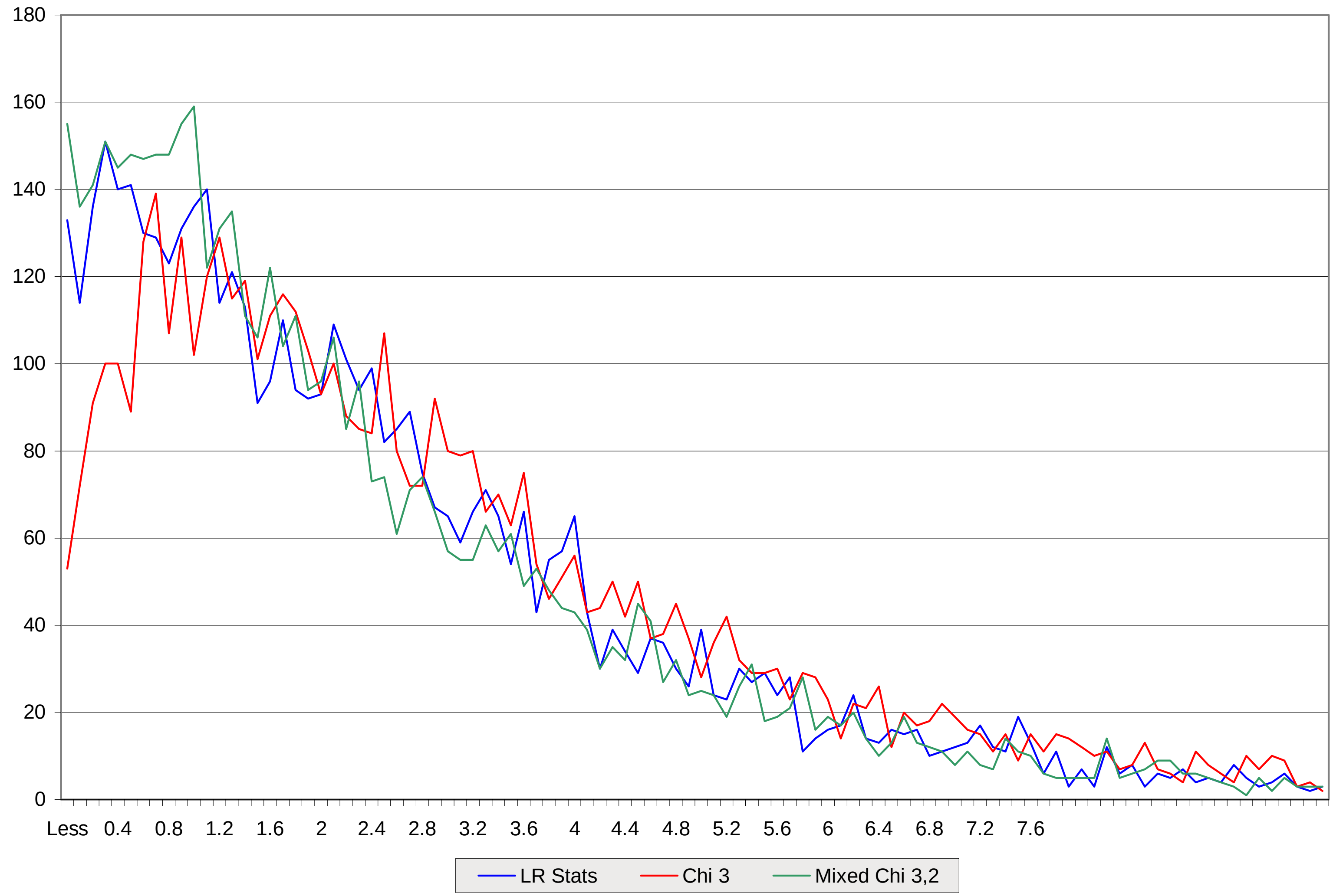


Figure 13: The distribution of the Likelihood Ratio Test, N = 500



The distribution of LR statistics for the test of H_o: $\delta = 1$ using the RE ML estimator with a stationary data generating process, true δ = 1, N = 500, compared with $\chi^2(2)$, $\chi^2(3)$, and a mixed distribution of $\chi^2(2)$ and $\chi^2(3)$.

Figure 14: The distribution of the Likelihood Ratio Test, N = 100



The distribution of LR statistics for the test of H_o: $\delta = 1$ using the RE ML estimator with a stationary data generating process, true $\delta = 1$, N = 100, compared with $\chi^2(2)$, $\chi^2(3)$, and a mixed distribution of $\chi^2(2)$ and $\chi^2(3)$.

