STATISTICAL INFERENCE IN A RANDOM COEFFICIENT PANEL MODEL

LAJOS HORVÁTH AND LORENZO TRAPANI

ABSTRACT. This paper studies the asymptotics of the Weighted Least Squares (WLS) estimator of the autoregressive root in a panel Random Coefficient Autoregression (RCA). We show that, in an RCA context, there is no "unit root problem": the WLS estimator is always asymptotically normal, irrespective of the average value of the autoregressive root, of whether the autoregressive coefficient is random or not, and of the presence and degree of cross dependence. Our simulations indicate that the estimator has good properties, and that confidence intervals have the correct coverage even for sample sizes as small as $(N, T) = (10, 25)$. We illustrate our findings through two applications to macroeconomic and financial variables.

Keywords: Random Coefficient Autoregression, Panel Data, WLS estimator, common factors

JEL Codes: C13, C23.

1. INTRODUCTION

In this paper, we study the asymptotics for the Weighted Least Squares (WLS) estimator of the autoregressive coefficient φ in the following Random Coefficient Autoregressive (RCA) panel model:

$$(1.1) \qquad y_{i,t} = (\varphi + b_{i,t}) y_{i,t-1} + u_{i,t}, \text{ with } 1 \leq t \leq T \text{ and } 1 \leq i \leq N.$$

In a time series setting, RCA models have been popular for a very long time, chiefly due to their flexibility and analytical tractability - we refer to the monograph by Nicholls and Quinn (1983) for an excellent introduction to the topic, mainly in the field of biostatistics, and to the contributions by Swamy (1970), Feige and Swamy (1974), and Hsiao (1975).

Recently, also due to the increasing availability of large datasets, models with random coefficient have been applied in the context of panel data analysis (see Hsiao and Pesaran, 2004). Although slope heterogeneity may be desirable in a panel context, a specification with fixed heterogeneous slopes may yield a loss of efficiency due to the penalty it imposes onto the degree of freedom. This is evident e.g. in the context of forecasting, where several studies by Baltagi and his co-authors point out that, whilst models with homogeneous slopes are often rejected by the data, they could however yield superior predictive performances - see e.g. Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson and Pirotte (2002) and Baltagi, Bresson, Griffin and Pirotte (2002). This explains the importance of

models that, on the one hand, do not impose the restriction of slope homogeneity but, on the other hand, do not overparameterise the model. Similar considerations also hold for the panel unit root literature (see e.g. the survey by Breitung and Pesaran, 2008): whilst earlier contributions proposed tests for the null that $\varphi = 1$ and $b_{i,t} = 0$ for all i and t under the null, with $|\varphi| < 1$ and $b_{i,t} = 0$ under the alternative, more recent contributions consider the same null hypothesis but entertain the possibility that, under the alternative, φ differs across units. However, in this case the model becomes overparameterised under the alternative. A possible solution is to use the RCA panel model (1.1): this was originally proposed by Ng (2008), in order to estimate the fraction of units that have a unit root, and it was subsequently fully exploited, to construct a test for unit root, by Westerlund and Larsson (2012). Of course, the dynamics of $y_{i,t}$ in (1.1) is controlled by $\varphi + b_{i,t}$, and not by φ alone. Thus, when $Var(b_{i,t}) > 0$, having for example $\varphi = 1$ does not mean having a unit root; rather, as illustrated in the seminal contribution by Granger and Swanson (1997), it means having a "stochastic unit root" series, i.e. a series which is non-stationary on average, having periods of explosive and stationary dynamics.[1] Results on the estimation of φ are already available in a time series context: Aue *et al.* (2006) and Berkes *et al.* (2009) use the quasi-maximum likelihood (QML) method to estimate the regression coefficient, showing that the asymptotic distribution of the estimated φ is normal irrespective of the value of φ, as long as $Eb_{i,t}^2 > 0$. Hill and Peng (2014) propose an Empirical Likelihood (EL) estimator which affords standard normal inference even when $b_{i,t} = 0$ - that is, even when there is no coefficient randomness. However, in a large panel context, both approaches could be problematic, due to computational issues (the QML estimator), or to the possible presence of dependence across units (the EL estimator).

Hence, in this contribution we complement the existing results in the literature by developing a full-fledged asymptotic theory for the WLS estimator of φ in (1.1). The WLS estimator avoids the issues described above, and it has been shown, in a time series context, to work well in comparison with the maximum likelihood approach (Schick, 1996; Koul and Schick, 1996). We show that in general, thanks to the self-normalised nature of the WLS estimator, there is no "unit root problem" in case of the RCA model defined by (1.1). Further, we study an estimator of the asymptotic variance of the WLS estimator which is consistent under any degree of cross sectional dependence; hence, it is possible to normalise the estimated φ and recover standard normal inference, with no need for any prior inference on the possible presence of a factor structure. We also show that, contrary to the time series case, the suitably normalised WLS estimator converges to a normal distribution even when the autoregressive root is not random, i.e. when $b_{i,t} = 0$ for all $1 \leq i \leq N$ and $1 \leq t \leq T$; the only case in which asymptotic normality does not hold is the case of a constant explosive root φ in presence of strong common factors. From a

---

[1]We are grateful to an anonymous Referee for pointing this interpretation out to us.

technical point of view, the asymptotics derived in this paper requires that $\min \{N, T\} \to \infty$. Earlier studies in the context of dynamic panels (e.g. Arellano and Bond, 1991) focused on the case of fixed T and $N \to \infty$. In such context, having $|\varphi| > 1$ is inconsequential. Thus, as far as the applicability of our setup is concerned, the unified estimation theory developed here is needed only when $T \to \infty$.

Our findings have important implications in various areas where panel data are routinely employed. In the context of a standard dynamic panel model, the RCA model avoids, as is well known, the heterogeneity bias problem (Swamy, 1970); moreover, not having the unit root problem is advantageous *per se*. Thus, our results are bound to prove useful, for example, in the context of panel studies of hyperinflation (Juselius and Mladenovic, 2002), or in the literature on bubbles (Banerjee *et al.*, 2012).

The remainder of the paper is organised as follows. We lay out the model and the assumptions, and discuss the estimation techniques in Section 2. The asymptotics is reported in Section 3. Particularly, in Section 3.1 we show that the WLS estimator is consistent; in Section 3.2 we study the limiting distribution. Extensions to the cases of more complex models are reported in Section 4. The properties of the WLS estimator are illustrated through a Monte Carlo exercise (Section 5), and a set of empirical applications (Section 6). Section 7 concludes. Technical Lemmas and the proofs of the results in Section 3 are in Appendix; all other proofs are in the Supplement (Horváth and Trapani, 2016).

NOTATION. We denote the ordinary limit as "$\to$"; convergence in probability with "$\xrightarrow{P}$"; convergence in distribution as "$\xrightarrow{\mathcal{D}}$"; $N(\mu, a)$ denotes a normally distributed random variable with mean μ and variance a; "$\equiv$" denotes definitional equality; $\|\cdot\|$ denotes the Euclidean norm. Other notation is introduced further ahead in the paper, when needed.

2. Estimation and main assumptions

In this section, we introduce the WLS estimator of φ, and we spell out the main assumptions needed for the asymptotics. Recall model (1.1), given by

$$y_{i,t} = (\varphi + b_{i,t}) \, y_{i,t-1} + u_{i,t}.$$

As far as the error term $u_{i,t}$ is concerned, we consider a factor structure to capture the possible presence of strong cross sectional dependence. Following Ng (2008) we write

(2.1) $$u_{i,t} = e_{i,t} + \gamma_i v_t, \quad 1 \le i \le N, 1 \le t \le T,$$

i.e. $u_{i,t}$ is decomposed into two terms: $e_{i,t}$ depends only on unit i while the term v_t is common for all panels.

The WLS estimator, $\hat{\varphi}$, is the solution to the minimisation problem (see Janečková and Prašková, 2003)

$$(2.2) \qquad \min_s \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{(y_{i,t} - sy_{i,t-1})^2}{1 + y_{i,t-1}^2} = \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{(y_{i,t} - \hat{\varphi}y_{i,t-1})^2}{1 + y_{i,t-1}^2}.$$

The weighing function in (2.2) is not designed to attain efficiency. Indeed, this would require employing, as weights, $1/\left[\left(Eu_{i,t}^2\right) + \left(Eb_{i,t}^2\right)y_{i,t-1}^2\right]$, or a feasible version thereof; however, as shown in Chan, Li and Peng (2012), first order asymptotics holds when setting e.g. $Eu_{i,t}^2 = Eb_{i,t}^2 = 1$, so that it can be expected that consistency will be ensured when using (2.2). Define, for convenience,

$$\hat{\varphi} \equiv \hat{\varphi}_{N,T} = \frac{A_{N,T}}{B_{N,T}},$$

with

$$(2.3) \qquad A_{N,T} \equiv \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{y_{i,t}y_{i,t-1}}{1 + y_{i,t-1}^2} \quad \text{and} \quad B_{N,T} \equiv \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{y_{i,t-1}^2}{1 + y_{i,t-1}^2}.$$

Consider the following assumption:

Assumption 1. It holds that: *(i)* (a) for every $i = 1, ..., N$, $\{b_{i,t}, -\infty < t < \infty\}$ is *i.i.d.* across t; (b) for every $i = 1, ..., N$, $\{e_{i,t}, -\infty < t < \infty\}$ is *i.i.d.* across t; (c) $\{v_t, -\infty < t < \infty\}$ is *i.i.d.* across t; *(ii)* (a) for every $i = 1, ..., N$, $E(b_{i,0}) = E(e_{i,0}) = 0$, and also $E(v_0) = 0$; (b) for every $i = 1, ..., N$, $E(b_{i,0}^2) = \sigma_i^2$ and $E(e_{i,0}^2) = \tau_i^2$, and $E(v_0^2) = 1$; furthermore $N^{-1} \sum_{i=1}^{N} (\sigma_i^2 + \tau_i^2) = O(1)$; *(iii)* (a) $\{e_{i,t}\}_{t=-\infty}^{+\infty}$, $\{b_{i,t}\}_{t=-\infty}^{+\infty}$ and $\{v_t\}_{t=-\infty}^{+\infty}$ are three mutually independent groups for $i = 1, ..., N$; (b) $y_{i,0}$ is independent of $\{e_{i,t}, b_{i,t}, v_t\}$ for $i = 1, ..., N$; *(iv)* the γ_i's are independent across i and such that (a) $\gamma_i(N,T) \xrightarrow{\mathcal{D}} \bar{\gamma}_i$ for all $i = 1, ..., N$ as $N, T \to \infty$; (b) $N^{-1} \sum_{i=1}^{N} |\gamma_i| = O_P(1)$ as $\min(N,T) \to \infty$; (c) γ_i is independent of $\{y_{i,0}, e_{i,t}, b_{i,t}, v_t\}$ for $i = 1, ..., N$.

Assumption 1 contains a set of regularity conditions, as high level as possible. Parts *(i)* and *(iii)* stipulate that all the variables are serially independent (part *(i)*), and cross-sectionally independent (part *(iii)*). Cross sectional dependence among the $y_{i,t}$s is allowed for through the presence of a factor structure in the error - the term $\gamma_i v_t$ in equation (2.1); note that we do not require estimation of either γ_i or v_t, which are both treated as nuisance parameters, or of the number of common factors if v_t is multidimensional. The theory developed here can be extended to accommodate for serial dependence, with minor modifications to the main arguments of the proofs.

Part *(ii)* of Assumption 1 states that our theory requires the (minimal) assumption on the existence of second moments. A sufficient condition for parts *(ii)* and *(iii)* to hold is that $Ee_{i,0}^2 < c_1$, $Eb_{i,0}^2 < c_2$ and $E|\gamma_i| < c_3$ for all $i = 1, ..., N$, for some constants c_1, c_2 and c_3. By part (b) of the assumption, the initial values $y_{i,0}$ can be constants or random variables, as long as they are independent of the future error terms; no moment restrictions are needed on the initial conditions.

Finally, part *(iv)* takes into account the possibility that the correlation between units may decay; the limit in part (a), $\bar{\gamma}_i$, can be a constant (even 0 or ∞); however, part (b), intuitively, stipulates that we can only have very few large loadings.

Equations (1.1) and (2.1) can be solved explicitly, resulting in

$$(2.4) \qquad y_{i,t} = y_{i,0} \prod_{s=1}^{t}(\varphi + b_{i,s}) + \sum_{s=1}^{t} u_{i,s} \prod_{z=s}^{t-1}(\varphi + b_{i,z+1})$$

$$= y_{i,0} \prod_{s=1}^{t}(\varphi + b_{i,s}) + \sum_{s=1}^{t} e_{i,s} \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}) + \gamma_i \sum_{s=1}^{t} v_s \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}).$$

According to (2.4), the $y_{i,t}$s can be decomposed into three parts. The first term shows the effect of the initial value $y_{i,0}$; the second term is independent of all the other units, while the last one contains all the dependence of the ith unit to the other units.

Consider equation (1.1). According to the value taken by $b_{i,t}$, each of the $y_{i,t}$ can be: stationary (which, heuristically, corresponds to the AR root $\varphi + b_{i,t}$ being "smaller than one"); explosive (which, heuristically, corresponds to the case of $\varphi + b_{i,t}$ being "larger than one"); or on the boundary (which, heuristically, corresponds to the AR root $\varphi + b_{i,t}$ being "equal to one").

We now discuss the three regimes in detail.

