Econometric Theory, **31**, 2015, 539–559.
doi:10.1017/S0266466614000656

ASYMPTOTICALLY UMP PANEL UNIT ROOT TESTS—THE EFFECT OF HETEROGENEITY IN THE ALTERNATIVES

I. Gaia Becheri
Delft University of Technology

Feike C. Drost and Ramon van den Akker
Tilburg University

In a Gaussian, heterogeneous, cross-sectionally independent panel with incidental intercepts, Moon, Perron, and Phillips (2007, *Journal of Econometrics* 141, 416–459) present an asymptotic power envelope yielding an upper bound to the local asymptotic power of unit root tests. In case of homogeneous alternatives this envelope is known to be sharp, but this paper shows that it is not attainable for heterogeneous alternatives. Using limit experiment theory we derive a sharp power envelope. We also demonstrate that, among others, one of the likelihood ratio based tests in Moon et al. (2007, *Journal of Econometrics* 141, 416–459), a pooled generalized least squares (GLS) based test using the Breitung and Meyer (1994, *Applied Economics* 25, 353–361) device, and a new test based on the asymptotic structure of the model are all asymptotically UMP (Uniformly Most Powerful). Thus, perhaps somewhat surprisingly, pooled regression-based tests may yield optimal tests in case of heterogeneous alternatives. Although finite-sample powers are comparable, the new test is easy to implement and has superior size properties.

1. INTRODUCTION

Just as for univariate time series, the presence or absence of unit roots in a panel data model can have crucial economic policy implications. Therefore, in the last two decades, a lot of attention has been given to test for the presence of unit roots in panel data; see, for example, the surveys Banerjee (1999), Baltagi and Kao (2000), Choi (2006), Breitung and Pesaran (2008), and Westerlund and Breitung (2013).

In this context it is important to know the (asymptotic) power of panel unit root tests. For cross-sectionally independent panels, local asymptotic powers of tests

The authors wish to thank the referees and a co-editor for constructive comments. In particular, we are grateful for suggesting optimality of a test based on Breitung and Meyer (1994). Address correspondence to Ramon van den Akker, Econometrics Group, CentER, Tilburg University, Tilburg, the Netherlands; e-mail: R.vdnAkker@TilburgUniversity.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

have been considered in, for example, Breitung (2000), Moon and Perron (2008), Harris, Harvey, Leybourne, and Sakkas (2010), and Madsen (2010); (asymptotic) optimality for Gaussian panels has been studied in Moon, Perron, and Phillips (2007). For models with cross-sectional dependence, local asymptotic powers have been derived in, for example, Moon and Perron (2004), Breitung and Das (2005), and Westerlund and Larsson (2012).

This paper reconsiders the Gaussian, heterogeneous, cross-sectionally independent panel in Moon et al. (2007; MPP henceforth) allowing for incidental intercepts. In this setup, with ρ_i the autoregression parameter of panel unit i, the unit root null hypothesis corresponds to $\rho_i = 1$ for all i. Local alternatives are of the form $\rho_i = 1 + hH_i/(T\sqrt{n})$, where $h < 0$, n denotes the number of panel units, T is the time series dimension, and $H_1, \ldots, H_n$ are i.i.d. (independent and identically distributed) *unobserved* random variables with mean 1 and unknown distribution. Note that the random perturbations H_i yield heterogeneity in the alternatives. Using an *auxiliary model* where both panel units and random perturbations H_i are observed, MPP derived an asymptotic power envelope for this testing problem. This power envelope was shown to be sharp in case one would be able to observe both the panel units and the perturbations. Thus, in particular, their power envelope was shown to be attainable if attention is restricted to homogeneous alternatives (i.e., H_i degenerated at 1). More generally, using only the panel units, the MPP power envelope determines an upper bound to the local asymptotic power of unit root tests.

The main contribution of this paper is twofold. First, we show that the upper bound in MPP is not sharp in case of heterogeneous alternatives (i.e., when H_i is not degenerated), i.e., it is not possible to construct a test, based on the panel units only, with an asymptotic local power function that attains the MPP upper bound in at least one point. Using limit experiment theory (à la Le Cam) we derive an attainable, sharp power envelope. Our derivation also clarifies the precise role of heterogeneous alternatives. Second, we show that, among others, one of the likelihood ratio based tests in Moon et al. (2007), a pooled generalized least squares (GLS) based test using the Breitung and Meyer (1994) device, and a new test based on the asymptotic structure of the model are all asymptotically Uniformly Most Powerful (UMP). The new test is easy to implement, its critical values are readily available with excellent size properties, and the finite-sample power is comparable to the other tests.

Our motivation to use limit experiment theory is twofold. First, due to the unobserved random perturbations, it is not possible to obtain closed-form likelihood ratios. The "standard" approach to derive a power envelope, applying the Neyman–Pearson lemma in the finite-sample situation and passing to the limit afterward, is therefore not feasible. Second, using "Le Cam calculus," in particular Le Cam's third lemma, the local asymptotic power of any test can be obtained without triangular array arguments. For unit root problems in (univariate) time series, limit experiment theory has been exploited by Jansson (2008) and Hallin, Van den Akker, and Werker (2011) (for the continuous-valued case) and Drost, Van den Akker, and Werker (2009) (for the integer-valued case).[1] To the best of

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

our knowledge, this paper is the first to consider the limit experiment for panel data models with unit roots.

The paper is organized as follows. Section 2 describes the model setup and Section 3 presents an "information-loss" result which is crucial to our derivation of limit experiments. In Section 4 we discuss our main results: Section 4.1 contains the results on the limit experiments and power envelopes, Section 4.2 discusses the optimality of tests, and in Section 4.3 we report the results of a Monte Carlo study. Section 5 concludes. The proofs are gathered in the Appendix.

2. SETUP

We assume the observations Y_{it}, $i = 1, \dots, n$ and $t = 1, \dots, T$, to be generated by the component model

$$Y_{it} = m_i + Y_{it}^0, \tag{1}$$

$$Y_{it}^0 = \rho_i Y_{i,t-1}^0 + \sigma_i \varepsilon_{it}, \tag{2}$$

where m_i is a (deterministic) intercept, i.e., fixed effect, and $Y_{i0}^0 = 0$. The autoregression coefficients ρ_i are assumed to be generated by the random coefficient structure

$$\rho_i = 1 + \frac{h}{T\sqrt{n}} H_i, \tag{3}$$

where H_i is random with mean 1 and the (deterministic) parameter h describes the (mean) deviation from the unit root. Assumption 1 describes our precise conditions on the perturbations H_i and the idiosyncratic shocks ε_{it}.

Assumption 1.

(a) The innovations ε_{it}, $i, t \in \mathbb{N}$, are i.i.d. $\mathcal{N}(0, 1)$.

(b) The perturbations H_i, $i \in \mathbb{N}$, are i.i.d. with mean 1 and independent of the idiosyncratic shocks ε_{it}, $i, t \in \mathbb{N}$. Moreover, the moment generating function of H_1 exists on an open interval containing 0.

(c) The deterministic scale parameters σ_i are strictly positive, i.e., $\sigma_i > 0$ for $i \in \mathbb{N}$.

Remark 1. For univariate time series it is known (see, for example, Elliott, 1999; Müller and Elliott, 2003) that the initial value can have a nonnegligible impact on the asymptotic behavior of unit root tests. We refer to Section 6.2 in MPP for a discussion of the effect of initial values on power envelopes and to Harris et al. (2010) for the effect of initial values on the local asymptotic power of the Im, Pesaran, and Shin (2003) test. As the main goal of this paper is to unveil the effect of heterogeneity in the alternatives on the (feasible) power envelope, we do not pursue an analysis of initial values here and assume $Y_{i0}^0 = 0$ just as in MPP.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

We are interested in testing the unit root hypothesis[2]

$$\mathrm{H}_0 : h = 0 \text{ versus } \mathrm{H}_a : h < 0. \tag{4}$$

Under the null hypothesis each individual time series $\{Y_{it}, t \in \mathbb{N}\}$ contains a unit root. Under the alternative, heterogeneity in the ρ_i is introduced via the perturbations H_i. Assumption 1 allows for $P(H_1 = 0) > 0$, which means that a (random) fraction of the time series $\{Y_{it}, t \in \mathbb{N}\}$ still contains a unit root under the alternative. By considering $P(H_1 = 1) = 1$ we can also cover homogenous alternatives, i.e., $\rho_i = \rho < 1$ for all i. Finally, note that even $P(H_1 < 0) > 0$ is allowed for, yielding a (random) fraction of explosive time series $\{Y_{it}, t \in \mathbb{N}\}$. This last setting, however, seems to be less relevant from an empirical point of view.