Table 4: The Size and Power of the Likelihood Ratio Test Based on RE ML

	$\delta_0 = 1$			$\delta_0 = 0.95$			$\delta_0 = 0.9$		
Significance	1%	5%	10%	1%	5%	10%	1%	5%	10%
σ_η^2=2, $\sigma^2_{q_0}$ =2, ν=1									
N=500									
$\chi^2(3)$	0.82	3.66	7.22	91.08	97.62	98.98	100	100	100
$\chi^2(2)$	1.98	8.02	15.60	95.80	99.14	99.66	100	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.04	5.10	9.86	92.98	98.42	99.38	100	100	100
N=100									
$\chi^2(3)$	0.88	3.88	8.14	14.84	34.08	46.03	67.26	86.58	92.58
$\chi^2(2)$	2.22	9.04	16.54	25.18	48.42	61.38	79.82	93.34	96.88
Mix $\chi^2(3)$, $\chi^2(2)$	1.24	5.88	11.04	18.46	39.56	52.74	72.48	90.04	94.60
σ_η^2=2, $\sigma^2_{q_0}$ =4, ν=1									
N=500									
$\chi^2(3)$	0.86	3.10	7.06	100	100	100	100	100	100
$\chi^2(2)$	1.72	8.08	15.90	100	100	100	100	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.16	4.94	10.12	100	100	100	100	100	100
N=100									
$\chi^2(3)$	0.84	3.96	7.96	77.76	91.68	95.44	100	100	100
$\chi^2(2)$	2.16	8.74	16.82	87.04	96.10	98.30	100	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.30	5.50	10.84	81.90	93.62	97.24	100	100	100
σ_η^2=2, $\sigma^2_{q_0}$ =1, ν=1									
N=500									
$\chi^2(3)$	0.78	3.85	7.56	62.82	81.76	88.84	99.96	100	100
$\chi^2(2)$	2.00	8.62	15.66	74.86	89.92	94.64	99.92	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.18	5.38	10.36	67.58	85.66	91.84	99.94	100	100
N=100									
$\chi^2(3)$	0.78	4.02	8.00	6.46	20.64	30.76	39.38	64.08	75.18
$\chi^2(2)$	2.20	9.24	16.74	13.78	32.68	45.78	52.70	76.82	85.28
Mix $\chi^2(3)$, $\chi^2(2)$	1.20	5.56	11.04	8.86	25.24	36.62	44.04	69.38	79.82
σ_η^2=2, $\sigma^2_{q_0}$ =2, ν=1									
N=500									
$\chi^2(3)$	0.82	3.66	7.22	91.08	97.62	98.98	100	100	100
$\chi^2(2)$	1.98	8.02	15.60	95.80	99.14	99.66	100	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.04	5.10	9.86	92.98	98.42	99.38	100	100	100
N=100									
$\chi^2(3)$	0.88	3.88	8.14	14.84	34.08	46.30	67.26	86.58	92.58
$\chi^2(2)$	2.22	9.04	16.54	25.18	48.42	61.38	79.82	93.34	96.88
Mix $\chi^2(3)$, $\chi^2(2)$	1.24	5.88	11.04	18.46	39.56	46.30	72.48	90.04	94.60
σ_η^2=2, $\sigma^2_{q_0}$ =2, ν=4									
N=500									
$\chi^2(3)$	0.82	3.66	7.22	65.02	83.04	89.88	99.96	100	100
$\chi^2(2)$	1.98	8.02	15.60	76.36	90.68	95.04	100	100	100
Mix $\chi^2(3)$, $\chi^2(2)$	1.04	5.40	9.86	69.32	86.58	92.48	99.98	100	100
N=100									
$\chi^2(3)$	0.88	3.88	8.14	6.74	20.82	31.36	44.06	68.86	78.84
$\chi^2(2)$	2.22	9.04	16.54	13.90	33.36	46.00	58.42	80.54	88.66
Mix $\chi^2(3)$, $\chi^2(2)$	1.24	5.88	11.04	8.98	22.56	37.60	49.56	73.50	83.28

Notes: Table 5 reports the size and power of the likelihood ratio test for the Random Effects estimator using $\chi^2(3)$, $\chi^2(2)$, and a mixture of $\chi^2(3)$ and $\chi^2(2)$ statistics over 1000 simulations with the number of observations equal to N and the true value of delta equal to δ_0 for varying values of σ_η^2, $\sigma_{q_0}^2$, and ν.