Stationary units

Formally, we say that the ith unit is *stationary* if

$$(2.5) \qquad\qquad\qquad\qquad E \log |\varphi + b_{i,0}| < 0,$$

and we henceforth denote $C(1)$ to be a set containing the indices of the stationary units, i.e. the set of the i's for which (2.5) holds. Note that if condition (2.5) holds, then $y_{i,t}$ converges to a stationary solution as $t \to \infty$. The stationary solution (henceforth denoted as $\bar{y}_{i,t}$) is given by

$$(2.6) \qquad\qquad \bar{y}_{i,t} = \sum_{s=-\infty}^{t} e_{i,s} \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}) + \bar{\gamma}_i \sum_{s=-\infty}^{t} v_s \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}).$$

Aue *et al.* (2006) showed that the sums defining $\bar{y}_{i,t}$ are finite with probability one, and that there exists a $\kappa_i \in (0, 1]$ such that

$$(2.7) \qquad \delta_i \equiv E|\varphi + b_{i,0}|^{\kappa_i} < 1.$$

In order to study the contribution of the stationary units to the asymptotics of $\hat{\varphi}$ consider the following assumption:

Assumption 2. It holds that, as $N \to \infty$: *(i)*

$$\frac{1}{N} \sum_{i \in C(1)} \left[\frac{E|e_{i,0}|^{\kappa_i} + E|\bar{\gamma}_i|E|v_0|^{\kappa_i}}{1 - \delta_i} \right]^{1/(1+\kappa_i)} \frac{1}{\delta_i^{1/(1+\kappa_i)}} = O(1);$$

(ii) $N^{-1} \sum_{i \in C(1)} \left[(1 - \delta_i)^{-1} E|\gamma_i - \bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i} \right]^{1/(1+\kappa_i)} = o(1);$ *(iii)*

$$\frac{1}{N} \sum_{i \in C(1)} E \left\{ |\bar{\gamma}_i| \left[\frac{E|e_{i,0}|^{\kappa_i} + |\bar{\gamma}_i|E|v_0|^{\kappa_i}}{1 - \delta_i} \right]^{1/(1+\kappa_i)} \frac{1}{\delta_i^{1/(1+\kappa_i)}} \right\} = O(1);$$

(iv) $N^{-1} \sum_{i \in C(1)} [\kappa_i(1 - \delta_i)]^{-1} = O(1).$

Assumption 2 is rather technical, and it poses some regularity conditions on the stationary units. A sufficient condition for parts *(i)* and *(iii)* is that $E|v_0| < c_1$, $E|e_{i,0}| < c_2$ and $E\bar{\gamma}_i^2 < c_3$ and $0 < c_4 \leq \delta_i \leq c_5 < 1$ for all $i \in C(1)$, for some constants $c_1,..., c_5$. However, by Assumption 2, some of these moments could be tending to infinity, but in that case the number of stationary units should be small; similarly, if γ_i is nonrandom, the assumption implies that $\bar{\gamma}_i$ is finite for all $i \in C(1)$. By a similar logic, a sufficient condition for part *(ii)* is that γ_i converges to $\bar{\gamma}_i$ in L_1-norm, viz. $E|\gamma_i - \bar{\gamma}_i| = o(1)$. Finally, part *(iv)* stipulates that κ_i and $(1 - \delta_i)$ cannot be too small for too many units.

Explosive units

We say that the ith unit is *explosive* if

$$(2.8) \qquad E \log |\varphi + b_{i,0}| > 0,$$

and we denote $C(2)$ to be the set containing the indices of the explosive units, i.e. the set of the i's for which (2.8) holds. When (2.8) is satisfied, Berkes *et al.* (2009) prove that $|y_{i,t}| \to \infty$ at an exponential rate in probability as $t \to \infty$ (see also Lemma 7.4).

Henceforth, let

$$(2.9) \qquad \bar{\sigma}_i^2 \equiv \mathrm{var}(\log|\varphi + b_{i,0}|).$$

We need the following assumption when studying the asymptotics of explosive units.

Assumption 3. It holds that

$$\frac{1}{N} \sum_{i \in C(2)} \frac{\bar{\sigma}_i^3 + 1}{(E \log|\varphi + b_{i,0}|)^3} = O(1).$$

A set of sufficient conditions for Assumption 3 is that $\bar{\sigma}_i < c_1$ and $E \log|\varphi + b_{i,0}| > c_2$ for all $i \in C(2)$, for some constants c_1 and $c_2 > 0$. This entails that the values of the $\varphi + b_{i,0}$ are not too spread out (this also follows from Assumption 4*(ii)* below); further, the part that requires $E \log|\varphi + b_{i,0}| > 0$ is a way of ruling out too many "local-to-explosive" cases. Further assumptions that are needed for the case of explosive units will be spelt out when discussing the boundary case.

Boundary units

Finally, when it holds that

$$(2.10) \qquad E \log|\varphi + b_{i,0}| = 0,$$

we say that unit i is on the boundary between the stationary and the explosive behaviour. Indeed, if $\sigma_i = \mathrm{var}(b_{i,0}) = 0$, under (2.10) the ith unit would boil down to being a standard AR process with a unit root. We henceforth denote $C(3)$ to be the set containing the indices of the units on the boundary, i.e. for which (2.10) is satisfied.

When (2.10) holds, Berkes *et al.* (2009) show that $|y_{i,t}| \to \infty$ in probability as $t \to \infty$. However, in this case the rate of convergence to ∞ is slower than exponential (see Lemma 7.4).

Under (2.8) as well as (2.10), the $y_{i,t}$s are therefore unbounded. In both cases, we consider the following set of assumptions, again needed to study the impact of these units on the asymptotics of $\hat{\varphi}$.

Assumption 4. It holds that: *(i)* $u_{i,0}$ has a bounded density if $i \in C(2) \cup C(3)$; the upper bound is henceforth denoted as M_i; *(ii)* there exists a $\nu_i \in [\underline{v}, \overline{v}]$ with $\underline{v} > 0$ and $\underline{v} > 2$ such that $E|\log|\varphi +$

$b_{i,0}|^{\nu_i} < \infty$ for all $i \in C(2) \cup C(3)$; *(iii)* letting $m_i \equiv E|\log|\varphi + b_{i,0}| - E\log|\varphi + b_{i,0}||^{\nu_i}$, we have $N^{-1}\sum_{i\in C(2)\cup C(3)} E|\bar{\gamma}_i|M_i\left(1 + \bar{\sigma}_i^{-1} + m_i\right) = O(1)$ (recall that $\bar{\sigma}_i$ is defined in (2.9)).

3. ASYMPTOTICS

After spelling out the full set of assumptions needed for the consistency of $\hat{\varphi}$, in this section we report the asymptotic properties of $\hat{\varphi}$. Section 3.1 contains results on the consistency of the estimator; the limiting distribution is in Section 3.2.

3.1. **Consistency.** We start by showing that $\hat{\varphi}_{N,T}$ is a consistent estimator of φ. Henceforth, $\#A$ denotes the cardinality of a set A.

Theorem 3.1. *Under $Eb_{i,t}^2 > 0$, if Assumptions 1-4 hold and*

$$(3.1) \qquad \lim_{N\to\infty} \frac{1}{N}\left\{\sum_{i\in C(1)} E\frac{\bar{y}_{i,0}^2}{1+\bar{y}_{i,0}^2} + \#C(2) + \#C(3)\right\} = a_0 > 0,$$

as $\min(N,T) \to \infty$, we have $\hat{\varphi}_{N,T} \xrightarrow{P} \varphi$. The same result holds when $b_{i,t} = 0$, under Assumptions 1-3, 4(i)-(ii) and (3.1).

Remarks

Theorem 3.1 states that $\hat{\varphi}_{N,T}$ is always consistent for φ, irrespective of the value taken by φ. Thus, $\hat{\varphi}_{N,T}$ is consistent if all units are stationary, if some of them are nonstationary, and even if some or all of them are explosive. Similarly, the results hold irrespective of whether $Eb_{i,t}^2 > 0$ or whether $b_{i,t} = 0$. In this respect, $\hat{\varphi}_{N,T}$ does not have the typical "boundary problems" which are encountered in the unit root literature (see for example Phillips, 1987). Finally, (3.1) requires that a_0 is nonzero, which is a non-degeneracy condition to rule out that the denominator of $\hat{\varphi}_{N,T}$ converges to zero. Equation (3.1) always holds, unless there are "too many" units with $\bar{y}_{i,0} = 0$ and the number of units with $|\varphi| \geq 1$ is very small. Note that, when $b_{i,t} = 0$, the condition is automatically satisfied for $|\varphi| \geq 1$, with $a_0 = 1$.

3.2. **Limiting distribution.** In this section, we study the asymptotic distribution of the suitably normed $\hat{\varphi}_{N,T} - \varphi$. The main results of this section are: the limiting distribution, the rates of convergence, and the computation of the norming sequences. We show that these depend (when $Eb_{i,t}^2 > 0$) on

$$(3.2) \qquad r_N \equiv E\left[\sum_{i\in C(1)} \gamma_i \frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\right]^2 = \sum_{i\in C(1)}\sum_{j\in C(1)} E\left\{\gamma_i\gamma_j \frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\frac{\bar{y}_{j,0}}{1+\bar{y}_{j,0}^2}\right\}.$$

The quantity r_N is determined by the amount of cross sectional dependence across the *stationary* units only. When $Eb_{i,t}^2 > 0$, cross sectional dependence has an impact on the asymptotics of $\hat\varphi_{N,T}$ only through the stationary units. Note that if no factor structure is present in the error term $u_{i,t}$, i.e. if $\gamma_i = 0$ for every $i \in C(1)$, then $r_N = 0$. On the other hand, an upper bound for r_N is $\sum_{i \in C(1)} \sum_{j \in C(1)} E|\gamma_i \gamma_j| = O(N^2)$, which represents strong cross sectional dependence.

In order to derive the asymptotic distribution of $\hat\varphi_{N,T}$, we need the following assumption, which strengthens some of the moment conditions in Assumption 1.

Assumption 5. Let $\epsilon > 0$. It holds that: *(i)* $E|v_0|^{2+\epsilon} < \infty$ and $N^{-1} \sum_{i=1}^N E|\gamma_i|^{2+\epsilon} = O(1)$; *(ii)* $N^{-1} \sum_{i=1}^N E|b_{i,0}|^{2+\epsilon} = O(1)$ and $N^{-1} \sum_{i=1}^N E|e_{i,0}|^{2+\epsilon} = O(1)$.

The main results of this section are the following two theorems. The first one deals with the case of genuinely random autoregressive root (i.e. $Eb_{i,t}^2 > 0$), whereas the second one considers the properties of the estimator when $b_{i,t} = 0$.

Theorem 3.2. *Let Assumptions 1-4 with $Eb_{i,t}^2 > 0$, and (3.1) hold.*

(i) If Assumption 5(i) is satisfied and $\lim_{N\to\infty} \frac{r_N}{N^2} = a_1 > 0$, then, as $\min(N,T) \to \infty$, we have

$$(3.3) \qquad \sqrt{T}\left(\hat\varphi_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{a_1}{a_0^2}\right).$$

(ii) If Assumption 5(i) is satisfied and $\lim_{N\to\infty} \frac{r_N}{N} = \infty$, then, as $\min(N,T) \to \infty$, we have

$$(3.4) \qquad \sqrt{\frac{N^2 T}{r_N}}\left(\hat\varphi_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{a_0^2}\right).$$

(iii) If Assumptions 5(i)-(ii) are satisfied and $\limsup_{N\to\infty} \frac{r_N}{N} < \infty$, then, as $\min(N,T) \to \infty$, we have

$$(3.5) \qquad \sqrt{NT}\left(\hat\varphi_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{a_2}{a_0^2}\right),$$

where

$$\lim_{N\to\infty} \frac{1}{N}\left\{ \sum_{i\in C(1)} \sigma_i^2 E\left(\frac{\bar y_{i,0}^2}{1+\bar y_{i,0}^2}\right)^2 + \sum_{i\in C(2)\cup C(3)} \sigma_i^2 + \sum_{i\in C(1)} \tau_i^2 E\left(\frac{\bar y_{i,0}}{1+\bar y_{i,0}^2}\right)^2 + r_N \right\} = a_2.$$

Remarks

Theorem 3.2 reports the rates of convergence and the limiting distribution of $\hat{\varphi}_{N,T}$ under various degrees of cross sectional dependence, depending on the value taken by r_N. The estimator always has a normal distribution, but the rate of convergence is affected by cross sectional dependence.

Part *(i)* of the Theorem considers the case of strong cross sectional dependence. An example in which part *(i)* holds is if $\gamma_i = 1$ for all units i. In this case, we show that $\hat{\varphi}_{N,T}$ is $\sqrt{T}$-convergent. Note that the limiting distribution is completely determined by the stationary units: units that are explosive or on the boundary do not have an impact on (3.3).

Turning to part *(iii)* of the Theorem, this holds when nearly all the loadings of the stationary units go to 0. This also includes the important case of $\gamma_i = 0$ for $1 \leq i \leq N$, i.e. cross sectional independence. Indeed, if the $\bar{\gamma}_i$s are nonrandom, with $\bar{\gamma}_i = \bar{\gamma}_j = 0$, then $\bar{y}_{i,t}$ and $\bar{y}_{j,t}$ are independent (and independent of γ_i) and therefore

$$\frac{r_N}{N} = \frac{1}{N}\left(\sum_{i \in C(1)} \gamma_i E \frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\right)^2;$$

if, further, the distribution of $e_{i,0}$ is symmetric around 0, then $E\frac{\bar{y}_{i,t}}{1+\bar{y}_{i,t}^2} = 0$, so that $\bar{\gamma}_i = 0$ suffices for part *(iii)* to hold, regardless the rate of convergence of γ_i to 0. However, if the distribution of $e_{i,0}$ is asymmetric, γ_i must converge to 0 at a faster rate than $N^{-1/2}$. This condition is known in the panel literature as having "weak factors" - see Onatski (2012). A special case of part *(iii)* of the Theorem is the case where $\lim_{N\to\infty} r_N/N = 0$. In such case, equation (3.5) specialises into $\sqrt{NT}\left(\hat{\varphi}_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N\left(0, a_3/a_0^2\right)$, where

$$\lim_{N\to\infty} \frac{1}{N}\left\{\sum_{i \in C(1)} \sigma_i^2 E\left(\frac{\bar{y}_{i,0}^2}{1+\bar{y}_{i,0}^2}\right)^2 + \sum_{i \in C(2)\cup C(3)} \sigma_i^2 + \sum_{i \in C(1)} \tau_i^2 E\left(\frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\right)^2\right\} = a_3.$$

Finally, part *(ii)* of the Theorem is the case in between strong cross sectional dependence (part *(i)*), and very weak cross sectional dependence (part *(iii)*). This case can be illustrated by considering, as an example, the case where the loadings γ_i are non-random with $\gamma_i = N^{-\delta}$ for some $\delta \in \left(0, \frac{1}{2}\right)$.