The normalization by $\sqrt{n}T$ in the specification for ρ_i is motivated by the asymptotic structure of the model as this rate of localization yields contiguous alternatives to the null hypothesis. Throughout we use joint limit theory (i.e., $T, n \to \infty$ jointly); see Phillips and Moon (1999) for precise definitions and a detailed exposition.

To unveil the effect of heterogeneous alternatives on the power envelope, we consider two models. The first one is an *auxiliary model* where both the panel units Y_{it} and the perturbations H_i are observed. Of course, these perturbations are not observed in the model of interest: only the panel units Y_{it} are observed. The MPP power envelope is generated from the auxiliary model and is therefore an upper bound to the local asymptotic power of tests in the model of interest. The next section presents a general tool to quantify the loss of information related to not observing the perturbations.

3. AN INFORMATION LOSS RESULT

This section presents an information loss result that will be essential to derive two limit experiments associated to the panel unit root model in Section 4. This general result quantifies the gain in statistical information if one also observes auxiliary data. In our panel unit root model this corresponds to (not) observing the perturbations H_i in addition to the panel units Y_{it}.

To determine a limit experiment we need to derive an asymptotic expansion of likelihood ratios. The following lemma, inspired by general results by Le Cam and Yang (1988) on information loss in i.i.d. models, will allow us to derive the limit experiment for the model of interest from the limit experiment of the auxiliary model (with additional data).

LEMMA 3.1. *Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be a sequence of σ-fields and let, for all $n \in \mathbb{N}$, $(\mathcal{X}_{ni})_{1\le i\le n}$ and $(\mathcal{J}_{ni})_{1\le i\le n}$ be $\mathcal{G}_n$-measurable random variables. And let $(U_i)_{i\in\mathbb{N}}$ be i.i.d. random variables independent of $\mathcal{G}_n$ for all n. Suppose*

(i) *the moment generating function of U_1 exists;*

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

(ii)

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_{ni}^2 = O_P(1) \text{ and } \max_{i=1,\ldots,n}\frac{|\mathcal{X}_{ni}|}{\sqrt{n}} = o_P(1); \tag{5}$$

(iii) *for all* $n \in \mathbb{N}$ *and* $i = 1, \ldots, n$, $\mathcal{J}_{ni} \geq 0$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{J}_{ni} - \frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_{ni}^2 = o_P(1) \text{ and } \max_{i=1,\ldots,n}\frac{\mathcal{J}_{ni}}{n} = o_P(1). \tag{6}$$

Then, defining

$$L_n^{\mathcal{X},U} = \exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\mathcal{X}_{ni} - \frac{1}{2n}\sum_{i=1}^{n}U_i^2\mathcal{J}_{ni}\right),$$

and

$$L_n^{\mathcal{X}} = \mathrm{E}\left[L_n^{\mathcal{X},U} \mid \mathcal{G}_n\right],$$

we have

$$L_n^{\mathcal{X}} = m_n \exp\left((\mathrm{E}U_1)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{X}_{ni} - (\mathrm{E}U_1)^2\frac{1}{2n}\sum_{i=1}^{n}\mathcal{J}_{ni}\right),$$

with $m_n \xrightarrow{p} 1$.

We note that the result is of independent interest and could, for example, be exploited in other (panel) unit root models (with alternatives generated by random unobservable perturbations).

Informally, the probability limits of $n^{-1}\sum_{i=1}^{n}U_i^2\mathcal{J}_{ni}$ and $(\mathrm{E}U_1)^2 n^{-1}\sum_{i=1}^{n}\mathcal{J}_{ni}$ can be interpreted as the Fisher information in the auxiliary model and the model of interest, respectively. Since $\mathrm{E}U_1^2 > (\mathrm{E}U_1)^2$, if U is not degenerated, we see that there is indeed an information loss in case of heterogeneous alternatives. This heuristic reasoning will be formalized in Section 4.1 where we show that both the auxiliary and the model of interest are Locally Asymptotically Normal (LAN).

4. MAIN RESULTS

This section presents our main results. Section 4.1 contains the results on the limit experiments and asymptotic power envelopes and in Section 4.2 we demonstrate that one of the likelihood ratio based tests in Moon et al. (2007), a pooled GLS based test using the Breitung and Meyer (1994) device, and a new test based on the asymptotic structure of the model are all asymptotically UMP. Finally, to assess the finite-sample behavior of these tests, Section 4.3 contains the results of a Monte Carlo study. Throughout we use the 'Δ-operator' to denote differencing across the time dimension, i.e., $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

4.1. Limit Experiments and Power Envelopes

In this section we assume that the parameters m_i and σ_i are known. We derive two limit experiments: the limit experiment of the auxiliary model, which corresponds to observing both Y_{it} and H_i, and the limit experiment of the model of interest, which corresponds to observing the panel units Y_{it} only. By applying general limit experiment theory, we obtain the power envelopes of the models as a corollary. In Section 4.2 we will show that the power envelopes do not change in case the incidental intercepts m_i and the scale parameters σ_i are nuisance parameters.

By considering limit experiments one basically determines the "limiting statistical structure" instead of limiting distributions of given (test) statistics. Using the Asymptotic Representation Theorem, which basically says that you cannot do better along the sequence of experiments than the best procedure in the limit experiment, one can analyze optimality via the limit experiment. As the auxiliary model and the model of interest turn out to be both LAN, which means that the limit experiment is a Gaussian shift experiment, the power envelopes readily follow.

We first derive the limit experiment for the auxiliary model. To this end, we denote the joint law of Y_{it} and H_i, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, by $\tilde{P}_h^{(n,T)}$. The following proposition shows that the auxiliary model is of the LAN type. To formulate the proposition we first introduce the partial sum processes,[3]

$$W_i^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \frac{1}{\sigma_i} \Delta Y_{it}, \quad u \in [0,1].$$

Note that these partial sum processes are measurable with respect to the observations Y_{it}, $i = 1, \ldots, n$ and $t = 1, \ldots, T$ (recall that $Y_{i0} = m_i$ is known in this section). Moreover, as already suggested by the notation, the partial sum processes approximate independent Brownian motions under $\tilde{P}_0^{(n,T)}$.

PROPOSITION 4.1. *Let Assumption 1 hold and let $h_n \to h \in \mathbb{R}$. Then, under $\tilde{P}_0^{(n,T)}$,*

$$\log \frac{\mathrm{d}\tilde{P}_{h_n}^{(n,T)}}{\mathrm{d}\tilde{P}_0^{(n,T)}} = h_n \tilde{\Delta}_{n,T} - \frac{1}{2} h_n^2 \tilde{J}_{n,T}, \tag{7}$$

where the central-sequence $\tilde{\Delta}_{n,T}$ and the finite-sample Fisher information $\tilde{J}_{n,T}$ are given by

$$\tilde{\Delta}_{n,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_i X_i^{(T)} \text{ and } \tilde{J}_{n,T} = \frac{1}{n} \sum_{i=1}^{n} H_i^2 J_i^{(T)},$$

with

$$X_i^{(T)} = \int_0^1 W_i^{(T)}(u-) \, \mathrm{d} W_i^{(T)}(u) \text{ and } J_i^{(T)} = \int_0^1 \left(W_i^{(T)}(u-) \right)^2 \mathrm{d}u.$$

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Moreover, under $\tilde{P}_{h_n}^{(n,T)}$, $\tilde{\Delta}_{n,T} \xrightarrow{d} \mathcal{N}(h\tilde{J}, \tilde{J})$ *and* $\tilde{J}_{n,T} \xrightarrow{p} \tilde{J} = \frac{1}{2}\mathrm{E}H_1^2$ *as* $(T, n \to \infty)$.