Table 5: The Size and Power of the Likelihood Ratio Test Based on HPT ML

	$\delta^0 = 1$			$\delta^0 = 0.95$			$\delta^0 = 0.9$		
Significance	1%	5%	10%	1%	5%	10%	1%	5%	10%
σ_η^2=2, $\sigma_{q_0}^2$ =2, ν=1									
N=500									
$\chi^2(2)$	0.72	3.14	6.52	45.58	66.10	76.76	88.06	96.94	98.48
$\chi^2(1)$	2.26	9.26	17.80	60.46	82.10	89.22	95.90	99.08	99.70
Mix $\chi^2(1)$, $\chi^2(2)$	0.96	4.96	9.48	48.82	73.10	82.14	91.74	97.92	99.10
N=100									
$\chi^2(2)$	0.66	3.60	6.90	4.96	15.22	23.70	14.34	32.44	44.30
$\chi^2(1)$	2.24	10.22	18.94	12.08	30.04	42.30	27.38	52.26	64.60
Mix $\chi^2(1)$, $\chi^2(2)$	1.14	5.08	10.42	7.22	19.88	30.46	18.06	39.24	52.66
σ_η^2=2, $\sigma_{q_0}^2$ =4, ν=1									
N=500									
$\chi^2(2)$	0.72	3.14	6.52	2.70	8.80	16.30	15.26	34.68	46.86
$\chi^2(1)$	2.26	9.26	17.80	6.76	21.54	33.02	30.02	54.70	68.00
Mix $\chi^2(1)$, $\chi^2(2)$	0.96	4.96	9.48	3.82	12.92	21.78	20.42	41.56	55.20
N=100									
$\chi^2(2)$	0.66	3.60	6.90	1.14	4.82	9.12	2.92	9.66	16.48
$\chi^2(1)$	2.24	10.22	18.94	3.76	12.58	22.34	7.18	22.52	33.88
Mix $\chi^2(1)$, $\chi^2(2)$	1.14	5.08	10.42	1.80	7.18	12.92	4.06	13.58	22.84
σ_η^2=2, $\sigma_{q_0}^2$ =1, ν=1									
N=500									
$\chi^2(2)$	0.72	3.14	6.52	61.56	81.70	88.96	99.88	99.98	100
$\chi^2(1)$	2.26	9.26	17.80	78.06	92.20	96.28	99.96	100	100
Mix $\chi^2(1)$, $\chi^2(2)$	0.96	4.94	9.48	68.08	86.34	93.32	99.94	100	100
N=100									
$\chi^2(2)$	0.66	3.60	6.90	7.54	20.16	30.30	37.34	62.24	73.84
$\chi^2(1)$	2.24	10.22	18.94	16.76	37.08	50.64	57.02	80.28	88.24
Mix $\chi^2(1)$, $\chi^2(2)$	1.14	5.08	10.42	10.24	26.00	37.58	44.46	68.96	80.42
σ_η^2=2, $\sigma_{q_0}^2$ =2, ν=2									
N=500									
$\chi^2(2)$	0.72	3.14	6.52	61.56	81.70	88.96	99.88	99.98	100
$\chi^2(1)$	2.26	9.26	17.80	78.06	92.20	96.28	99.96	100	100
Mix $\chi^2(1)$, $\chi^2(2)$	0.96	4.94	9.48	68.08	86.34	93.32	99.94	100	100
N=100									
$\chi^2(2)$	0.66	3.60	6.90	7.54	20.16	30.30	37.34	62.24	73.84
$\chi^2(1)$	2.24	10.22	18.94	16.76	37.08	50.64	57.02	80.28	88.24
Mix $\chi^2(1)$, $\chi^2(2)$	1.14	5.08	10.42	10.24	26.00	37.58	44.46	68.96	80.42
σ_η^2=2, $\sigma_{q_0}^2$ =2, ν=4									
N=500									
$\chi^2(2)$	0.72	3.14	6.52	66.48	84.66	91.16	99.96	100	100
$\chi^2(1)$	2.26	9.26	17.80	81.32	93.94	97.16	100	100	100
Mix $\chi^2(1)$, $\chi^2(2)$	0.96	4.94	9.48	72.62	89.02	94.96	99.98	100	100
N=100									
$\chi^2(2)$	0.66	3.60	6.90	8.40	22.14	31.80	46.00	70.28	80.96
$\chi^2(1)$	2.24	10.22	18.94	17.84	39.32	53.02	65.24	85.98	92.00
Mix $\chi^2(1)$, $\chi^2(2)$	1.14	5.08	10.42	11.14	27.80	39.72	53.52	76.76	86.16

Notes: Table 6 reports the size and power of the likelihood ratio test for the HPT estimator using $\chi^2(2)$, $\chi^2(1)$, and a mixture of $\chi^2(2)$ and $\chi^2(1)$ statistics over 1000 simulations with the number of observations equal to N and the true value of delta equal to δ_0 for varying values of σ_η^2, $\sigma_{q_0}^2$, and ν.