We now consider the asymptotic properties of $\hat{\varphi}_{N,T}$ when $b_{i,t} = 0$; we show that the rates of convergence depend on

$$(3.6) \qquad\qquad r'_N \equiv \sum_{i=1}^{N}\sum_{j=1}^{N}\left(E\left|\gamma_i\gamma_j\right|^2\right)^{1/2}.$$

Consider the notation $C_N = \max\left\{\frac{1}{\sqrt{N}}, \sqrt{\frac{r'_N}{N^2}}\right\}$.

Theorem 3.3. *Under $b_{i,t} = 0$, let Assumptions 1-3, 4(i) and (3.1) hold. As $\min(N,T) \to \infty$*

(i) Under the Assumptions of Theorem 3.2 with $b_{i,t} = 0$ and

$$\lim_{N \to \infty} \frac{1}{N} \left\{ \sum_{i=1}^{N} \tau_i^2 E \left(\frac{\bar{y}_{i,0}}{1 + \bar{y}_{i,0}^2} \right)^2 + r_N \right\} = a_2' > 0,$$

the results of Theorem 3.2 hold for $|\varphi| < 1$.

(ii) If $\varphi = 1$, $\hat{\varphi}_{N,T} - 1 = O_P \left(T^{-2/3} C_N \right)$; further, if Assumption 5 holds,

$$(3.7) \qquad NT \left[\sum_{i=1}^{N} \tau_i^2 \left(\sum_{t=2}^{T} \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \right)^2 + \left(\sum_{i=1}^{N} \gamma_i \sum_{t=2}^{T} \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \right)^2 \right]^{-1/2} \left(\hat{\varphi}_{N,T} - 1 \right) \xrightarrow{\mathcal{D}} N(0,1).$$

(iii) If $|\varphi| > 1$, then $\hat{\varphi}_{N,T} - \varphi = O_P \left(T^{-1} C_N \right)$; further, if Assumption 5 holds and $r_N'/N \to 0$

$$(3.8) \qquad NT \left[\sum_{i=1}^{N} \tau_i^2 \left(\sum_{t=2}^{T} \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \right)^2 \right]^{-1/2} \left(\hat{\varphi}_{N,T} - \varphi \right) \xrightarrow{\mathcal{D}} N(0,1).$$

Remarks

Theorem 3.3 considers the case in which φ is constant over time and homogeneous across units: the WLS estimator $\hat{\varphi}_{N,T}$ is always consistent.

Part *(i)* is, in essence, the same as in Theorem 3.2. Part *(ii)* considers the panel unit root case: in this case the WLS estimator of φ has a rate of convergence that is faster than $T^{-1/2}$. The rate provided by the theorem is not the sharpest possible, and it is indeed only an upper bound. Equation (3.7) stipulates that, under all circumstances, the WLS estimator of φ converges to a normal distribution, so that even in this case there is no "unit root problem". Technically, this is due to the fact that, although $|y_{i,t}| \to \infty$ in probability, this is not at an exponential rate; thus, the variance of the term that leads the asymptotics still diverges as $T \to \infty$.

As far as part *(iii)* is concerned, the rate of convergence is T; the impact of the cross sectional dimension on the rate of convergence is the same as in the case of a genuinely random coefficient model. However, asymptotic normality holds when cross sectional dependence is weak (i.e. under $r_N'/N \to 0$). Intuitively, this is a consequence of being in a panel data context: the cross sectional averaging affords a CLT to hold, even in those cases in which, in a single time series case, it would be impossible to show convergence to normality (see the comments in Hill and Peng, 2014). Conversely, when there is strong cross sectional dependence, $\hat{\varphi}_{N,T} - \varphi$ does not, in general, converge to a normal. Heuristically, this is due to the fact that the variance of the leading term stays bounded as $T \to \infty$, which is a degenerate case - see e.g. Davidson (1993). This is a limitation of WLS-based inference, although it may be argued that it corresponds to a quite restrictive case: φ is larger than 1 (thereby having a genuinely explosive model),

and it is the same across all units. This also entails that an estimation technique based on WLS which removes the factor structure $\gamma_i v_t$ will yield asymptotic normality under all possible cases.

Finally, it is instructive to compare the rates of convergence provided in the theorem with those that one would have in a pure time series setup, when using a standard OLS estimator. In such a case, the rate of convergence is the same as in Theorem 3.3 for the case of a stationary series, viz. $|\varphi| < 1$. When $\varphi = 1$, it is well known that the OLS estimator of φ is T-consistent; as mentioned above, the rate provided in the theorem is not the sharpest one. The biggest discrepancy, however, is found in the case of an explosive root, viz. $|\varphi| > 1$: the T-consistency of the WLS estimator can be contrasted with the OLS estimator (see Wang and Yu, 2015), which converges at a rate $O_P(\varphi^T)$. Intuitively, this is due to the use of the weight $1/(1 + y_{i,t-1}^2)$, which is designed to hold down both numerator and denominator of $\hat{\varphi}_{N,T}$.

The main result in Theorems 3.2 and 3.3 is that, modulo the exception detailed in Theorem 3.3, the suitably normalised estimation error $\hat{\varphi}_{N,T} - \varphi$ converges to a normal. We now discuss the estimation of the asymptotic variance of $\hat{\varphi}_{N,T} - \varphi$. Our main result is that there exists one estimator which is always consistent, with no need to know the amount of cross sectional dependence, or whether $Eb_{i,t}^2 > 0$ or $b_{i,t} = 0$. Thus, $\hat{\varphi}_{N,T} - \varphi$ can always be normalised by such estimator, and the normalised quantity will always converge to a standard normal.

The estimation of the asymptotic variance is based on the weighted residuals

$$(3.9) \qquad z_{i,t} = \left(y_{i,t} - \hat{\varphi}_{N,T} y_{i,t-1}\right) \frac{y_{i,t-1}}{1 + y_{i,t-1}^2}.$$

Define $U_{N,T} \equiv \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} z_{i,t} z_{j,t}$; we propose the following "universal" estimator of the asymptotic variance of $\hat{\varphi}_{N,T} - \varphi$:

$$(3.10) \qquad V_{N,T} = \frac{U_{N,T}}{B_{N,T}^2}.$$

Theorem 3.4. *Let the Assumptions of Theorems 3.2 and 3.3 hold, and assume further that, for some $1 \leq c_2 < \infty$*

$$(3.11) \qquad \sum_{i \in C(1)} \sum_{j \in C(1)} E \left| \gamma_i \gamma_j \frac{\bar{y}_{i,0}}{1 + \bar{y}_{i,0}^2} \frac{\bar{y}_{j,0}}{1 + \bar{y}_{j,0}^2} \right| \leq c_2 r_N.$$

Then, as $\min(N,T) \to \infty$, we have $V_{N,T}^{-1/2} \left(\hat{\varphi}_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N(0,1)$.

Remarks

Theorem 3.4 illustrates once again that there is no boundary problem in case of a panel RCA, apart from the case discussed in Theorem 3.3. Interestingly, the theorem stipulates that the same random normalization can be used regardless the structure of the units: the random norming by $U_{N,T}$ is the same regardless of how strong is the cross correlation between the units, and of the proportion of stationary versus nonstationary (boundary or explosive) units. Thus, in contrast to autoregressive processes, if we wish to test for $H_0 : \varphi > \varphi_0$, the asymptotic normal limit can be used regardless the value of φ_0, even if $\varphi_0 \geq 1$. As mentioned in Theorem 3.3, the only case in which standard normal inference is not valid is when $b_{i,t} = 0$ and $\varphi > 1$ in the presence of a pervasive factor structure.

4. EXTENSIONS: INTRODUCING DETERMINISTICS AND COVARIATES

In this section, we show that our results are essentially unchanged when considering extensions of the basic model such as the presence of individual effects and covariates. Specifically, we consider the case where one observes

$$(4.1) \qquad\qquad\qquad y_{i,t}^* = \alpha_i + y_{i,t},$$

with

$$(4.2) \qquad\qquad y_{i,t} = (\varphi + b_{i,t})y_{i,t-1} + \beta_i' x_{i,t} + u_{i,t}, \quad 1 \leq i \leq N \quad \text{and} \quad 1 \leq t \leq T;$$

in (4.2), $u_{i,t}$ is defined in (2.1). No assumptions are needed on the α_is, since these are removed prior to estimating φ; thus, the individual effects can be fixed, random, and correlated or not with the other covariates. Similarly, we allow for some flexibility in the unit specific regressors $x_{i,t}$ - these can e.g. be correlated with $y_{i,t-1}$, and as far as serial dependence is concerned, we only assume that they are stationary. Henceforth, the number of regressors $x_{i,t}$ is referred to as h.

Model (4.1)-(4.2) nests several popular specifications. The leading examples are: (a) the "classical" fixed or random effects dynamic panel regression, which corresponds to the case $b_{i,t} = 0$ and $\gamma_i = 0$; (b) the panel model with time effects (although an important restriction in our case is that correlation between the time effect and $y_{i,t-1}$ is ruled out) which corresponds to having $b_{i,t} = 0$ and $\gamma_i = \gamma$; (c) in the most general case, our model is similar to the ones considered in Bai (2009) and Pesaran (2006), with the addition of considering the presence of the weakly endogenous regressor $y_{i,t-1}$ (see also Song, 2013; Chudik and Pesaran, 2015; Moon and Weidner, 2015), and of randomness in the slope of its coefficient; in our case, though, we need to rule out the correlation between $x_{i,t}$ and $\gamma_i v_t$.

As in the previous sections, the focus of our analysis is the estimation of φ only. Firstly, in order to remove individual effects, let

$$(4.3) \qquad \tilde{y}_{i,t} = y_{i,t}^* - y_{i,0}^* = y_{i,t} - y_{i,0},$$

so that the α_is are treated as nuisance parameters.[2] As far as covariates are concerned, the likelihood maximisation problem can be formalised as

$$(4.4) \qquad \min_{\varphi, \beta_1, ..., \beta_N} S\left(\varphi, \beta_1, ..., \beta_N\right),$$

with

$$S\left(\varphi, \beta_1, ..., \beta_N\right) = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\left(\tilde{y}_{i,t} - \varphi \tilde{y}_{i,t-1} - \beta_i' x_{i,t}\right)^2}{1 + \tilde{y}_{i,t-1}^2}.$$

From (4.4), it follows that the infeasible estimator of β_i is, for $1 \le i \le N$

$$(4.5) \qquad \tilde{\beta}_i^{INF} = \left[\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2}\right]^{-1} \left[\sum_{t=2}^{T} \frac{x_{i,t} \left(\tilde{y}_{i,t} - \varphi \tilde{y}_{i,t-1}\right)}{1 + \tilde{y}_{i,t-1}^2}\right].$$

Hence

$$(4.6) \qquad \tilde{\varphi} \equiv \tilde{\varphi}_{N,T} = \frac{\tilde{A}_{N,T}}{\tilde{B}_{N,T}},$$

where

$$(4.7) \qquad \tilde{A}_{N,T} = A_{N,T} - \sum_{i=1}^{N} \left[\sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t-1}}{1 + \tilde{y}_{i,t-1}^2}\right]' \left[\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2}\right]^{-1} \left[\sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t}}{1 + \tilde{y}_{i,t-1}^2}\right];$$

$$(4.8) \qquad \tilde{B}_{N,T} = B_{N,T} - \sum_{i=1}^{N} \left[\sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t-1}}{1 + \tilde{y}_{i,t-1}^2}\right]' \left[\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2}\right]^{-1} \left[\sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t-1}}{1 + \tilde{y}_{i,t-1}^2}\right].$$

The definitions of $\tilde{A}_{N,T}$ and $\tilde{B}_{N,T}$ can be contrasted with those of $A_{N,T}$ and $B_{N,T}$ provided in (2.3). Note that, based on this approach, it is possible to have a feasible estimator of the β_is, defined as

$$(4.9) \qquad \tilde{\beta}_i = \left[\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2}\right]^{-1} \left[\sum_{t=2}^{T} \frac{x_{i,t} \left(\tilde{y}_{i,t} - \tilde{\varphi}_{N,T} \tilde{y}_{i,t-1}\right)}{1 + \tilde{y}_{i,t-1}^2}\right].$$

---

[2]The scheme proposed in (4.3) is not the only way of dealing with unit specific effects. A more natural approach would be based on defining the vectors $\tilde{x}_{i,t} = \left[1, x_{i,t}'\right]'$ and $\tilde{\beta}_i = [\alpha_i, \beta_i']'$, and estimate the slopes $\tilde{\beta}_i$ in

$$y_{i,t}^* = (\varphi + b_{i,t}) y_{i,t-1} + \tilde{\beta}_i' \tilde{x}_{i,t} + u_{i,t},$$

based on (4.4). In such a case, one would be able to estimate the average of the α_is (see Section 4.1). On the other hand, the individual effects α_i would have to satisfy the same assumptions as β_i; for example, they would have to be independent of $y_{i,t-1}$.