In empirical applications we, of course, only observe Y_{it}. Denote the law of Y_{it}, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, by $P_h^{(n,T)}$. Then we have the following relation between the likelihood ratios of the model of interest and the auxiliary model discussed above:

$$\frac{\mathrm{d}P_h^{(n,T)}}{\mathrm{d}P_0^{(n,T)}} = \mathrm{E}\left[\left.\frac{\mathrm{d}\tilde{P}_h^{(n,T)}}{\mathrm{d}\tilde{P}_0^{(n,T)}}\right| \mathcal{F}_{n,T}\right], \tag{8}$$

with $\mathcal{F}_{n,T} := \sigma(Y_{it}, i = 1, \ldots, n, t = 1, \ldots, T)$. To determine the limit experiment we need to obtain the asymptotic behavior of the conditional expectation as $(T, n \to \infty)$. Recall that an expansion holds under the joint limit $(T, n \to \infty)$ if and only if it holds for all nondecreasing sequences $T = T(n) \to \infty$ as $n \to \infty$. Therefore we can consider an arbitrary sequence $T(n) \to \infty$ in the following. Inserting (7) in (8) we obtain, under $P_0^{(n,T)}$,

$$\frac{\mathrm{d}P_{h_n}^{(n,T)}}{\mathrm{d}P_0^{(n,T)}} = \mathrm{E}\left[\left.\exp\left(\frac{h_n}{\sqrt{n}}\sum_{i=1}^{n} H_i X_i^{(T)} - \frac{h_n^2}{2n}\sum_{i=1}^{n} H_i^2 J_i^{(T)}\right)\right| \mathcal{F}_{n,T}\right].$$

This conditional expectation cannot be calculated explicitly as we do not impose explicit distributional assumptions on H_i. Exploiting Lemma 3.1, using $\mathcal{G}_n = \mathcal{F}_{n,T}$, $\mathcal{X}_{ni} = X_i^{(T)}$, $\mathcal{J}_{ni} = J_i^{(T)}$, and $U_i = H_i$, we show that the model of interest is also LAN. We organize the result in the following proposition.

PROPOSITION 4.2. *Let Assumption 1 hold and* $h_n \to h$. *Then, under* $P_0^{(n,T)}$ *as* $(T, n \to \infty)$,

$$\log \frac{\mathrm{d}P_{h_n}^{(n,T)}}{\mathrm{d}P_0^{(n,T)}} = h\Delta_{n,T} - \frac{1}{2}h^2 J_{n,T} + o_P(1), \tag{9}$$

with central-sequence $\Delta_{n,T} = n^{-1/2}\sum_{i=1}^{n} X_i^{(T)}$ *and empirical Fisher-information* $J_{n,T} = n^{-1}\sum_{i=1}^{n} J_i^{(T)}$. *Furthermore, under* $P_{h_n}^{(n,T)}$, $\Delta_{n,T} \xrightarrow{d} \mathcal{N}(hJ, J)$ *and* $J_{n,T} \xrightarrow{p} J = \frac{1}{2}$ *as* $(T, n \to \infty)$.

Remark 1. If $P(H_1 = 1) < 1$, i.e., the perturbations are nonhomogeneous, the Fisher information J is strictly less than the Fisher information $\tilde{J}$. Of course, $J = \tilde{J}$ for homogeneous alternatives.

The propositions imply that the auxiliary model and the model of interest are both of the LAN type, i.e., the limit experiment is a Gaussian shift experiment.[4] An application of the Asymptotic Representation Theorem, see, for example, Van der Vaart (1991, Thm. 7.2) or Van der Vaart (2000, Chap. 15), yields the (asymptotic) power envelope for testing hypothesis (4).

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

COROLLARY 4.1. *Let Assumption 1 hold, $\alpha \in (0,1)$, and denote $z_\alpha = \Phi^{-1}(1-\alpha)$.*

(a) *Consider a test $\tilde{\psi}_{n,T} = \tilde{\psi}_{n,T}(Y_{11},\dots,Y_{nT},H_1,\dots,H_n)$ and denote its power under $\tilde{P}_h^{(n,T)}$ by $\tilde{\pi}_{n,T}(h)$. If $\tilde{\psi}_{n,T}$ is of level α, i.e., $\limsup_{n,T\to\infty} \tilde{\pi}_{n,T}(0) \le \alpha$, we have, for all $h \le 0$,*

$$\limsup_{n,T\to\infty} \tilde{\pi}_{n,T}(h) \le \Phi\left(-z_\alpha - h\sqrt{\tilde{J}}\right). \tag{10}$$

Moreover, the test statistic $\tilde{\psi}^\star_{n,T} = 1\{\tilde{J}_{n,T}^{-1/2}\tilde{\Delta}_{n,T} \le -z_\alpha\}$ attains this upper bound (for all h).

(b) *Consider a test $\psi_{n,T} = \tilde{\psi}_{n,T}(Y_{11},\dots,Y_{nT})$ of level α with power $\pi_{n,T}(h)$. Then we have, for all $h \le 0$,*

$$\limsup_{n,T\to\infty} \pi_{n,T}(h) \le \Phi\left(-z_\alpha - h\sqrt{J}\right). \tag{11}$$

Moreover, the test statistic $\psi^\star_{n,T} = 1\{J^{-1/2}\Delta_{n,T} \le -z_\alpha\}$ attains this upper bound (for all h).

Remark 2.

(a) Remark 1 yields

$$\Phi\left(-z_\alpha - h\sqrt{J}\right) \le \Phi\left(-z_\alpha - h\sqrt{\tilde{J}}\right), \quad h \le 0,$$

and equality only holds if H_1 is degenerated at 1 (homogeneous alternatives). The power envelope in MPP (see their Theorem 9) is given by the right-hand side of this display and, therefore, it is *not sharp in case of heterogeneous alternatives*. Only if one *could use the unobserved perturbations* $H_1,\dots,H_n$ it is possible to attain this envelope.

(b) The power envelope (11) does not depend on the distribution of the perturbations H_i, i.e., the power envelope for heterogeneous alternatives coincides with the power envelope for homogeneous alternatives. Intuitively, this means potential additional statistical information generated by heterogeneous alternatives (i.e., $\tilde{J} - J > 0$) is not present in the model of interest. This result actually formalizes the following conjecture in Breitung and Pesaran (2008): "This suggests that if there is no information about variation of H_i, then a test cannot be improved by taking into account a possible heterogeneity of the alternative."

(c) The incidental intercepts m_i and scale parameters σ_i do not affect the level of the power envelopes (10) and (11).

The corollary completes the picture for the model in which m_i and σ_i are known. Of course, in empirical applications the parameters m_i and σ_i are

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

(unknown) nuisance parameters; the test $\psi^{\star}_{n,T}$ is infeasible. Nevertheless, (11) still yields an upper bound to the (asymptotic) power of valid tests. In the next subsection we show that this bound is also attainable in case the nuisance parameters m_i and σ_i are unknown. Thus the power envelope is unaffected.

4.2. Asymptotically UMP Tests

In this section we consider the situation where the incidental intercepts m_i and the scale parameters σ_i are nuisance parameters. Throughout, we impose the following additional assumption, which is standard in the panel unit root literature and is needed to control the heterogeneity in σ_i.

Assumption 2. $n/T \to 0$.[5]

Asymptotic optimal tests are not unique. Here we discuss three examples of optimal tests, i.e., their local asymptotic power functions attain the bound (11). Two of these tests are already available in the literature and are known to be point-optimal for homogeneous alternatives. Surprisingly, these tests are also optimal for heterogeneous alternatives. The third optimal test is motivated by the LAN structure. Of course, many other equivalent versions exist; we do not strive for completeness. In fact, the results demonstrate that the Gaussian panel unit root problem is adaptive with respect to the parameters m_i and σ_i. We obtain this result, since the three tests are invariant with respect to these nuisance parameters.

Section 4.2.1 discusses a likelihood ratio based statistic of MPP, Section 4.2.2 discusses a test that is motivated by the LAN structure, and Section 4.2.3 discusses a pooled GLS-based t-test that uses the Breitung and Meyer (1994) device to get rid of the incidental intercepts.

4.2.1. MPP Likelihood Ratio Test. First, we reconsider the test based on the statistic $\hat{V}_{fe1,nT}$ of MPP, page 424. The implementation of this test requires, besides the data Y_{it}, the specification of a sequence C_i, $i \in \mathbb{N}$, of nonnegative random variables. Reason for this is that $\hat{V}_{fe1,nT}$ is based on a (quasi) likelihood ratio statistic for testing the null hypothesis $\rho_i = 1$ versus the alternatives $\rho_i = 1 - C_i/(\sqrt{n}T)$. Under our Assumptions 1 and 2, and some additional regularity, an application of Theorems 9 and 10 in MPP shows, under $\tilde{P}_h^{(n,T)}$,

$$\hat{V}_{fe1,nT} \xrightarrow{d} \mathcal{N}\left(h\mathrm{E}C_1H_1, 2\mathrm{E}C_1^2\right).$$

A combination with Corollary 4.1 yields the following conclusions.

(1) In case one observes both Y_{it} and H_i, i.e., when one is in the auxiliary model, choosing $C_i = H_i$ yields an (asymptotically) optimal test. Because of this property, MPP called the $\hat{V}_{fe1,nT}$-test point optimal.
(2) In the feasible case, where one only observes the panel units Y_{it}, choosing $C_i = c > 0$ yields an (asymptotically) optimal test.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Remark 3. Conclusion (2) is new to the literature. From MPP it follows that the local asymptotic power of the test coincides with the power envelope for homogeneous alternatives. Corollary 4.1 implies that the test is also optimal for heterogeneous alternatives.

4.2.2. Central-Sequence Based Test. As the panel unit root model turned out to be LAN (see Proposition 4.2), we know from general theory that an optimal test can be constructed on the basis of the central-sequence $\Delta_{n,T}$. A complication we have to face is that we cannot observe $\Delta_{n,T}$, so we will use estimates of the scale parameters σ_i instead. This usually affects the asymptotic behavior, but because of adaptivity (see Proposition 4.3) this is not the case here.

We estimate σ_i^2, $i = 1, \ldots, n$, by

$$\hat{\sigma}_{i,nT}^2 = \frac{1}{T-1} \sum_{t=2}^{T} (\Delta Y_{it})^2 \tag{12}$$

and introduce

$$\hat{\Delta}_{n,T} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=3}^{T} \left(\sum_{s=2}^{t-1} \frac{1}{\hat{\sigma}_{i,nT}} \Delta Y_{is} \right) \frac{1}{\hat{\sigma}_{i,nT}} \Delta Y_{it}. \tag{13}$$

Now the general theory on LAN experiments suggests to consider the test statistic

$$t_{\Delta,nT} = J^{-1/2} \hat{\Delta}_{n,T} = \sqrt{2} \hat{\Delta}_{n,T}. \tag{14}$$

PROPOSITION 4.3. *Let Assumptions 1 and 2 hold. Then we have, for all* $h \in \mathbb{R}$ *and under* $P_h^{(n,T)}$ *as* $(T, n \to \infty)$,

$$\hat{\Delta}_{n,T} = \Delta_{n,T} + o_P(1). \tag{15}$$

Remark 4.

(a) From (15) and Corollary 4.1(b), it follows that the test $\psi_{n,T} = 1\{t_{\Delta,nT} \leq -z_\alpha\}$ is asymptotically UMP.

(b) As we have shown that our model is LAN, an application of Le Cam's first lemma shows that we only have to prove (15) under $P_0^{(n,T)}$ (and thus avoid "triangular array calculations").

4.2.3. GLS Variant of the Pooled Test Using the Breitung–Meyer Device. Following Breitung and Meyer (1994), note that (1) and (2) imply

$$Y_{it} - Y_{i1} = \rho_i (Y_{i,t-1} - Y_{i1}) + \sigma_i (\varepsilon_{it} + (\rho_i - 1)\varepsilon_{i1}).$$

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Given estimators $\tilde{\sigma}_{i,nT}$ of the scale parameters σ_i, this leads to the pooled Feasible GLS (FGLS) estimator

$$\hat{\rho}_{n,T} = \frac{\sum_{i=1}^{n}\sum_{t=3}^{T}\frac{1}{\tilde{\sigma}_{i,nT}^2}(Y_{it}-Y_{i1})(Y_{i,t-1}-Y_{i1})}{\sum_{i=1}^{n}\sum_{t=3}^{T}\frac{1}{\tilde{\sigma}_{i,nT}^2}(Y_{i,t-1}-Y_{i1})^2}$$

and the associated t-statistic is given by

$$t_{\mathrm{PGLS},nT} = \left(\sum_{i=1}^{n}\sum_{t=3}^{T}\frac{1}{\tilde{\sigma}_{i,nT}^2}(Y_{i,t-1}-Y_{i1})^2\right)^{-1/2}\left(\hat{\rho}_{n,T}-1\right).$$

Typically Ordinary Least Squares (OLS) residuals, possibly based on a pooled OLS estimator, are used to estimate the scale parameters σ_i in the first stage of the FGLS procedure. However, to link the pooled FGLS estimator to the central-sequence based test statistic defined in (14), we use $\tilde{\sigma}_{i,nT} = \hat{\sigma}_{i,nT}$ instead, see (12). Then, with

$$\begin{aligned}\hat{J}_{n,T} &= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=3}^{T}\left(\frac{1}{\hat{\sigma}_{i,nT}}(Y_{i,t-1}-Y_{i1})\right)^2 \\ &= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=3}^{T}\frac{1}{\hat{\sigma}_{i,nT}^2}\left(\sum_{s=2}^{t-1}\Delta Y_{is}\right)^2, \end{aligned} \tag{16}$$

we have

$$t_{\mathrm{PGLS},nT} = \frac{\hat{\Delta}_{n,T}}{\sqrt{\hat{J}_{n,T}}} = \sqrt{\frac{J}{\hat{J}_{n,T}}}\,t_{\Delta,nT}. \tag{17}$$

The following proposition shows that the estimator $\hat{J}_{n,T}$ is consistent for $J = 1/2$.

PROPOSITION 4.4. *Let Assumptions 1 and 2 hold. Then we have, for all* $h \in \mathbb{R}$ *and under* $P_h^{(n,T)}$ *as* $(T, n \to \infty)$,

$$\hat{J}_{n,T} = J + o_P(1).$$

Remark 5.

(a) A combination of Propositions 4.3 and 4.4 yields, even in case of heterogeneous alternatives, optimality of the pooled GLS test $1\{t_{\mathrm{PGLS},nT} \le -z_\alpha\}$.

(b) We expect this result not to be affected in case residuals from a first stage (possibly pooled) OLS procedure are used to estimate the scale parameters σ_i.

(c) It is important to use the Breitung–Meyer device, i.e., to center the observations in the FGLS regression by the first observation, $Y_{it} \to Y_{it} - Y_{i1}$. Alternatively, centering by the time series means, would, after correcting for bias, basically lead to the Levin, Lin, and Chu (2002) test. This test is not asymptotically optimal, see MPP, page 425.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

4.3. Monte Carlo Experiment

To assess the finite-sample performances of the three tests discussed in Section 4.2, we conduct a small Monte Carlo experiment using the setup of MPP: as data generating process (DGP) we consider (1)–(3) for $h \in \{0, -1, -2, -4\}$ and four distributions for the i.i.d. perturbations: (i) $H_1 \sim U[0, 2]$, (ii) $H_1 \sim \chi_1^2$, (iii) $H_1 \sim \chi_4^2/4$, and (iv) $H_1 = 1$; as sample sizes we consider for each $n \in \{10, 25, 100\}$ three values for T: $T \in \{50, 100, 250\}$. Throughout we use a 5% significance level and 100,000 replications.

Before we turn to the analysis of sizes and powers, we first present the following comments on the implementation.

(a) We implement the (quasi) likelihood ratio test $\hat{V}_{fe1,nT}$ of MPP using $C_i = 1$. And we use $\tilde{\sigma}_{i,nT} = \hat{\sigma}_{i,nT}$ in the implementation of the pooled FGLS test based on the Breitung–Meyer device.