We discuss, as a by-product, estimation and inference on the β_is, and on their average, in Section 4.1. Consider now the following assumption on the covariates $x_{i,t}$, which complements the existing assumptions.

Assumption 6. It holds that: *(i)* for $i \in C(1)$, $x_{i,t} = f\left[\xi_{i,t}, \xi_{i,t-1}...\right]$ where (a) $E|x_{i,0}|^4 < \infty$; (b) $f : \mathbb{R}^{h' \times \infty} \to \mathbb{R}^h$ is a measurable function; (c) $\{\xi_{i,t}\}$, $1 \leq t \leq T$ is a sequence of *i.i.d.* (across t) random variables with values in $\mathbb{R}^{h'}$; (d) letting $x_{i,t}^{(m)} = f\left[\xi_{i,t}, ..., \xi_{i,t-m}, \xi_{i,t,t-m-1}^{(m)}, ...\right]$ with $\xi_{i,t,j}^{(m)}$ a sequence of *i.i.d.* (across i and t) copies of $\xi_{i,0}$, it holds that $E\left\|x_{i,t} - x_{i,t}^{(m)}\right\| = O\left(r_m\right)$, where $r_m = e^{-m}$; *(ii)* (a) $\{\beta_i\}$, $\{x_{i,t}\}$ and $\{e_{i,t}, \gamma_i v_t, b_{i,t}\}$ are three mutually independent groups for $t = 1, ..., T$ and $i = 1, ..., N$; (b) $y_{i,0}$ is independent of $\left\{e_{i,t}, b_{i,t}, \gamma_i v_t, \beta_i' x_{i,t}\right\}$ for $i = 1, ..., N$; *(iii)* (a)

$$\frac{1}{N} \sum_{i \in C(1)} \left[\frac{E|\beta_i' x_{i,0}|^{\kappa_i} + E|e_{i,0}|^{\kappa_i} + E|\bar{\gamma}_i| E|v_0|^{\kappa_i}}{1 - \delta_i}\right]^{1/2(1+\kappa_i)} \frac{1}{\delta_i^{1/2(1+\kappa_i)}} = O(1);$$

(b) $N^{-1} \sum_{i \in C(1)} \left[(1 - \delta_i)^{-1} E|\gamma_i - \bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i}\right]^{1/2(1+\kappa_i)} = o(1);$ (c)

$$\frac{1}{N} \sum_{i \in C(1)} E\left\{|\bar{\gamma}_i| \left[\frac{E|\beta_i' x_{i,0}|^{\kappa_i} + E|e_{i,0}|^{\kappa_i} + |\bar{\gamma}_i| E|v_0|^{\kappa_i}}{1 - \delta_i}\right]^{1/2(1+\kappa_i)} \frac{1}{\delta_i^{1/2(1+\kappa_i)}}\right\} = O(1);$$

(d) $E\|\beta_i\|^{\kappa_i} < \infty$ for $i \in C(1)$; *(iv)* $\beta_i' x_{i,0} + u_{i,0}$ has a bounded density if $i \in C(2) \cup C(3)$; *(v)* (a) $\max_{i,t} E\|x_{i,t}\|^{8+\epsilon} < \infty$ for every $i = 1, ..., N$ and (b) $E|e_{i,0}|^{4+\epsilon} < \infty$ for every $i = 1, ..., N$.

Part *(i)* of the assumption states that the $x_{i,t}$s are Bernoulli shifts, and therefore they are stationary processes that are, possibly, serially dependent; and that there exists an m-dependent approximation $x_{i,t}^{(m)}$ which, according to part (d), is close enough to $x_{i,t}$. This way of modelling dependence has been employed in several contexts (see e.g. Section 21 in Billingsley, 1968); as illustrated in Aue *et al.* (2009), it has the advantages of being mathematically tractable and of nesting several popular models, such as multivariate ARMA and a wide variety of GARCH models. Note that we only need to make an explicit assumption on the time dependence of the $x_{i,t}$s for the stationary units. By part *(ii)*, the $x_{i,t}$s are not required to be cross-sectionally independent, even though they are required to be independent of the error term $e_{i,t}$ and of the common factor structure v_t. Parts *(iii)* and *(iv)* extend Assumptions 2 and 4 (respectively), so that technical results such as Lemmas 7.2-7.4 hold in presence of covariates also; note that we do not need to assume that $x_{i,t}$ and $y_{i,t-1}$ are independent.

Define now

$$D_{i,t} \equiv \tilde{y}_{i,t-1} - x_{i,t}' \left[\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2}\right]^{-1} \sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t-1}}{1 + \tilde{y}_{i,t-1}^2},$$

and consider the notation

$$\tilde{r}_N = E\left[\sum_{i\in C(1)} \gamma_i \frac{D_{i,0}}{1+\bar{y}_{i,0}^2}\right]^2.$$

The following results characterise the consistency and the limiting distribution of $\tilde{\varphi}_{N,T}$, and they are the counterpart to Theorems 3.1 and 3.2 respectively.

Theorem 4.1. *Under $Eb_{i,t}^2 > 0$, let Assumptions 1-5 and 6(i)-(ii)-(iii)-(iv)-(v)(a) hold. If*

$$\lim_{N\to\infty} \frac{1}{N}\left\{\sum_{i\in C(1)} E\frac{\bar{y}_{i,0}D_{i,0}}{1+\bar{y}_{i,0}^2} + \#C(2) + \#C(3)\right\} = \tilde{a}_0 > 0,$$

as $\min(N,T)\to\infty$ it holds that $\tilde{\varphi}_{N,T} \xrightarrow{P} \varphi$, with the same rates of convergence as in Theorem 3.2. Further, it holds that

(4.10)
$$\sqrt{\frac{\tilde{a}_0^2}{\tilde{a}_1}}\left(\hat{\varphi}_{N,T} - \varphi\right) \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$\lim_{N\to\infty} \frac{1}{N}\left\{\sum_{i\in C(1)} \sigma_i^2 E\left(\frac{\bar{y}_{i,0}D_{i,0}}{1+\bar{y}_{i,0}^2}\right)^2 + \sum_{i\in C(2)\cup C(3)} \sigma_i^2 + \sum_{i\in C(1)}^{N} \tau_i^2\left(\frac{D_{i,0}}{1+\tilde{y}_{i,0}^2}\right)^2 + \tilde{r}_N\right\} = \tilde{a}_1.$$

Let $C_N' = \max\left\{\frac{1}{\sqrt{N}}, \sqrt{\frac{r_N'}{N^2}}\right\}$.

Theorem 4.2. *Under $b_{i,t} = 0$, let Assumptions 1-6 hold. As $\min(N,T)\to\infty$*

(i) If, further

$$\lim_{N\to\infty} \frac{1}{N}\left\{\sum_{i=1}^{N} \tau_i^2 E\left(\frac{D_{i,0}}{1+\bar{y}_{i,0}^2}\right)^2 + \tilde{r}_N\right\} = \tilde{a}_0' > 0,$$

the results of Theorem 4.1 hold for $|\varphi| < 1$.

(ii) If $|\varphi| = 1$, $\tilde{\varphi}_{N,T} - 1 = O_P\left(T^{-21/40}C_N'\right)$ and

(4.11)
$$NT\left\{\sum_{i=1}^{N}\tau_i^2\sum_{t=2}^{T}\left(\frac{D_{i,t}}{1+\tilde{y}_{i,t-1}^2}\right)^2 + \sum_{t=2}^{T}\left(\sum_{i=1}^{N}\gamma_i\frac{D_{i,t}}{1+\tilde{y}_{i,t-1}^2}\right)^2\right\}^{-1/2}\left(\hat{\varphi}_{N,T}-1\right) \xrightarrow{\mathcal{D}} N(0,1).$$

(iii) If $|\varphi| > 1$, under the conditions of Theorem 3.3 and Assumption 6, it holds that $\tilde{\varphi}_{N,T} - \varphi = O_P\left(T^{-1}C_N'\right)$; further, if $r_N'/N \to 0$ it holds that

(4.12)
$$NT\left\{\sum_{i=1}^{N}\tau_i^2\sum_{t=2}^{T}\left(\frac{D_{i,t}}{1+\tilde{y}_{i,t-1}^2}\right)^2\right\}^{-1/2}\left(\tilde{\varphi}_{N,T}-\varphi\right) \xrightarrow{\mathcal{D}} N(0,1).$$

Theorems 4.1 and 4.2 state, essentially, that results in the presence of covariates are the same as for the baseline case of model (1.1): the only difference is in the asymptotic variances, which change to reflect the presence of the covariates $x_{i,t}$. Indeed, as one may expect, it can be shown that results would be exactly the same as in Section 3 if $x_{i,t}$ and $y_{i,t-1}$ were assumed to be independent.

Turning to the estimation of the asymptotic variance of $\tilde{\varphi}_{N,T}$, we use $\tilde{V}_{N,T} = \frac{\tilde{U}_{N,T}}{\tilde{B}_{N,T}^2}$, where $\tilde{U}_{N,T} \equiv \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{z}_{i,t} \tilde{z}_{j,t}$ and

$$(4.13) \qquad \tilde{z}_{i,t} = (\tilde{y}_{i,t} - \tilde{\varphi}_{N,T} \tilde{y}_{i,t-1} - \tilde{\beta}_i' x_{i,t}) \frac{D_{i,t}}{1 + \tilde{y}_{i,t-1}^2}.$$

Theorem 4.3. *Under the conditions of Theorem 3.4, with (3.11) replaced by*

$$\sum_{i \in C(1)} \sum_{j \in C(1)} E \left| \gamma_i \gamma_j \frac{D_{i,0}}{1 + \bar{y}_{i,0}^2} \frac{D_{j,0}}{1 + \bar{y}_{j,0}^2} \right| \le c_2 r_N,$$

with $1 \le c_2 < \infty$, and under Assumption 6, as $\min(N, T) \to \infty$ it holds that $\tilde{V}_{N,T}^{-1/2} (\tilde{\varphi}_{N,T} - \varphi) \xrightarrow{\mathcal{D}} N(0, 1)$.

4.1. **Estimation and inference on β_i.** Based on (4.9), it is possible to study the estimation of the mean of the individual specific slopes β_i. To this end, we assume a random coefficient model for the β_is, similar e.g. to the one assumed in Pesaran (2006; see in particular Assumption 4)

$$(4.14) \qquad \beta_i = \bar{\beta} + \ddot{\beta}_i.$$

Building on (4.14), a possible way of estimating $\bar{\beta}$ is to use

$$(4.15) \qquad \widehat{\bar{\beta}} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2} \right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \frac{x_{i,t} \left(\tilde{y}_{i,t} - \tilde{\varphi}_{N,T} \tilde{y}_{i,t-1} \right)}{1 + \tilde{y}_{i,t-1}^2} \right],$$

whose covariance matrix can be estimated by

$$
\begin{aligned}
\hat{V}_\beta &= \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2} \right]^{-1} \times \left[\sum_{i=1}^{N} \left(\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2} \right) \left(\tilde{\beta}_i - \widehat{\bar{\beta}} \right) \left(\tilde{\beta}_i - \widehat{\bar{\beta}} \right)' \left(\sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2} \right) + \right. \\
&\quad \left. + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \frac{\left(\tilde{y}_{i,t} - \tilde{\varphi}_{N,T} \tilde{y}_{i,t-1} - \tilde{\beta}_i' x_{i,t} \right) G_{i,t}}{1 + \tilde{y}_{i,t-1}^2} \frac{\left(\tilde{y}_{j,t} - \tilde{\varphi}_{N,T} \tilde{y}_{j,t-1} - \tilde{\beta}_j' x_{j,t} \right) G_{j,t}'}{1 + \tilde{y}_{j,t-1}^2} \right] \\
&\quad \times \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \frac{x_{i,t} x_{i,t}'}{1 + \tilde{y}_{i,t-1}^2} \right]^{-1},
\end{aligned}
$$

having defined $G_{i,t} \equiv x_{i,t} - D_{i,t} \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{x_{i,t} \tilde{y}_{i,t-1}}{1 + \tilde{y}_{i,t-1}^2}$.

Consider the following assumption, which complements Assumption 6.

Assumption 7. It holds that: *(i)* $\bar{\beta}$ is nonrandom with $\left\| \bar{\beta} \right\| < \infty$; *(ii)* $\ddot{\beta}_i$ is *i.i.d.* across i with (a) $E\left(\ddot{\beta}_i\right) = 0$ and (b) $E\left\| \ddot{\beta}_i \right\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

It holds that:

Theorem 4.4. *Let Assumptions 1-7 hold. As* $\min(N, T) \to \infty$

(a) if

$$(4.16) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{i \in C(1)} E \frac{x_{i,0}^2}{1 + \tilde{y}_{i,0}^2} = a_\beta > 0,$$

it holds that $\widehat{\bar{\beta}} - \bar{\beta} = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\sqrt{\frac{r_N}{N^2 T}}\right)$;

(b) if

$$(4.17) \qquad \lim_{N,T} \frac{1}{r_N'} \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E \left\| \gamma_i \gamma_j \frac{G_{i,t}}{1 + \tilde{y}_{i,t-1}^2} \frac{G_{j,t}'}{1 + \tilde{y}_{j,t-1}^2} \right\| = +\infty,$$

it holds that $\widehat{\bar{\beta}} - \bar{\beta} = O_P\left(\frac{1}{\sqrt{N}}\right) + o_P\left(\sqrt{\frac{r_N}{N^2}}\right)$;

(c) if

$$(4.18) \qquad \lim_{N,T} \frac{1}{r_N'} \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E \left\| \gamma_i \gamma_j \frac{G_{i,t}}{1 + \tilde{y}_{i,t-1}^2} \frac{G_{j,t}'}{1 + \tilde{y}_{j,t-1}^2} \right\| = c' < \infty,$$

it holds that $\widehat{\bar{\beta}} - \bar{\beta} = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\sqrt{\frac{r_N}{N^2}}\right)$.