(b) Just as in MPP, we choose m_i i.i.d. $\mathcal{N}(0, 1)$ independent of σ_i i.i.d. $U[1/2, 3/2]$ (both independent of the innovations ε_{it}). But these specifications are, in fact, irrelevant, because the three tests are invariant with respect to the nuisance parameters m_i and σ_i. For the central-sequence based test and the pooled GLS test this is immediate. For the MPP (quasi) likelihood ratio test this can be seen from the decomposition of this test on pages 447 and 449 in MPP.

(ci) For the central-sequence test we use adjusted critical values by replacing $-z_\alpha$ by $(2n)^{-1/2}(F_{\chi_n^2}^{(\mathrm{inv})}(\alpha) - n)$. This is motivated from small n and large T analysis, since

$$t_{\Delta,nT} \xrightarrow{d} \sqrt{2}n^{-1/2} \sum_{i=1}^{n} \int_0^1 W_i(u)\,\mathrm{d}\,W_i(u) \stackrel{d}{=} (2n)^{-1/2}(\chi_n^2 - n) \text{ as } T \to \infty.$$

Since the leading term in the expansion of the MPP test, see MPP, pages 447 and 449, is given by the central-sequence, we consider both $-z_\alpha$ and the adjusted critical value in its implementation.

(cii) For the pooled GLS test one could try to apply a similar procedure, based on

$$t_{PGLS,nT} \xrightarrow{d} \textstyle\sum_{i=1}^{n} \int_0^1 W_i(u)\,\mathrm{d}\,W_i(u) \Big/ \Big(\sum_{i=1}^{n} \int_0^1 W_i^2(u)\,\mathrm{d}u\Big)^{1/2} \text{ as } T \to \infty.$$

However, critical values of this limiting distribution are not readily available for $n > 1$. For $n = 1$ the Dickey–Fuller distribution is obtained.

Table 1 in Appendix B presents sizes: "MPP" refers to the MPP (quasi) likelihood ratio test using $-z_\alpha$ as critical value and "MPP*" uses the adjusted critical value, "CS" refers to the central-sequence based test, and "PGLS" to the pooled GLS procedure. The CS test is slightly conservative, but even for small n, its size is close to the nominal level. The MPP test is more conservative, while MPP* is

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

a bit oversized. The PGLS test is also oversized, this could be expected from the Dickey–Fuller type limiting distribution.

Table 2 presents size-corrected rejection frequencies; thus the results for MPP and MPP* coincide. In view of (b) above the null distribution of the tests does not depend on nuisance parameters: thus, size-corrected frequencies actually correspond to (nominal) rejection frequencies. From our theoretical results we know that the asymptotic powers of the tests are all the same, since they are all asymptotically UMP. Inspecting the rows of Table 2 shows that in finite-samples, also for small n, the differences are small. Asymptotically, the distribution of the perturbations H_i has no effect on the powers. Of course, in finite-samples this distribution can have an effect. For $h = -1$ this effect is small or even negligible, but for $h = -4$ it is more pronounced.

5. CONCLUSION AND DISCUSSION

This paper reconsiders the asymptotic power envelope for the unit root hypothesis in a Gaussian, heterogeneous, cross-sectionally independent panel with incidental intercepts. We present the power envelope for panel unit root tests in case of heterogeneous alternatives and demonstrate that one of the (quasi) likelihood ratio tests of Moon et al. (2007), a pooled GLS test using the Breitung and Meyer (1994) device, and a new test based on the central-sequence are all asymptotically UMP. The results imply, confirming a conjecture in Breitung and Pesaran (2008), that the statistical information due to heterogeneous alternatives cannot be exploited.

Using exact critical values, Monte Carlo results show that small sample power differences are almost negligible. If one wants to avoid simulating critical values, the new central-sequence test is easy to implement with readily available (χ^2 based) critical values having excellent size properties (also for small n).

A natural question is whether the results of this paper carry over to the situation in which there is cross-sectional dependence. If we ignore the possible presence of fixed effects, a cross-sectional dependence version of the model for the observations $Y_{\cdot,t} = (Y_{1t}, \ldots, Y_{nt})'$ is given by $Y_{\cdot,t} = \Xi Y_{\cdot,t-1} + \varepsilon_{\cdot,t}$, $t = 1, \ldots, T$, where $Y_{\cdot,0} = 0_n$ and $\varepsilon_{\cdot,1}, \ldots, \varepsilon_{\cdot,T}$ are i.i.d. $\mathcal{N}(0_n, \Omega_n)$ with Ω_n nonsingular. Let the $n \times n$ matrix of autoregression coefficients be generated by the random coefficient structure $\Xi = I_n + \frac{h}{\sqrt{nT}} \Omega_n^{1/2} \mathcal{H}_n \Omega_n^{-1/2}$, where $\mathcal{H}_n = \mathrm{diag}(H_1, \ldots, H_n)$ and the random perturbations H_i satisfy Assumption 1(b). For this setting it is straightforward, by rotating the model back to the cross-section independence case, to verify that (11) yields an upper bound to the asymptotic power.

NOTES

1. We wish to stress that the limit experiment of the panel unit root model (Gaussian shift, LAN) is fundamentally different from the limit experiment for the univariate unit root problem (which is a Brownian functional experiment). This is also the reason that an asymptotically uniformly most powerful test does exist for the panel unit root model, while this is not the case in the univariate situation.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

2. Our results can also be used to derive the (asymptotic) power envelope for the two-sided alternative $H_a : h \neq 0$ or to test a unit root versus an explosive alternative, i.e., $H_a : h > 0$.

3. We adopt the convention that an empty sum is equal to 0.

4. Local Asymptotic Normality and convergence of experiments is in most books (see, for example, Van der Vaart, 2000) presented for 'univariate' limits $n \to \infty$. The abstract theory, however, is formulated in terms of nets and the results can thus be applied in the joint-limit setting as well; see, for example, Van der Vaart (1991).

5. Proposition 4.2 is, of course, not affected by Assumption 2.

REFERENCES

Baltagi, B.H. & C. Kao (2000) Nonstationary panels, cointegration in panels and dynamic panels, a survey. In B. Baltagi (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. *Advances in Econometrics*, vol. 15, pp. 7–52. Amsterdam.

Banerjee, A. (1999) Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics* 61, 607–629.

Breitung, J. (2000) The local power of some unit root tests for panel data. In B. Baltagi (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. *Advances in Econometrics*, vol. 15, pp. 161–178. Amsterdam.

Breitung, J. & S. Das (2005) Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica* 59, 414–433.

Breitung, J. & W. Meyer (1994) Testing for unit roots in panel data: Are wages on different bargaining levels cointegrated? *Applied Economics* 25, 353–361.

Breitung, J. & M.H. Pesaran (2008) Unit roots and cointegration in panels. In L. Mátyás & P. Sevestre (eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Kluwer.

Choi, I. (2006) Nonstationary panels. In K. Patterson & T.C. Mills (eds.), *Palgrave Handbook of Econometrics* 1, pp. 511–539. Palgrave Macmillan.

Drost, F.C., R. van den Akker, & B.J.M. Werker (2009) The asymptotic structure of nearly unstable non negative integer-valued AR(1) models. *Bernoulli* 15, 297–324.

Elliott, G. (1999) Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review* 40, 767–784.

Hallin, M., R. van den Akker, & B.J.M. Werker (2011) A class of simple distribution-free rank-based unit root tests. *Journal of Econometrics* 163, 200–214.

Harris, D., D.I. Harvey, S.J. Leybourne, & N.D. Sakkas (2010) Local asymptotic power of the Im-Pesaran-Shin panel unit root test and the impact of initial observations. *Econometric Theory* 26, 311–324.

Im, K.S., M.H. Pesaran, & Y. Shin (2003) Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115, 53–74.

Jansson, M. (2008) Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica* 76, 1103–1142.

Le Cam, L. & G.L. Yang (1988) On the preservation of local asymptotic normality under information loss. *Annals of Statistics* 16, 483–520.

Levin, A., C.-F. Lin, & C.-S.J. Chu (2002) Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.

Madsen, E. (2010) Unit root inference in panel data models where the time-series dimension is fixed: A comparison of different tests. *Econometrics Journal* 13, 63–94.

Moon, H.R. & B. Perron (2004) Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122, 81–126.

Moon, H.R. & B. Perron (2008) Asymptotic local power of pooled *t*-ratio tests for unit roots in panels with fixed effects. *Econometrics Journal* 11, 80–104.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Moon, H.R., B. Perron, & P.C.B. Phillips (2007) Incidental trends and the power of panel unit root tests. *Journal of Econometrics* 141, 416–459.
Müller, U.K. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.
Phillips, P.C.B. & H.R. Moon (1999) Limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
Van der Vaart, A.W. (1991) An asymptotic representation theorem. *International Statistical Review* 59, 97–121.
Van der Vaart, A.W. (2000) *Asymptotic Statistics*. Cambridge University Press.
Westerlund, J. & J. Breitung (2013) Lessons from a decade of IPS and LLC. *Econometric Reviews* 32, 547–591.
Westerlund, J. & R. Larsson (2012) Testing for a unit root in a random coefficient panel data model. *Journal of Econometrics* 167, 254–273.

APPENDIX

A. Proofs

A.1. Proof of Lemma 3.1

The result is immediate if $\sigma_U^2 = \text{var}(U_1) = 0$. So we consider $\sigma_U^2 > 0$. Let, for some $\tilde{\eta} > 0$, ϕ denote the moment generating function of $U_1 - \text{E}(U_1)$ on the interval $(-\tilde{\eta}, \tilde{\eta})$. We have $m_n = m_{n1} m_{n2}$ for

$$m_{n1} = \exp\left(-\frac{\sigma_U^2}{2n}\sum_{i=1}^{n}\mathcal{J}_{ni}\right),$$

$$m_{n2} = \text{E}\left[\exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i - \text{E}U_1)\mathcal{X}_{ni}\right)\exp\left(-\frac{1}{2n}\sum_{i=1}^{n}(U_i^2 - \text{E}U_1^2)\mathcal{J}_{ni}\right)\middle|\,\mathcal{G}_n\right].$$

Introduce, for $a \in \mathbb{R}$,

$$M_n(a) = \text{E}\left[\exp\left(\frac{a}{\sqrt{n}}\sum_{i=1}^{n}(U_i - \text{E}U_i)\mathcal{X}_{ni}\right)\middle|\,\mathcal{G}_n\right],$$

and decompose $m_n = m_{n1}m_{n2} = m_{n1}M_n(1) + m_{n1}(m_{n2} - M_n(1))$. To enhance readability we organize the proof of $m_n \xrightarrow{p} 1$ in two steps. In Step A we show $m_{n1}M_n(1) \xrightarrow{p} 1$, and in Step B we establish $m_{n2} - M_n(1) \xrightarrow{p} 0$. This also yields $m_{n1}(m_{n2} - M_n(1)) \xrightarrow{p} 0$ since $m_{n1} \leq 1$.

Step A By (6) it suffices (as $\phi'(0) = 0$ and $\phi''(0) = \sigma_U^2$) to show $\log(M_n(1)) - \frac{1}{2}\phi''(0)\frac{1}{n}\sum\mathcal{X}_{ni}^2 \xrightarrow{p} 0$. Let $\epsilon > 0$ arbitrary. Choose $\eta \in (0, \tilde{\eta})$ such that, for all $|x| \leq \eta$,

$$|\log(1+x) - x| \leq \epsilon|x|,$$

$$\left|\phi(x) - 1 - \frac{1}{2}\phi''(0)x^2\right| \leq \epsilon x^2,$$

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

and $(\frac{1}{2}\phi''(0)+\epsilon)\eta \le 1$. Then on the event $\mathcal{A}_n = \{\max_{i=1,\dots,n} |\mathcal{X}_{ni}|/\sqrt{n} \le \eta\}$,

$$\left|\phi\left(\frac{\mathcal{X}_{ni}}{\sqrt{n}}\right)-1\right| \le \left(\frac{1}{2}\phi''(0)+\epsilon\right)\frac{\mathcal{X}_{ni}^2}{n} \le \left(\frac{1}{2}\phi''(0)+\epsilon\right)\eta^2 \le \eta.$$

Thus, still on the event $\mathcal{A}_n$,

$$\begin{aligned}\left|\log(M_n(1)) - \frac{1}{2}\phi''(0)\sum_{i=1}^n \frac{\mathcal{X}_{ni}^2}{n}\right| &\le \sum_{i=1}^n \left|\log\left(\phi\left(\frac{\mathcal{X}_{ni}}{\sqrt{n}}\right)\right) - \frac{1}{2}\phi''(0)\frac{\mathcal{X}_{ni}^2}{n}\right| \\ &= \sum_{i=1}^n \left|\log\left(\phi\left(\frac{\mathcal{X}_{ni}}{\sqrt{n}}\right)\right) - \left\{\phi\left(\frac{\mathcal{X}_{ni}}{\sqrt{n}}\right)-1\right\}\right| \\ &\quad + \sum_{i=1}^n \left|\phi\left(\frac{\mathcal{X}_{ni}}{\sqrt{n}}\right) - 1 - \frac{1}{2}\phi''(0)\frac{\mathcal{X}_{ni}^2}{n}\right| \\ &\le \epsilon\left(1+\frac{1}{2}\phi''(0)+\epsilon\right)\frac{1}{n}\sum_{i=1}^n \mathcal{X}_{ni}^2.\end{aligned}$$

Using $n^{-1}\sum_{i=1}^n \mathcal{X}_{ni}^2 = O_P(1)$ and $\mathbb{P}(\mathcal{A}_n) \to 1$, we obtain the desired convergence to zero.

Step B For each $\epsilon > 0$ choose $M > 0$ and N such that, for the event

$$\mathcal{B}_n = \left\{\frac{1}{n}\sum_{i=1}^n \mathcal{J}_{ni} \le M\right\} \cap \left\{\max_{i=1,\dots,n}\frac{\mathcal{J}_{ni}}{n} \le 1\right\} \cap \left\{\max_{i=1,\dots,n}\frac{|\mathcal{X}_{ni}|}{\sqrt{n}} < \tilde{\eta}/2\right\} \in \mathcal{G}_n,$$

we have $\mathbb{P}(\mathcal{B}_n^c) \le \epsilon$ when $n \ge N$. Part B follows if $|m_{n2} - M_n(1)|1_{\mathcal{B}_n} \xrightarrow{p} 0$. We have, using Cauchy–Schwarz,

$$\mathrm{E}|m_{n2} - M_n(1)|1_{\mathcal{B}_n} \le \sqrt{\mathrm{E}r_n 1_{\mathcal{B}_n}}\sqrt{\mathrm{E}M_n(2)1_{\mathcal{B}_n}},$$

where

$$r_n = \mathrm{E}\left[\left(\exp\left(-\frac{1}{2n}\sum_{i=1}^n (U_i^2 - \mathrm{E}U_1^2)\mathcal{J}_{ni}\right) - 1\right)^2 \middle| \mathcal{G}_n\right].$$

On the event $\mathcal{B}_n$ $M_n(2)$ is uniformly bounded (which follows by similar arguments as in Step A), so $\mathrm{E}M_n(2)1_{\mathcal{B}_n}$ is bounded in n. We show that $\mathrm{E}r_n 1_{\mathcal{B}_n} \to 0$. The conditional Markov inequality implies, for all $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (U_i^2 - \mathrm{E}U_1^2)\mathcal{J}_{ni}\right| > \delta \middle| \mathcal{G}_n\right) \le \frac{1}{\delta^2}\mathrm{var}(U_1^2)\left(\max_{i=1,\dots,n}\frac{|\mathcal{J}_{ni}|}{n}\right)\frac{1}{n}\sum_{i=1}^n \mathcal{J}_{ni},$$

which converges to 0 in probability by (5) and (6). As $r_n 1_{\mathcal{B}_n}$ is uniformly bounded an application of the continuous mapping theorem, the law of iterated expectations, and the bounded convergence theorem yields $\mathrm{E}r_n 1_{\mathcal{B}_n} \to 0$. Conclude that $\mathrm{E}|m_{n2} - M_n(1)|1_{\mathcal{B}_n} \to 0$, which concludes Step B and the proof. ■