Further, if (4.16) or (4.17) hold, then $\hat{V}_\beta^{-1/2}\left(\widehat{\bar{\beta}} - \bar{\beta}\right) \xrightarrow{\mathcal{D}} N(0, I_h)$; *the same holds under (4.18) if, in addition,* $\frac{r_N'}{N} \to 0$.

Theorem 4.4 states that, in essence, $\widehat{\bar{\beta}}$ is always consistent, save for the case in which there are too many explosive units, and there is strong cross dependence, which corresponds to part *(iii)*. The rates of convergence differ according as (4.16), (4.17) or (4.18) hold. As in the previous theorems, the asymptotic normality of $\widehat{\bar{\beta}}$ holds for all cases considered; the only exception is when the number of boundary units is small, the number of stationary units is very small, and, in addition, there is strong cross sectional dependence. However, in principle an estimation technique that accounts for cross sectional dependence would restore the asymptotic normality and consistency for $\widehat{\bar{\beta}}$ even in this case.

5. Monte Carlo simulations

In this section, we present some evidence on the properties of the WLS estimator of φ from synthetic data. In particular, we consider the bias and the Mean Squared Error (MSE) of $\hat{\varphi}_{N,T}$, and we also analyse the empirical coverage of 95% confidence intervals, in order to evaluate the quality of the estimator of the asymptotic variance.

We base all experiments on (1.1) and (2.1), viz.

$$y_{it} = (\varphi + b_{it}) y_{it-1} + u_{it},$$

$$u_{it} = e_{it} + \gamma_i v_t.$$

As far as φ is concerned, we consider the following grid of values: $\varphi \in \{-1.5, -1, -0.5, 0, 0.5, 1, 1.5\}$.

In a first set of experiments, we consider the case of $Eb_{i,t}^2 > 0$. The individual coefficient random shocks $b_{i,t}$ have been generated as *i.i.d.* across i and t with $b_{i,t} \sim \sigma N(0,1)$. We ran various experiments with different values of σ, also allowing for heteroskedasticity, but results do not change much; thus, we report only the case corresponding to $\sigma = 1$. Similarly, as far as the idiosyncratic component of the error term $u_{i,t}$ is concerned, this is generated as *i.i.d.* across i and t, with $u_{i,t} \sim \tau N(0,1)$; no changes in the results were observed when considering a heteroskedastic design or several values of τ, and we therefore only report results for the case $\tau = 1$. As far as the common factors are concerned, we consider nonrandom, homogeneous loadings, i.e. $\gamma_i = \gamma$; introducing randomness and/or heterogeneity was found to have no impact on the results. In order to consider the impact of cross sectional dependence on our estimator, we consider three sets of experiments with $\gamma \in \{0, 1, 10\}$. The case $\gamma = 0$ is covered by equation (3.5), and it should correspond to $\hat{\varphi}$ being $\sqrt{NT}$-convergent, which is the fastest attainable rate. The common factor v_t is generated as *i.i.d.* with $v_t \sim N(0,1)$. Finally, in order to assess the impact of initial conditions on the results, we considered various possible initialisations for $y_{i,0}$; no changes were noticed across experiments, and in this chapter we report results corresponding to the case of $y_{i,0}$ generated as *i.i.d.* across i with $y_{i,0} \sim N(0,1)$.

Finally, we use combinations of (N,T) from $\{10, 20, 40, 80\} \times \{25, 50, 100\}$.

Let $\hat{\varphi}_j$ denote the estimate of the true value of φ (say φ_0) for iteration j of the Monte Carlo experiment, with $j = 1, ..., MC$. We report the following measures

(5.1)
$$bias = \frac{1}{MC} \sum_{j=1}^{MC} (\hat{\varphi}_j - \varphi),$$

$$(5.2) \qquad MSE \quad = \quad \frac{1}{MC} \sum_{j=1}^{MC} \left(\hat{\varphi}_j - \varphi \right)^2.$$

In addition to this, in order to assess the finite sample validity of the estimator of the asymptotic variance suggested by Theorem 2.1, we also consider the empirical rejection frequency (for a nominal size of 5%) of a t-test for $\varphi = \varphi_0$. This is tantamount to verifying the empirical coverage of 95% confidence intervals for φ_0. In our simulations, we set $MC = 2000$; this entails that the empirical rejection frequencies reported here have a 95% confidence interval given by $[0.04, 0.06]$.

[Insert Tables 1-3 somewhere here]

The tables shows that the estimator $\hat{\varphi}$, and the random norming suggested by Theorem 3.4, have excellent properties even for very small samples.

Considering first bias and MSE as defined in (5.1) and (5.2), we note that, as expected, they decline as either N or T increases. The trend is similar across the tables, thereby suggesting that the presence and pervasiveness of common factors does not impact on the decline of either bias or MSE. Observing the numbers in the table, the rate of decline of the MSE is the same as either N or T increases. This changes when $\gamma = 10$ (Table 3), and the impact of N becomes less significant in decreasing the MSE - this however can be expected by virtue of the fact that the asymptotics is driven by T only. Indeed, although not predicted by the theory, the WLS estimator seems to have the desirable property that its quality improves as N increases even in the presence of cross sectional dependence. The MSE and the bias do not seem to be affected by the value of φ (one, minor, exception could be the case $(N, T) = (10, 25)$ in Table 1), which confirms that the estimator proposed in this paper, due to its self-normalised nature, is not affected by unit or explosive roots. As far as the bias is concerned, we note that, although in the Tables there are only raw numbers, it seems to be rather small when compared with the value of φ; this is true even for the case $(N, T) = (10, 25)$.

Turning to the empirical rejection frequencies, as pointed out above these can be viewed as an assessment on the quality of the estimated asymptotic variance of $\hat{\varphi}$, especially since the bias is quite small. In general, the empirical rejection frequencies do not change across the Tables, showing that the estimator of the asymptotic variance is, as can be expected, not affected by the presence of common factors. The empirical size of the t-tests is always close to the nominal one, with few exceptions in the case of small T: typically, when $T \geq 50$, the problem disappears across all experiments.

Finally, we also ran a set of experiments for the case $b_{i,t} = 0$. The setup of the Monte Carlo exercise is the same as above (apart from setting $b_{i,t} = 0$), although we report only the cases of $\gamma_i = \gamma = 0$

and $\gamma_i = \gamma = 10$, to better illustrate the difference in the results brought about by cross sectional dependence. Also, we only report results for $N \geq 20$ and $T \geq 50$; this is essentially in order to save space: results for $N = 10$ and $T = 25$ are usually good with the exception of the case $|\varphi| = 1.5$, where $N = 10$ does not seem to be sufficient to ensure that the CLT holds.

[**Insert Table 4 somewhere here**]

The results in Table 4 can be contrasted with Theorem 3.3, and, when $|\varphi| < 1$, with the results in Tables 1 and 3. The MSE appears to be lower, but this is due to the "natural" effect of having $\sigma_i^2 = 0$, so that slope heterogeneity does not contribute to the asymptotic variance of the WLS estimator. Considering first the MSEs for $|\varphi| \geq 1$, when there is no cross sectional dependence ($\gamma = 0$), the results in Table 4 show a great improvement with respect to those in Table 1 for all cases where $|\varphi| \geq 1$. When there is cross sectional dependence ($\gamma = 10$), the MSE improves (compared to Table 3) when $|\varphi| > 1$, and also and when $|\varphi| = 1$ and $N \geq 40$ - the results confirm the faster convergence of the WLS estimator in presence of a homogeneous root (unit or explosive), although it should be noted that increasing N alone does not yield almost any improvements. The faster rates of convergence also emerge when comparing numbers within Table 4: the MSE for the cases of $|\varphi| < 1$ are one or two orders of magnitude larger than in the case of $|\varphi| = 1$ or $|\varphi| > 1$ respectively. All results worsen as we move from $\gamma = 0$ to $\gamma = 10$, as a consequence of cross sectional dependence. Turning to the empirical rejection frequencies, these are always within the confidence interval $[0.04, 0.06]$ when $|\varphi| < 1$, with few exceptions. Similarly, confidence intervals have almost always the correct coverage when $|\varphi| = 1$; the same results can be observed when $|\varphi| > 1$ and $\gamma = 0$. As predicted by Theorem 3.3, when $|\varphi| > 1$ and there is strong cross sectional dependence ($\gamma = 10$), the CLT fails and this is evident from the severely undersized empirical rejection frequencies.

A final comment on the simulations. We ran a separate, unreported exercise where we assess the robustness of results when altering the weighing scheme in (2.2) to $1/\left(a + y_{i,t-1}^2\right)$ and adding the fixed effects $\alpha_i \sim N(1,1)$. We tried several values of a, namely $a \in \{0.1, 0.5, 1, 2, 10\}$, showing that results remain virtually unchanged across all choices of a.

6. EMPIRICAL APPLICATION

In this section, we illustrate the results derived above by considering two applications of (1.1).

We firstly estimate the average autoregressive root in an RCA model applied to several macroeconomic and financial time series for several EU countries. Specifically, we consider the following series: log

of the GDP; log of the M2 money aggregate; log of the main Equity Index; short term interest rate (expressed in percentage); and the log of the Industrial Production. Data are quarterly, and in order to show the impact of different sample sizes on the WLS estimator and its confidence intervals, we consider the following time spans:

- **-:** log of the GDP (raw data are expressed in millions of $): from 1960Q1 until 2013Q1;

- **-:** log of the M2 money aggregate: (raw data are expressed in local currency, except Germany, expressed in billions of €): from 1999Q1 until 2013Q1;

- **-:** log of the Equity index (raw data are based at 100 in 2005): from 1994Q1 to 2013Q1;

- **-:** short term interest rate (data are expressed in percentage): from 1992Q4 to 2013Q1;

- **-:** log of the Industrial Production index: from 1990Q1 until 2013Q1.

As far as the cross sectional sample size is concerned, we consider the following countries: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden and United Kingdom. Thus, in total we have $N = 18$; some of these countries have incomplete datasets for some of the series considered, and in such case we omit them from the panel. In Table 5 below, we specify the true cross sectional sample size for each exercise.

In addition to the exercise described above, we also apply our estimator to verify whether there is a bubble in the UK housing market. The idea that the exuberant dynamics in asset prices could be well represented by an autoregressive process with a root larger than 1 has been exploited in various contributions, e.g. Phillips *et al.* (2011). According to Banerjee *et al.* (2012), a bubble in model (1.1) would be present when $\varphi > 1$; a similar analysis is also contained in Charemza and Deadman (1995). In our application, we consider UK quarterly data from 1997Q1 to 2008Q1, so that $T = 45$ - there is general consensus that, in that period, house prices had an exuberant growth which should signal the presence of a bubble. Specifically, data are (logs of) the Nationwide house price index, and are disaggregated at regional level; in total, we have $N = 13$ regions - North, Yorkshire and the Humber, North West, East Midlands, West Midlands, East Anglia, South East, Outer Metropolitan Area, London, South West, Wales, Scotland and Northern Ireland.

For both exercises, we use the rebased versions of the series as defined in (4.3). We report: the estimated φ; its standard error; the 95% confidence interval; and, finally, the test statistic for the null $H_0 : \varphi \leq 1$ and its p-value.

[Insert Table 5 somewhere here]

Consider the first panel of the table. For all the series, with the exceptions of the short term interest rate, it holds that $\varphi \geq 1$, based on the confidence intervals computed using (3.10). Interestingly, the short term interest rate seems to have "near unit root" behaviour, based on the estimated *average* autoregressive root; indeed, this only represents an average behaviour, and some countries are bound to be in the "boundary case" described above, or even, possibly, in the explosive one. Similarly, the results seem to indicate that, as far as the demand for money M2, and the log of industrial production are concerned, φ is significantly larger than 1: this is based on the t-test for φ; confidence intervals also reinforce this finding. Again, we note that even for these series φ is very close to the boundary. As a general comment, confidence intervals are, in general, short, which confirms the idea that the panel-based approach is bound to improve inference, and the findings in the simulations in Section 5. The results indicate that the absence of the unit root problem in the WLS-based estimate of φ is advantageous in this setting.

Turning to the second panel of the table, this contains the inference on the presence of a bubble in the UK housing market. We find significant evidence that $\varphi > 1$, thereby indicating that a bubble was indeed present on the UK housing market. This is made evident by the rejection, at the 5% level, of the null of no explosive behaviour $H_0 : \varphi \leq 1$, and also by confidence intervals. It can however be noted that the average root is relatively close to 1; this is consistent with the literature on bubbles (Phillips and Yu, 2009; Phillips *et al.*, 2011), where the underlying autoregressive process is modelled with a local-to-explosive unit.

7. Conclusions

In this paper, we have studied the WLS-based estimation of the average autoregressive root in a panel RCA model. We have shown that the estimator is always consistent, irrespective of the true value of the root φ, of whether the autoregressive root is genuinely random or fixed, and on the possible presence and extent of cross sectional dependence. Indeed, our paper proposes a "universal" estimator of the asymptotic variance of the estimated φ. When normalising the WLS estimator by the proposed estimator of the asymptotic variance, standard normal inference is recovered, with the only exception of panels with a common, explosive root and strong cross sectional dependence. The robustness of the WLS estimator comes, however, at a price: rates of convergence are somehow sacrificed, since the weighing scheme employed serves the purpose of anchoring down summations involving $y_{i,t}$ in the cases where $|\varphi| \geq 1$, making them of comparable magnitude with the stationary case. This is quite evident in the cases where there is no randomness in the autoregressive root ($b_{i,t} = 0$), and in particular in the case of an explosive root ($|\varphi| > 1$), where the estimator is shown to be T-consistent as opposed to having

the exponential rate which is typical of OLS. Nonetheless, simulations show that the estimator has very good properties, even for small samples, and that confidence intervals based on the estimated asymptotic variance have the correct coverage almost under all circumstances. Although the focus of this paper is mainly theoretical, we show how the estimator can be applied to several macro and financial time series; in addition, Section 6 also discusses the application of our methodology to testing for bubbles.