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

A.2. Proof of Proposition 4.1

We have

$$\log \frac{\mathrm{d}\tilde{P}_{h_n}^{(n,T)}}{\mathrm{d}\tilde{P}_0^{(n,T)}} = \frac{h_n}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{H_i}{\sigma_i^2}\left(Y_{i,t-1}-m_i\right)\Delta Y_{it} - \frac{h_n^2}{2}\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{H_i^2}{\sigma_i^2}\left(Y_{i,t-1}-m_i\right)^2 = h_n\tilde{\Delta}_{n,T}-\frac{1}{2}h_n^2\tilde{J}_{n,T}.$$

An application of the following lemma (with $U_i = H_i$) yields, under $\tilde{P}_0^{(n,T)}$, $\tilde{\Delta}_{n,T} \xrightarrow{d} \mathcal{N}\left(0,\tilde{J}\right)$ and $\tilde{J}_{n,T} \xrightarrow{p} \tilde{J} = \frac{1}{2}\mathrm{E}H_1^2$ as $(T,n\to\infty)$. An application of Le Cam's third and first lemma yields, under $\tilde{P}_{h_n}^{(n,T)}$, $\tilde{\Delta}_{n,T} \xrightarrow{d} \mathcal{N}\left(h\tilde{J},\tilde{J}\right)$ and $\tilde{J}_{n,T} \xrightarrow{p} \tilde{J}$ as $(T,n\to\infty)$. This completes the proof. ■

LEMMA A.1. *Let Assumption 1 hold and let U_i be i.i.d. with mean 1 and finite fourth moment independent of ε_{it}, $i,t\in\mathbb{N}$. Then we have, with $X_i^{(T)}$ and $J_i^{(T)}$ as defined in Proposition 4.1 and Y_{it} generated under $h=0$, and as $(T,n\to\infty)$ and for any $\delta>0$,*

$$\operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}U_i^2J_i^{(T)} = \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}U_i^2\left(X_i^{(T)}\right)^2 = \frac{1}{2}\mathrm{E}U_1^2, \tag{A.1}$$

$$\lim \mathrm{E}U_1^2\left(X_1^{(T)}\right)^2 1\left\{\left|U_1X_1^{(T)}\right|>\delta\sqrt{n}\right\}=0, \tag{A.2}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_iX_i^{(T)} \xrightarrow{d} \mathcal{N}(0,\mathrm{E}U_1^2/2). \tag{A.3}$$

Proof. The convergence in probability (A.1) follows from $\mathrm{E}U_1^2J_1^{(T)} = \mathrm{E}U_1^2\left(X_1^{(T)}\right)^2 = [T(T-1)/(2T^2)]\mathrm{E}U_i^2$ and an application of the Markov inequality (using that $\operatorname{var}\left(X_1^{(T)}\right)^2$ and $\operatorname{var}J_1^{(T)}$ are both bounded and using that U_1 has a finite fourth moment).

Let $\delta>0$. Note that $I_T := \sum_{j=1}^{T}B(t_{j-1}^{(T)})\left(B(t_j^{(T)})-B(t_{j-1}^{(T)})\right)$, where $t_j^{(T)} = j/T$ and B is a standard Brownian motion, and $X_1^{(T)}$ have the same distribution. As $T\to\infty$ we have $I_T \to \int_0^1 B(u)\,\mathrm{d}B(u) =: I$ in L_2. This implies $I_T^2 \to I^2$ in L_1. Hence the collection of random variables $\{I_T^2, T\in\mathbb{N}\}$ is uniformly integrable. As uniform integrability of a collection of random variables is determined by the marginal laws, we can conclude that the sequence $\left\{\left(X_1^{(T)}\right)^2, T\in\mathbb{N}\right\}$ is uniformly integrable. This yields $\mathrm{E}\left[\left(X_1^{(T)}\right)^2 1\left\{\left|U_1X_1^{(T)}\right|>\delta\sqrt{n}\right\}\mid U_1\right]\to 0$. As this conditional expectation is also

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

bounded (we already noted that $\mathrm{E}\left(X_1^{(T)}\right)^2 = O(1)$) an application of the dominated convergence theorem yields (A.2).

As the $U_i X_i^{(T)}$ are i.i.d. with mean zero and the Lindeberg condition (A.2) holds, an application of the joint-limit central limit theorem (see Phillips and Moon, 1999, Thm. 2) yields (A.3). ■

A.3. Proof of Proposition 4.2

We first verify the conditions to Lemma 3.1. Recall that $\mathcal{G}_n = \mathcal{F}_{n,T(n)}$, $\mathcal{X}_{ni} = X_i^{(T(n))}$, $\mathcal{J}_{ni} = J_i^{(T(n))}$, and $U_i = H_i$. Condition i) is satisfied because of Assumption 1. The first part of (5) and the first part of (6) follow from Lemma A.1 (using $U_i = 1$). The second part of (5) is a consequence of (A.2) (using $U_i = 1$). And the second part of (6) follows from, for $\epsilon > 0$ and with probabilities and expectations calculated under $P_0^{(n,T(n))}$,

$$
\begin{aligned}
P\left(\max_{i=1,\ldots,n} \frac{J_i^{(T(n))}}{n} > \epsilon\right) &\leq \frac{1}{n}\sum_{i=1}^{n} \mathrm{E} J_i^{(T(n))} 1\left\{J_i^{(T(n))} > \epsilon n\right\} \\
&= \mathrm{E}\left[\int_0^1 \left(W_1^{(T(n))}(u-)\right)^2 \mathrm{d}u 1 \right. \\
&\qquad \left. \times \left\{\int_0^1 \left(W_1^{(T(n))}(u-)\right)^2 \mathrm{d}u > \epsilon n\right\}\right] \\
&\to 0,
\end{aligned}
$$

which follows along similar lines as used in the proof of Lemma A.1. The lemma thus yields, under $P_0^{(n)}$,

$$
\begin{aligned}
\log \frac{\mathrm{d}P_{h_n}^{(n)}}{\mathrm{d}P_0^{(n)}} &= \log \mathrm{E}\left[\exp\left(\frac{h_n}{\sqrt{n}}\sum_{i=1}^{n} H_i X_i^{(T(n)} - \frac{h_n^2}{2n}\sum_{i=1}^{n} H_i^2 J_i^{(T(n))}\right)\middle| \mathcal{F}_n\right] \\
&= \frac{h_n}{\sqrt{n}}\sum_{i=1}^{n} X_i^{(T(n))} - \frac{1}{2}\frac{h_n^2}{n}\sum_{i=1}^{n} J_i^{(T(n))} + o_P(1) = h_n \Delta_n - \frac{h_n^2}{2} J_{n,T(n)} \\
&\quad + o_P(1).
\end{aligned}
$$

Another application of Lemma A.1 (with $U_i = 1$) yields $\Delta_n \xrightarrow{d} \mathcal{N}(0, J)$ and $J_{n,T(n)} \xrightarrow{p} J = 1/2$. An application of Le Cam's third and first lemma yields, under $\tilde{P}_{h_n}^{(n,T)}$, $\tilde{\Delta}_{n,T} \xrightarrow{d} \mathcal{N}(hJ, J)$ and $J_{n,T} \xrightarrow{p} J$ as $(T, n \to \infty)$. This concludes the proof. ■

A.4. Proof of Proposition 4.3

As noted in Remark 4 we only need to prove (15) for $h = 0$. In the following, all probabilities and expectations are evaluated under $P_0^{(n,T)}$.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Decompose $\hat{\Delta}_{n,T} - \Delta_{n,T} = I_{n,T} - II_{n,T}$ with

$$I_{n,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} - 1 \right) \int_0^1 W_i^{(T)}(u-) \, \mathrm{d} W_i^{(T)}(u),$$

$$II_{n,T} = \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \varepsilon_{i1} \sum_{t=2}^{T} \varepsilon_{it} + \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(\frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} - 1 \right) \varepsilon_{i1} \sum_{t=2}^{T} \varepsilon_{it}.$$

To see that $I_{n,T}$ converges to 0 in probability we first apply Cauchy–Schwarz to obtain

$$\left| I_{n,T} \right|^2 \leq \sum_{i=1}^{n} \left(\frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} - 1 \right)^2 \times \frac{1}{n} \sum_{i=1}^{n} \left(\int_0^1 W_i^{(T)}(u-) \, \mathrm{d} W_i^{(T)}(u) \right)^2.$$