Finally, we note that several interesting questions are still outstanding. As mentioned in Section 2, the weighing scheme proposed in (2.2) is not aimed at achieving efficiency, which would require employing $1/\left[\left(E\gamma_i^2 + \tau_i^2\right) + \sigma_i^2 y_{i,t-1}^2\right]$ as a weight function. However, since φ is estimated consistently using the weights proposed in (2.2), the variance $\left(E\gamma_i^2 + \tau_i^2\right) + \sigma_i^2 y_{i,t-1}^2$ could be also estimated equation by equation, thereby casting the WLS estimator in a two-step, or even iterative, procedure, although unreported Monte Carlo evidence based on using the weight $1/\left(a + y_{i,t-1}^2\right)$ showed virtually no change for different values of $a > 0$. Further, in this paper we do not try to eliminate the common component $\gamma_i v_t$, unlike e.g. Pesaran (2006) and Bai (2009; see also Song, 2013). Indeed, this feature of the WLS estimator is the reason why we require the assumption that the regressors $x_{i,t}$ are independent of $\gamma_i v_t$. However, the main technical results in the paper (the coupling arguments used in Lemmas 7.2-7.3, and the concentration inequalities in Lemma 7.4) do not require any assumptions on the (in)dependence between $x_{i,t}$ and v_t, and therefore they automatically hold even in this case. Adapting the WLS estimator to the presence of interactive effects is therefore possible, by modifying the estimation problem (4.4) into

$$\min_{\varphi,\beta_1,...,\beta_N,\gamma_1,...,\gamma_N,v_1,....,v_T} S\left(\varphi,\beta_1,...,\beta_N,\gamma_1,...,\gamma_N,v_1,.....,v_T\right),$$

with

$$S\left(\varphi,\beta_1,...,\beta_N,\gamma_1,...,\gamma_N,v_1,.....,v_T\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\left(\tilde{y}_{i,t} - \varphi\tilde{y}_{i,t-1} - \beta_i' x_{i,t} - \gamma_i v_t\right)^2}{1 + \tilde{y}_{i,t-1}^2};$$

the estimation of φ and the β_is (and of the factor-loading structure) can be carried out by iterative concentration of the likelihood, as in Bai (2009) and Song (2013). This extension is currently under investigation by the authors.

Acknowledgement

Appendix: technical results and proofs

This appendix contains the proofs of the results in Section 3; prior to reporting the proofs, we lay out some technical lemmas whose proofs can be found in the Supplement; all the proofs of results in Section 4 are in the Supplement.

We often employ the following notation:

$$(7.1) \qquad A_{N,T} = \varphi B_{N,T} + A_{N,T}(1) + A_{N,T}(2) + A_{N,T}(3),$$

where $A_{N,T}(1) \equiv \sum_{i=1}^{N} \sum_{t=2}^{T} b_{i,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}$; $A_{N,T}(2) \equiv \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \frac{y_{i,t-1}}{1+y_{i,t-1}^2}$; and $A_{N,T}(3) \equiv \sum_{i=1}^{N} \sum_{t=2}^{T} \gamma_i v_t \frac{y_{i,t-1}}{1+y_{i,t-1}^2}$.

Lemma 7.1. *Under Assumption 1(i)-(ii)-(iii)-(iv)(b), we have* $EA_{N,T}(1) = EA_{N,T}(2) = EA_{N,T}(3) = 0$. *Also*

 (i): $EA_{N,T}^2(1) \leq T \sum_{i=1}^{N} \sigma_i^2$;

 (ii): $EA_{N,T}^2(2) \leq T \sum_{i=1}^{N} \tau_i^2$;

 (iii): $EA_{N,T}^2(3) \leq T \left(\sum_{i=1}^{N} E|\gamma_i| \right)^2$.

Lemma 7.2. *Under Assumptions 1 and 2, for all* $i \in C(1)$, *it holds that*

$$(7.2) \qquad E|y_{i,t} - \bar{y}_{i,t}|^{\kappa_i} \leq q_{i,t},$$

with

$$q_{i,t} = E|y_{i,0}|^{\kappa_i} \delta_i^t + \frac{E|e_{i,0}|^{\kappa_i} + E|\bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i}}{1 - \delta_i} \delta_i^t + E|\gamma_i - \bar{\gamma}_i|^{\kappa_i} \frac{E|v_0|^{\kappa_i}}{1 - \delta_i}.$$

Also, let $g(\cdot)$ *and* $g'(\cdot)$ *be a function and its first derivative, both bounded on the real line, and* C *is a constant that only depends on* g; *it holds that*

$$(7.3) \qquad E|g(y_{i,t}) - g(\bar{y}_{i,t})| \leq C q_{i,t}^{1/(1+\kappa_i)}.$$

Lemma 7.3. *Under Assumptions 1 and 2, for all* $\beta \geq 1$, *it holds that*

$$\frac{1}{NT} \sum_{i \in C(1)} \sum_{t=1}^{T} \left(\frac{y_{i,t}^2}{1+y_{i,t}^2} \right)^\beta = \frac{1}{N} \sum_{i \in C(1)} E \left(\frac{\bar{y}_{i,0}^2}{1+\bar{y}_{i,0}^2} \right)^\beta + o_P(1).$$

Lemma 7.4. *Let* c_0 *denote an absolute constant. Under Assumption 1:*

(i) if $i \in C(2)$, $Eb_{i,t}^2 > 0$ *and Assumptions 3 and 4 hold then for all* $x > 0$ *and* $t \geq [2(\bar{\sigma}_i + 1)/E \log |\varphi + b_{i,0}|]^3$, *it holds that*

$$P\{|y_{i,t}| \leq x\} \leq 2x M_i \left\{ \exp(-t^{3/(2(1+\nu_i))}) + t^{-2} + c_0 2^{2\nu_i} m_i t^{-(\nu_i - 2)/(2(1+\nu_i))} \right\}.$$

(ii) if $i \in C(2)$, $b_{i,t} = 0$, and Assumptions 3 and 4(i) hold, it holds that, for some $\nu' \geq 4$ such that $E|u_{i,0}|^{\nu'} \leq m'_i$

$$P\{|y_{i,t}| \leq x\} \leq \sqrt{\frac{2}{\pi}} \int_0^{x/\tau_i \sqrt{t}} \exp\left(-\frac{1}{2}u^2\right) du + \frac{tc_0}{x^{\nu'}} 2^{\nu'} m'_i.$$

(iii) if $i \in C(3)$ and $Eb_{i,t}^2 > 0$, and Assumption 4 holds then for all $x > 0$ and $t = 1, 2, ...$, it holds that

$$P\{|y_{i,t}| \leq x\} \leq 2x M_i \left\{ \exp(-t^{3/(2(1+\nu_i))}) + \left[\frac{1}{\bar{\sigma}_i} + c_0 2^{2\nu_i} m_i\right] t^{-(\nu_i-2)/(2(1+\nu_i))} \right\}.$$

(iv) if $i \in C(3)$ and $b_{i,t} = 0$ and Assumption 4(i) holds then for all $x > 0$ and $t = 1, 2, ...$, it holds that

$$P\{|y_{i,t}| \leq x\} \leq 2x |\varphi| M_i \left\{ e^{-t \ln|\varphi|} \right\}.$$

Lemma 7.5. *Under Assumptions 1, 3 and 4, for all $\beta \geq 1$ it holds that*

$$\frac{1}{NT} \sum_{i \in C(2)} \sum_{t=2}^T \left(\frac{y_{i,t}^2}{1 + y_{i,t}^2}\right)^\beta = \frac{\#C(2)}{N} + o_P(1).$$

Lemma 7.6. *Let Assumptions 1-3 hold; further, if either $Eb_{i,t}^2 > 0$ and Assumption 4 holds, or $b_{i,t} = 0$ and Assumption 4(i) holds, then, for all $\beta \geq 1$*

$$\frac{1}{NT} \sum_{i \in C(3)} \sum_{t=2}^T \left(\frac{y_{i,t}^2}{1 + y_{i,t}^2}\right)^\beta = \frac{\#C(3)}{N} + o_P(1).$$

Lemma 7.7. *Under Assumptions 1-4, it holds that $\frac{B_{N,T}}{NT} \xrightarrow{P} a_0$, where a_0 is defined in (3.1).*

Theorem 7.1. *Let $S_t = \{S_{t,T}, \mathcal{H}_{t,T}\}, 1 \leq t \leq T$ be zero–mean, square integrable martingale array with differences $x_{t,T}, 1 \leq t \leq T$, and $\mathcal{H}_{t,T} \subseteq \mathcal{H}_{t+1,T}, 0 \leq t < T$. Suppose that, as $T \to \infty$,*

$$(7.4) \qquad \sum_{t=1}^T E\left(x_{t,T}^2 | \mathcal{H}_{t-1,T}\right) \xrightarrow{P} a > 0,$$

$$(7.5) \qquad \sum_{t=1}^T E\left(|x_{t,T}|^{2+\epsilon} | \mathcal{H}_{t-1,T}\right) = o_P(1),$$

with some $\epsilon > 0$. Then, as $T \to \infty$, $S_T \xrightarrow{\mathcal{D}} N(0, a)$.

We are now ready to report the proofs of the main results in the paper.

Proof of Theorem 3.1. It follows from Lemma 7.1 and Assumptions 1(ii)(b) and 1(iv)(b) that $N^{-1}T^{-1/2}A_{N,T} = O_P(1)$, as $\min(N, T) \to \infty$. Since $T \to \infty$ we conclude from Lemma 7.7 that $\frac{A_{N,T}}{B_{N,T}} = O_P(T^{-1/2}) =$

$o_P(1)$, so Theorem 3.1 now follows from the definition $\hat{\varphi}_{N,T}$. $\qquad\square$

Proof of Theorem 3.2. We start with part *(i)* of the theorem. According to Lemmas 7.1 and 7.7 it is enough to prove that $N^{-1}T^{-1/2}A_{N,T}(3) \xrightarrow{\mathcal{D}} N(0, a_1)$. This will be based on Theorem 7.1. Let $\mathcal{H}_t$, $t \geq 1$, be the σ–algebra generated by $\{y_{i,0}, e_{i,s}, b_{i,s}, \gamma_i, v_s, \ 1 \leq i \leq N, \ s \leq t\}$. We write $N^{-1}T^{-1/2}A_{N,T}(3) = \sum_{t=2}^{T} x_t$, where $x_t = \frac{1}{NT^{1/2}} v_t \sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2}$. Clearly, since $Ev_0^2 = 1$ and $y_{i,t-1}$ is $\mathcal{H}_{t-1}$–measurable,

$$E(x_t^2|\mathcal{H}_{t-1}) = \frac{1}{N^2T}\left(\sum_{i=1}^{N}\gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2 = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_i\gamma_j \frac{y_{i,t-1}}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}.$$

In order to apply Theorem 7.1, we need to show

$$(7.6) \qquad \sum_{t=2}^{T} E(x_t^2|\mathcal{H}_{t-1}) \xrightarrow{P} a_1.$$

Using Lemma 7.4 we get that

$$\frac{1}{N^2T}\left|\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_i\gamma_j \frac{y_{i,t-1}}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2} - \sum_{t=2}^{T}\sum_{i\in C(1)}\sum_{j\in C(1)}\gamma_i\gamma_j \frac{y_{i,t-1}}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right| = o_P(1),$$

i.e. the non–stationary units do not contribute to the limit, since $|y_{i,t}| \to \infty$ in probability as $t \to \infty$ for all $i \in C(2) \cup C(3)$. Next, repeating the arguments used in the proof of Lemma 7.3 we obtain that

$$\frac{1}{N^2T}\sum_{t=2}^{T}\sum_{i\in C(1)}\sum_{j\in C(1)}|\gamma_i\gamma_j|\left|\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2} - \frac{\bar{y}_{i,t-1}}{1+\bar{y}_{i,t-1}^2}\frac{\bar{y}_{j,t-1}}{1+\bar{y}_{j,t-1}^2}\right| = o_P(1),$$

i.e. we can replace $y_{i,t}$ with the stationary solution $\bar{y}_{i,t}$ in case of stationary units. Hence (7.6) is established if we show that

$$(7.7) \qquad \frac{1}{N^2T}\sum_{t=2}^{T}\sum_{i\in C(1)}\sum_{j\in C(1)}\gamma_i\gamma_j \frac{\bar{y}_{i,t-1}}{1+\bar{y}_{i,t-1}^2}\frac{\bar{y}_{j,t-1}}{1+\bar{y}_{j,t-1}^2} \xrightarrow{P} a_1.$$

Recall that the subscript "γ" denotes conditioning on $\{\gamma_i\}_{i=1}^{N}$. Elementary arguments give that

$$E_\gamma\left(\frac{1}{N^2T}\sum_{t=2}^{T}\sum_{i\in C(1)}\sum_{j\in C(1)}\bar{\gamma}_i\bar{\gamma}_j z_{i,j,t-1}\right)^2 = \frac{1}{N^4T^2}\sum_{s,t=2}^{T}\sum_{i\in C(1)}\cdots\sum_{\ell\in C(1)}\bar{\gamma}_i\bar{\gamma}_j\bar{\gamma}_k\bar{\gamma}_\ell E_\gamma z_{i,j,t-1}z_{k,\ell,s-1}$$

$$\leq \frac{2}{N^4T}\sum_{i\in C(1)}\cdots\sum_{\ell\in C(1)}\sum_{h=0}^{\infty}|\bar{\gamma}_i\bar{\gamma}_j\bar{\gamma}_k\bar{\gamma}_\ell||E_\gamma z_{i,j,0}z_{k,\ell,h}|$$

with

$$z_{i,j,t-1} = \frac{\bar{y}_{i,t-1}}{1+\bar{y}_{i,t-1}^2}\frac{\bar{y}_{j,t-1}}{1+\bar{y}_{j,t-1}^2} - E_\gamma \frac{\bar{y}_{i,t-1}}{1+\bar{y}_{i,t-1}^2}\frac{\bar{y}_{j,t-1}}{1+\bar{y}_{j,t-1}^2}.$$