As the second term is $O_P(1)$, $I_{n,T} = o_P(1)$ follows from

$$\sum_{i=1}^{n} \left(1 - \frac{\hat{\sigma}_{i,nT}^2}{\sigma_i^2} \right)^2 = o_P(1) \text{ and } \lim P \left(\sup_{1 \leq i \leq n} \frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} \leq 3/2 \right) = 1, \tag{A.4}$$

where the second statement in the display follows from the first one, and the first statement follows from

$$\mathrm{E} \left[\sum_{i=1}^{n} \left(\frac{\hat{\sigma}_{i,nT}^2}{\sigma_i^2} - 1 \right)^2 \right] = \sum_{i=1}^{n} \mathrm{var} \left(\frac{1}{T-1} \sum_{t=2}^{T} \varepsilon_{it}^2 \right) = 2 \frac{n}{T-1} \to 0.$$

As the first term of $II_{n,T}$ is centered and its variance, given by $n(T-1)/(nT^2)$, tends to 0 it follows that this term converges to 0 in probability. The second term of $II_{n,T}$ is seen to convergence to 0 in probability by similar arguments as for $I_{n,T}$. This completes the proof. ■

A.5. Proof of Proposition 4.4

By an application of Le Cam's first lemma we only have to prove the result under $P_0^{(n,T)}$. In the following, all probabilities and expectations are evaluated under $P_0^{(n,T)}$. Recall, see Proposition 4.2, $J_{n,T} = n^{-1} \sum_{i=1}^{n} J_i^{(T)} = n^{-1} \sum_{i=1}^{n} \int_0^1 \left(W_i^{(T)}(u-) \right)^2 \mathrm{d}u \xrightarrow{p} 1/2$. We have $\hat{J}_{n,T} - J_{n,T} = I_{n,T} - II_{n,T} - 2III_{n,T}$ with

$$I_{n,T} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} - 1 \right) J_i^{(T)}, \qquad II_{n,T} = \frac{T-1}{nT^2} \sum_{i=1}^{n} \frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} \varepsilon_{i1}^2,$$

$$III_{n,T} = \frac{1}{nT^2} \sum_{i=1}^{n} \frac{\sigma_i^2}{\hat{\sigma}_{i,nT}^2} \varepsilon_{i1} \sum_{t=2}^{T} \sum_{s=2}^{t-1} \varepsilon_{is}.$$

Since $n^{-1} \sum_{i=1}^{n} (J_i^{(T)})^2 = O_P(1)$, Cauchy–Schwarz and (A.4) yield $I_{n,T} = o_P(1)$. By bounding the L_1-norm it is easy to see that $II_{n,T} 1\{\sup_{1 \leq i \leq n} \sigma_i^2 / \hat{\sigma}_{i,nT}^2 \leq 3/2\} = o_P(1)$.

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

Invoking (A.4) thus yields $II_{n,T} = o_P(1)$. Finally, for $III_{n,T}$ we obtain

$$III_{n,T}^2 1\{\sup_{1\le i\le n} \sigma_i^2/\hat{\sigma}_{i,nT}^2 \le 3/2\} \le \frac{9}{4}\frac{1}{nT^4}\sum_{i=1}^{n}\varepsilon_{i1}^2\left(\sum_{t=2}^{T}\sum_{s=2}^{t-1}\varepsilon_{is}\right)^2.$$

The L_1-norm of the right-hand side is $o(1)$, so another application of (A.4) yields $III_{n,T} = o_P(1)$, which completes the proof. ■

B. Tables

TABLE 1. Sizes, based on 100,000 replications, of the MPP, MPP*, CS, and PGLS tests

n	T	MPP	MPP*	CS	PGLS
10	50	.028	.078	.042	.067
25	50	.038	.062	.042	.058
100	50	.043	.053	.041	.051
10	100	.029	.081	.047	.069
25	100	.039	.064	.046	.060
100	100	.045	.055	.046	.054
10	250	.028	.078	.047	.068
25	250	.040	.064	.048	.060
100	250	.046	.056	.048	.055
∞	∞	.05	.05	.05	.05

TABLE 2. Size-corrected powers, based on 100,000 replications, of the MPP, CS, and PGLS tests

$H_1 \sim U[0,2]$										
		$h=-1$			$h=-2$			$h=-4$		
n	T	MPP	CS	PGLS	MPP	CS	PGLS	MPP	CS	PGLS
10	50	.124	.124	.124	.237	.234	.239	.485	.472	.501
25	50	.137	.137	.137	.279	.277	.281	.601	.594	.610
100	50	.149	.149	.150	.327	.325	.328	.731	.729	.735
10	100	.126	.125	.124	.241	.237	.242	.500	.484	.515
25	100	.140	.140	.141	.288	.287	.288	.620	.613	.627
100	100	.154	.154	.153	.336	.336	.335	.747	.746	.749
10	250	.128	.128	.127	.250	.246	.251	.519	.504	.533
25	250	.140	.141	.141	.291	.288	.293	.631	.624	.639
100	250	.153	.153	.154	.341	.340	.341	.756	.755	.759
∞	∞	.174	.174	.174	.409	.409	.409	.882	.882	.882

Table continues on overleaf

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press

TABLE 2. *continued*

$H_1 \sim \chi_1^2$

		$h=-1$			$h=-2$			$h=-4$		
n	T	MPP	CS	PGLS	MPP	CS	PGLS	MPP	CS	PGLS
10	50	.113	.113	.114	.183	.180	.186	.307	.298	.319
25	50	.125	.124	.125	.216	.214	.218	.393	.386	.400
100	50	.139	.139	.140	.270	.269	.272	.546	.543	.551
10	100	.114	.114	.115	.186	.182	.189	.312	.301	.324
25	100	.128	.128	.128	.222	.220	.224	.403	.396	.410
100	100	.144	.145	.143	.277	.277	.277	.558	.557	.561
10	250	.116	.116	.116	.191	.187	.193	.323	.312	.334
25	250	.127	.127	.129	.222	.220	.226	.407	.401	.415
100	250	.143	.142	.143	.279	.278	.280	.564	.561	.568
∞	∞	.174	.174	.174	.409	.409	.409	.882	.882	.882

$H_1 \sim \chi_4^2/4$

		$h=-1$			$h=-2$			$h=-4$		
n	T	MPP	CS	PGLS	MPP	CS	PGLS	MPP	CS	PGLS
10	50	.123	.123	.123	.230	.227	.232	.468	.455	.483
25	50	.136	.136	.136	.271	.270	.273	.581	.575	.590
100	50	.149	.148	.149	.321	.320	.322	.712	.710	.715
10	100	.124	.124	.123	.233	.230	.235	.482	.467	.496
25	100	.139	.139	.140	.279	.279	.281	.599	.594	.606
100	100	.153	.154	.153	.328	.329	.328	.729	.728	.732
10	250	.126	.126	.126	.242	.239	.243	.501	.486	.513
25	250	.139	.139	.139	.282	.280	.285	.610	.604	.619
100	250	.152	.152	.152	.333	.333	.334	.739	.737	.741
∞	∞	.174	.174	.174	.409	.409	.409	.882	.882	.882

$H_1 = 1$

		$h=-1$			$h=-2$			$h=-4$		
n	T	MPP	CS	PGLS	MPP	CS	PGLS	MPP	CS	PGLS
10	50	.126	.126	.125	.255	.252	.256	.597	.580	.612
25	50	.141	.140	.141	.302	.301	.303	.711	.703	.717
100	50	.152	.152	.153	.344	.343	.345	.791	.790	.794
10	100	.128	.127	.126	.261	.258	.261	.617	.599	.633
25	100	.144	.145	.145	.312	.311	.312	.731	.725	.737
100	100	.156	.157	.156	.353	.354	.353	.809	.808	.810
10	250	.129	.130	.129	.272	.269	.270	.644	.626	.657
25	250	.143	.144	.144	.317	.315	.318	.746	.740	.754
100	250	.156	.156	.157	.359	.358	.359	.821	.819	.822
∞	∞	.174	.174	.174	.409	.409	.409	.882	.882	.882

https://doi.org/10.1017/S0266466614000656 Published online by Cambridge University Press