The stationary solutions $\bar{y}_{i,t}$ satisfy the equations

$$\bar{y}_{i,t} = \bar{y}_{i,0} \prod_{s=1}^{t}(\varphi + b_{i,s}) + \sum_{s=1}^{t} e_{i,s} \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}) + \bar{\gamma}_i \sum_{s=1}^{t} v_s \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}).$$

Let $\hat{y}_{1,0}, \ldots, \hat{y}_{N,0}$ be independent of $\{b_{i,s}, e_{i,s}, \gamma_i v_s\}$, $1 \leq i \leq N$, $s > 0$ and fulfill $\{\hat{y}_{1,0}, \ldots, \hat{y}_{N,0}\} \overset{\mathcal{D}}{=} \{\bar{y}_{1,0}, \ldots, \bar{y}_{N,0}\}$. Next we define

$$\hat{y}_{i,t} = \hat{y}_{i,0} \prod_{s=1}^{t}(\varphi + b_{i,s}) + \sum_{s=1}^{t} e_{i,s} \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}) + \bar{\gamma}_i \sum_{s=1}^{t} v_s \prod_{z=s}^{t-1}(\varphi + b_{i,z+1}).$$

It is clear from the definition that $\{\hat{y}_{1,t}, \ldots, \hat{y}_{N,t}, t > 0\} \overset{\mathcal{D}}{=} \{\bar{y}_{1,t}, \ldots, \bar{y}_{N,t}, t > 0\}$, and $\{\hat{y}_{1,t}, \ldots, \hat{y}_{N,t}, t > 0\}$ is independent of $\{\bar{y}_{1,0}, \ldots, \bar{y}_{N,0}\}$, conditional on γ_i, $1 \leq i \leq N$. Let

$$\hat{z}_{i,j,t-1} = \frac{\hat{y}_{i,t-1}}{1 + \hat{y}_{i,t-1}^2} \frac{\hat{y}_{j,t-1}}{1 + \hat{y}_{j,t-1}^2} - E_\gamma \frac{\bar{y}_{i,t-1}}{1 + \bar{y}_{i,t-1}^2} \frac{\bar{y}_{j,t-1}}{1 + \bar{y}_{j,t-1}^2}.$$

On account of the independence (conditional on $\{\gamma_i\}_{i=1}^N$) of $z_{i,j,0}$ and $\hat{z}_{k,\ell,h}$, and the facts that $Ez_{i,j,0} = 0$ and $|z_{i,j,0}| \leq 2$, we get $|E_\gamma z_{i,j,0} z_{k,\ell,h}| = |E_\gamma z_{i,j,0}(z_{k,\ell,h} - \hat{z}_{k,\ell,h})| \leq 2E_\gamma|z_{k,\ell,h} - \hat{z}_{k,\ell,h}|$. It follows from the definition of $\hat{z}_{k,\ell,h}$ that for all ϵ_1 and ϵ_2 we get

$$(7.8) \qquad E_\gamma|z_{k,\ell,h} - \hat{z}_{k,\ell,h}| \leq \epsilon_1 + 2P_\gamma\{|\bar{y}_{k,h} - \hat{y}_{k,h}| > \epsilon_1\} + \epsilon_2 + 2P_\gamma\{|\bar{y}_{\ell,h} - \hat{y}_{\ell,h}| > \epsilon_2\}.$$

By Hardy *et al.* (1959, p. 32) and (2.7) for all $1 \leq i \leq N$ we have

$$E_\gamma|\bar{y}_{i,h} - \hat{y}_{i,h}|^{\kappa_i} = E_\gamma \left| (\bar{y}_{i,0} - y_{i,0}) \prod_{s=1}^{h}(\varphi + b_{i,s}) \right|^{\kappa_i} = E_\gamma|\bar{y}_{i,0} - \hat{y}_{i,0}|^{\kappa_i} E_\gamma \left| \prod_{s=1}^{h}(\varphi + b_{i,s}) \right|^{\kappa_i}$$

$$\leq 2E_\gamma|\bar{y}_{i,0}|^{\kappa_i} \prod_{s=1}^{h} E_\gamma|\varphi + b_{i,s}|^{\kappa_i} \leq 2E_\gamma|\bar{y}_{i,0}|^{\kappa_i} \delta_i^h.$$

Using again (2.6) we conclude

$$E_\gamma|\bar{y}_{i,0}|^{\kappa_i} \leq \sum_{s=-\infty}^{0} E|e_{i,0}|^{\kappa_i} \delta_i^{|s|} + |\bar{\gamma}_i|^{\kappa_i} \sum_{i=-\infty}^{0} E|v_0|^{\kappa_i} \delta_i^{|s|} = \frac{E|e_{i,0}|^{\kappa_i} + |\bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i}}{1 - \delta_i}.$$

Applying now Markov's inequality we have for all $x > 0$

$$P_\gamma\{|\bar{y}_{i,h} - \hat{y}_{i,h}| > x\} = P_\gamma\{|\bar{y}_{i,h} - \hat{y}_{i,h}|^{\kappa_i} > x^{\kappa_i}\} \leq \frac{2}{x^{\kappa_i}} \frac{E|e_{i,0}|^{\kappa_i} + |\bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i}}{1 - \delta_i} \delta_i^h.$$

So by (7.8) there is an absolute constant C such that

$$E_\gamma|z_{k,\ell,h} - \hat{z}_{k,\ell,h}|$$

$$\leq C \left\{ \left[\frac{E|e_{k,0}|^{\kappa_k} + |\bar{\gamma}_k|^{\kappa_k} E|v_0|^{\kappa_k}}{1 - \delta_k} \delta_k^h \right]^{1/(1+\kappa_k)} + \left[\frac{E|e_{\ell,0}|^{\kappa_\ell} + |\bar{\gamma}_\ell|^{\kappa_\ell} E|v_0|^{\kappa_\ell}}{1 - \delta_\ell} \delta_\ell^h \right]^{1/(1+\kappa_\ell)} \right\}.$$

Thus we conclude

$$(7.9) \quad E_\gamma \left(\frac{1}{N^2 T} \sum_{t=2}^{T} \sum_{i \in C(1)} \sum_{j \in C(1)} \bar{\gamma}_i \bar{\gamma}_j z_{i,j,t-1} \right)^2$$

$$\leq \frac{C}{N^4 T} \left(\sum_{i \in C(1)} |\gamma_i| \right)^3 \sum_{\ell \in C(1)} |\bar{\gamma}_\ell| \left[\frac{E|e_{\ell,0}|^{\kappa_\ell} + |\bar{\gamma}_\ell|^{\kappa_\ell} E|v_0|^{\kappa_\ell}}{1 - \delta_\ell} \right]^{1/(1+\kappa_\ell)} \frac{1}{\delta_\ell^{1/(1+\kappa_\ell)}} = O_P \left(\frac{1}{T} \right).$$

Now (7.7) follows from Chebyshev's inequality, and from the cross sectional independence of the γ_is. Clearly, by Assumption 5(i)

$$E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) = \frac{E|v_0|^{2+\epsilon}}{T^{1+\epsilon/2}} \left(\frac{1}{N} \sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \right)^{2+\epsilon} \leq \frac{E|v_0|^{2+\epsilon}}{T^{1+\epsilon/2}} \left(\frac{1}{N} \sum_{i=1}^{N} |\gamma_i| \right)^{2+\epsilon},$$

which entails that $\sum_{t=2}^{T} E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) \xrightarrow{P} 0$. This, and (7.6), yield the first part of Theorem 3.2.

We continue with the proof of part (ii). Under part (ii) and Lemmas 7.1 and 7.7, we conclude that (3.4) is established if $\frac{A_{N,T}(3)}{(r_N T)^{1/2}} \xrightarrow{\mathcal{D}} N(0,1)$. According to Theorem 7.1 we only need to prove that

$$(7.10) \qquad \sum_{t=2}^{T} E(x_t^2|\mathcal{H}_{t-1}) \xrightarrow{P} 1.$$

$$(7.11) \qquad \sum_{t=2}^{T} E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) \xrightarrow{P} 0,$$

where $x_t = \frac{1}{(r_N T)^{1/2}} v_t \sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1 + y_{i,t-1}^2}$. As in the proof of the first part of Theorem 3.2, we have

$$E(x_t^2|\mathcal{H}_{t-1}) = \frac{1}{r_N T} \left(\sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \right)^2 = \frac{1}{r_N T} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \frac{y_{j,t-1}}{1 + y_{j,t-1}^2}.$$

It holds that

$$\frac{1}{r_N T} \left| \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \frac{y_{j,t-1}}{1 + y_{j,t-1}^2} - \sum_{t=2}^{T} \sum_{i \in C(1)} \sum_{j \in C(1)} \gamma_i \gamma_j \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \frac{y_{j,t-1}}{1 + y_{j,t-1}^2} \right|$$

$$\leq \frac{2}{r_N T} \sum_{j=1}^{N} |\gamma_j| \sum_{i \in C(2) \cup C(3)} \sum_{t=2}^{T} |\gamma_i| \frac{|y_{i,t-1}|}{1 + y_{i,t-1}^2}.$$

By Assumption $1(iv)$(b), $\sum_{j=1}^{N} |\gamma_j| = O_P(N)$. Using Lemma 7.4$(i)$ with $x = t^{(\nu_i - 2)/(4(\nu_i + 1))}$ we obtain

$$
E \sum_{i \in C(3)} \sum_{t=2}^{T} |\gamma_i| E_\gamma \frac{|y_{i,t-1}|}{1 + y_{i,t-1}^2} \leq 2 \sum_{i \in C(3)} \sum_{t=2}^{T} t^{(\nu_i - 2)/(4(\nu_i + 1))} E|\gamma_i|(1 + M_i) \Bigg\{ \exp(-t^{3/(2(1+\nu_i))})
$$

$$
+ \left[\frac{1}{\bar{\sigma}_i} + c_0 2^{2\nu_i} m_i + 1 \right] t^{-(\nu_i - 2)/(2(\nu_i + 1))} \Bigg\}
$$

$$
\leq CT^{1-(\underline{\nu}-2)/(4(\underline{\nu}+1))} \sum_{i \in C(3)} E|\gamma_i|(1 + M_i) \left[\frac{1}{\bar{\sigma}_i} + m_i + 1 \right]
$$

with some constant C. Similarly, Lemma 7.4(i) with $x = t^{(\nu_i - 2)/(4(\nu_i + 1))}$ yields

$$
E \sum_{i \in C(2)} \sum_{t=2}^{T} |\gamma_i| E_\gamma \frac{|y_{i,t-1}|}{1 + y_{i,t-1}^2} \leq CT^{1-(\underline{\nu}-2)/(4(\underline{\nu}+1))} \sum_{i \in C(2)} E|\gamma_i|(1 + M_i) \left[\frac{1}{\bar{\sigma}_i} + m_i + 1 \right]
$$

$$
+ \sum_{i \in C(2)} [2(\bar{\sigma}_i + 1)/E \log|\varphi + b_{i,0}|]^3.
$$

Next we replace $y_{i,t}$ with $\bar{y}_{i,t}$, $i \in C(1)$. Note

$$
\frac{1}{r_N T} \sum_{t=2}^{T} \sum_{i \in C(1)} \sum_{j \in C(1)} |\gamma_i \gamma_j| \left| \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} \frac{y_{j,t-1}}{1 + y_{j,t-1}^2} - \frac{\bar{y}_{i,t-1}}{1 + \bar{y}_{i,t-1}^2} \frac{\bar{y}_{j,t-1}}{1 + \bar{y}_{j,t-1}^2} \right|
$$

$$
\leq \frac{1}{r_N T} \sum_{j \in C(1)} |\gamma_j| \sum_{i \in C(1)} |\gamma_i| \sum_{t=1}^{T} \left| \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} - \frac{\bar{y}_{i,t-1}}{1 + \bar{y}_{i,t-1}^2} \right|;
$$

we have

$$
\sum_{i \in C(1)} |\gamma_i| \sum_{t=1}^{T} E_\gamma \left| \frac{y_{i,t-1}}{1 + y_{i,t-1}^2} - \frac{\bar{y}_{i,t-1}}{1 + \bar{y}_{i,t-1}^2} \right|
$$

$$
\leq C_2 \sum_{i \in C(1)} |\gamma_i| \left\{ \frac{(E|y_{i,0}|^{\kappa_i})^{1/(1+\kappa_i)}}{(1 - \delta_i)^{1/(1+\kappa_i)}} + \frac{(E|e_{i,0}|^{\kappa_i} + |\bar{\gamma}_i|^{\kappa_i} E|v_0|^{\kappa_i})^{1/(1+\kappa_i)}}{(1 - \delta_i)^{1/(1+\kappa_i)}} \right\} \frac{1}{1 - \delta_i^{1/(1+\kappa_i)}}
$$

$$
+ C_3 T \sum_{i \in C(1)} \left(|\gamma_i - \bar{\gamma}_i|^{\kappa_i} \frac{E|v_0|^{\kappa_i}}{1 - \delta_i} \right)^{1/(1+\kappa_i)}.
$$

Putting all together, it follows that (7.10) is proven if we show

$$
(7.12) \qquad \frac{1}{r_N T} \sum_{t=2}^{T} \sum_{i \in C(1)} \sum_{j \in C(1)} \gamma_i \gamma_j \frac{\bar{y}_{i,t-1}}{1 + \bar{y}_{i,t-1}^2} \frac{\bar{y}_{j,t-1}}{1 + \bar{y}_{j,t-1}^2} \xrightarrow{P} 1;
$$

this follows by repeating the proof of (7.7), whence

$$
\left(\frac{N^2}{r_N} \right)^2 E_\gamma \left(\frac{1}{N^2 T} \sum_{t=2}^{T} \sum_{i \in C(1)} \sum_{j \in C(1)} \bar{\gamma}_i \bar{\gamma}_j z_{i,j,t-1} \right)^2 = O \left[\frac{1}{T} \left(\frac{N^2}{r_N} \right)^2 \right],
$$

so that (7.12) is implied by the same argument as above.

The proof of (7.11) is very simple since

$$E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) = \frac{E|v_0|^{2+\epsilon}}{(r_N T)^{1+\epsilon/2}} \left(\sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^{2+\epsilon} \le (N^2)^{1+\epsilon/2} \frac{E|v_0|^{2+\epsilon}}{(r_N T)^{1+\epsilon/2}} \left(\frac{1}{N} \sum_{i=1}^{N} |\gamma_i| \right)^{2+\epsilon},$$

hence

$$E\left\{ \sum_{t=2}^{T} E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) \right\} = O\left[\frac{1}{T^{\epsilon/2}} \left(\frac{N^2}{r_N} \right)^{1+\epsilon/2} \right].$$

Now Markov's inequality gives (7.11).

In order to prove part *(iii)* of the theorem, we need to show that

(7.13)
$$\frac{1}{(NT)^{1/2}} \{A_{N,T}(1) + A_{N,T}(2) + A_{N,T}(3)\} \xrightarrow{\mathcal{D}} N(0, a_2),$$

where a_2 is defined in Theorem 3.2. First we write

$$\frac{1}{(NT)^{1/2}} \{A_{N,T}(1) + A_{N,T}(2) + A_{N,T}(3)\} = \sum_{t=2}^{T} x_t,$$

where

$$x_t = \frac{1}{(NT)^{1/2}} \left\{ \sum_{i=1}^{N} b_{i,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} + \sum_{i=1}^{N} e_{i,t} \frac{y_{i,t-1}}{1+y_{i,t-1}^2} + v_t \sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right\}.$$

Using Assumptions 1*(i)-(iii)*, we obtain that

$$E(x_t^2|\mathcal{H}_{t-1}) = \frac{1}{NT} \sum_{i=1}^{N} \sigma_i^2 \left(\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \right)^2 + \frac{1}{NT} \sum_{i=1}^{N} \tau_i^2 \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 + \frac{1}{NT} \left(\sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2.$$

We prove

(7.14)
$$\sum_{t=2}^{T} E(x_t^2|\mathcal{H}_{t-1}) \xrightarrow{P} a_2$$

The proof of (7.14) starts with

(7.15)
$$\frac{1}{NT} \sum_{t=2}^{T} \sum_{i\in C(1)} \sigma_i^2 \left(\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \right)^2 = \frac{1}{N} \sum_{i\in C(1)} \sigma_i^2 E\left(\frac{\bar{y}_{i,0}^2}{1+\bar{y}_{i,0}^2} \right)^2 + o_P(1).$$

This can be established easily by repeating the proof of Lemma 7.3. The first step is to show that

$$\frac{1}{NT} \sum_{t=2}^{T} \sum_{i\in C(1)} \sigma_i^2 \left| \left(\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \right)^2 - \left(\frac{\bar{y}_{i,t-1}^2}{1+\bar{y}_{i,t-1}^2} \right)^2 \right| = o_P(1)$$

which follows from Assumptions 1*(ii)*(b), 2*(i)* and 2*(ii)*. The truncation argument used in Lemma 7.3 gives that

$$E\left\{\frac{1}{NT}\sum_{t=2}^{T}\sum_{i\in C(1)}\sigma_i^2\left[\left(\frac{\bar{y}_{i,t-1}^2}{1+\bar{y}_{i,t-1}^2}\right)^2-E\left(\frac{\bar{y}_{i,t-1}^2}{1+\bar{y}_{i,t-1}^2}\right)^2\right]\right\}^2=o(1),$$

so Markov's inequality implies (7.15). The next step of the proof of (7.14) is to show that

$$(7.16) \qquad \frac{1}{NT}\sum_{t=2}^{T}\sum_{i\in C(2)\cup C(3)}\sigma_i^2\left(\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\right)^2=\frac{1}{N}\sum_{i\in C(2)}\sigma_i^2+o_P(1);$$

this can be proven along the lines of Lemmas 7.5 and 7.6.

Similarly to (7.15) one can verify that

$$(7.17) \qquad \frac{1}{NT}\sum_{t=2}^{T}\sum_{i\in C(1)}\tau_i^2\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2=\frac{1}{N}\sum_{i\in C(1)}\tau_i^2 E\left(\frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\right)^2+o_P(1),$$

Since $|y_{i,t}|\to\infty$ in probability as $t\to\infty$ for all $i\in C(2)\cup C(3)$, following the proofs of Lemmas 7.5 and 7.6 we obtain

$$(7.18) \qquad \frac{1}{NT}\sum_{t=2}^{T}\sum_{i\in C(2)\cup C(3)}\tau_i^2\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2=o_P(1).$$

To complete the proof of (7.14) we need to show only that

$$\frac{1}{NT}\sum_{t=2}^{T}\left(\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2=\frac{r_N}{N}+o_P(1).$$

The arguments used in the proofs of parts *(i)* and *(ii)* of Theorem 3.2 can be repeated to show that

$$\frac{1}{NT}\sum_{t=2}^{T}\left(\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2=\frac{1}{NT}\sum_{t=2}^{T}\left(\sum_{i\in C(1)}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2+o_P(1),$$

and

$$\frac{1}{NT}\sum_{t=2}^{T}\left(\sum_{i\in C(1)}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2=\frac{r_N}{N}+o_P(1).$$

This also completes the proof of (7.14).

In order to use Theorem 7.1, we now establish that

$$(7.19) \qquad \sum_{t=2}^{T}E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1})=o_P(1).$$

Rosenthal's inequality (cf. Petrov p. 59) yields that

$$(7.20) \quad E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) \leq \frac{C_1}{(NT)^{1+\epsilon/2}} \left\{ \left(\sum_{i=1}^{N} \sigma_i^2 \left[\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \right]^2 \right)^{1+\epsilon/2} + \sum_{i=1}^{N} E|b_{i,0}|^{2+\epsilon} \left| \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \right|^{2+\epsilon} \right. $$

$$+ \left(\sum_{i=1}^{N} \tau_i^2 \left[\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right]^2 \right)^{1+\epsilon/2} + \sum_{i=1}^{N} E|e_{i,0}|^{2+\epsilon} \left| \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right|^{2+\epsilon} $$

$$\left. + E|v_0|^{2+\epsilon} \left| \sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right|^{2+\epsilon} \right\} $$

$$\leq \frac{C_2}{T^{1+\epsilon/2}} \left\{ \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right)^{1+\epsilon/2} + \frac{1}{N} \sum_{i=1}^{N} E|b_{i,0}|^{2+\epsilon} + \left(\frac{1}{N} \sum_{i=1}^{N} \tau_i^2 \right)^{1+\epsilon/2} \right.$$

$$\left. + \frac{1}{N} \sum_{i=1}^{N} E|e_{i,0}|^{2+\epsilon} + \frac{1}{N} \left(\sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 \left(\frac{1}{N} \sum_{i=1}^{N} |\gamma_i| \right)^{\epsilon} \right\}. $$

Since $r_N/N = O(1)$ implies that $\frac{1}{N} \left(\sum_{i=1}^{N} \gamma_i \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 = O_P(1)$, (7.19) is an immediate consequence of Theorem 7.1. $\qquad \square$

Proof of Theorem 3.3. Part *(i)* of the Theorem is just a special case of Theorem 3.2. Consider parts *(ii)* and *(iii)*, and note that, by definition, in these cases $a_0 = 1$. The rate of convergence of $\hat{\varphi}_{N,T} - \varphi$ is driven by

$$(7.21) \qquad \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \frac{y_{i,t-1}}{1+y_{i,t-1}^2} + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \gamma_i v_t \frac{y_{i,t-1}}{1+y_{i,t-1}^2} = I + II. $$

Consider first part *(ii)*. Note, as a preliminary result, that based on Lemma 7.4*(ii)* we have

$$E \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 \leq \frac{1}{x^2} + P\left[|y_{i,t-1}| \leq x\right] = \frac{1}{x^2} + \sqrt{\frac{2}{\pi}} \int_0^{x/\tau_i \sqrt{t}} \exp\left(-\frac{1}{2}u^2\right) du + \frac{tc_0}{x^{\nu_i'}} 2^{\nu_i'} m_i', $$

so that, setting $x = t^{\alpha}$, $\sum_{t=2}^{T} E \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 = O\left(T^{1-2\alpha}\right) + O\left(T^{\alpha+\frac{1}{2}}\right) + O\left(T^{2-\alpha\nu'}\right)$; on account of Assumption 4*(i)*, this can be shown to be bounded by $O\left(T^{2/3}\right)$; this is not the sharpest bound, but it suffices for our purposes. Consider now I in (7.21); this has mean zero and its variance is

$$\frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} E\left(e_{i,t}^2\right) E \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \tau_i^2 \sum_{t=2}^{T} E \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 = O\left(\frac{T^{2/3}}{NT^2} \right), $$

whence $I = O_P\left(N^{-1/2}T^{-2/3}\right)$. Similarly, after repeated applications of the Cauchy-Schwartz inequality to the variance of II (which also has mean zero) we have

$$(7.22) \qquad \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left|\gamma_i\gamma_j\sum_{t=2}^{T}\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)\left(\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right)\right|$$

$$\leq \max_{1\leq i\leq N}\left|\sum_{t=2}^{T}E\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2\right|\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(E\left|\gamma_i\gamma_j\right|^2\right)^{1/2} = O\left(\frac{T^{2/3}}{T^2}\frac{r'_N}{N^2}\right),$$

by (3.6), which entails that $II = O_P\left(T^{-2/3}\sqrt{r'_N/N^2}\right)$. Putting all together, the rate of convergence follows. The limiting distribution can be found by setting

$$(7.23) \qquad x_t = s_N^{-1/2}\sum_{i=1}^{N}w_{i,T}^{-1/2}\left(e_{i,t}+\gamma_i v_t\right)\frac{y_{i,t-1}}{1+y_{i,t-1}^2},$$

where $s_N = \max\{N, r'_N\}$ and

$$(7.24) \qquad w_{i,T} = \sum_{t=2}^{T}E\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2;$$

thus, x_t is an MDS with

$$E(x_t^2|\mathcal{H}_{t-1}) = \frac{1}{s_N}\sum_{i=1}^{N}\tau_i^2 w_{i,T}^{-1}\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2 + \frac{1}{s_N}\left(\sum_{i=1}^{N}w_{i,T}^{-1/2}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2,$$

so that $\sum_{t=2}^{T}E(x_t^2|\mathcal{H}_{t-1})$ can be readily shown to be bounded. Also, by adapting (7.20)

$$E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1}) \leq C\frac{(Ns_N^{-1})^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\tau_i^2\left[\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right]^2\right|^{1+\epsilon/2} + \frac{1}{N^{1+\epsilon/2}}\sum_{i=1}^{N}E|e_{i,0}|^{2+\epsilon}\left|\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2+\epsilon}\right.$$

$$\left. + E|v_0|^{2+\epsilon}\left|\frac{1}{N}\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2+\epsilon}\right\}.$$

Consider $\sum_{t=2}^{T}E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1})$; by virtue of the above we have

$$\frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}\sum_{t=2}^{T}E\left|\frac{1}{N}\sum_{i=1}^{N}\tau_i^2\left[\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right]^2\right|^{1+\epsilon/2}$$

$$\leq \frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}\frac{1}{N}\sum_{i=1}^{N}\tau_i^{2+\epsilon}\sum_{t=2}^{T}E\left|\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2+\epsilon}$$

$$\leq \frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}\frac{1}{N}\sum_{i=1}^{N}\tau_i^{2+\epsilon}\sum_{t=2}^{T}E\left|\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^2 = O\left(\frac{\max_i w_{i,T}}{\min_i w_{i,T}}\min_i w_{i,T}^{-\epsilon/2}\right);$$

also it follows immediately that

$$\frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}\frac{1}{N^{1+\epsilon/2}}\sum_{i=1}^{N}E|e_{i,0}|^{2+\epsilon}\sum_{t=2}^{T}E\left|\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2+\epsilon}=O\left[Ns_N^{-(1+\epsilon/2)}\left(\frac{\max_i w_{i,T}}{\min_i w_{i,T}}\min_i w_{i,T}^{-\epsilon/2}\right)\right];$$

similarly note

$$\frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}E|v_0|^{2+\epsilon}\sum_{t=2}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2+\epsilon}$$

$$\leq\frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}E|v_0|^{2+\epsilon}\sum_{t=2}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2}\left|\frac{1}{N}\sum_{i=1}^{N}|\gamma_i|\right|^{\epsilon}$$

$$\leq\frac{(N/s_N)^{1+\epsilon/2}}{\min_i w_{i,T}^{1+\epsilon/2}}E|v_0|^{2+\epsilon}\left(\frac{1}{N}\sum_{i=1}^{N}|\gamma_i|^{\epsilon}\right)\sum_{t=2}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^{2};$$

by (7.22), $\sum_{t=2}^{T}E\left|\frac{1}{N}\sum_{i=1}^{N}\gamma_i\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right|^2=O\left[r_N'/N^2\left(\max_i w_{i,T}\right)\right]$; putting all together and using Assumptions 1*(iv)*(b) and 5, it holds that

$$(7.25)\qquad\qquad\sum_{t=2}^{T}E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1})=O\left(\frac{\max_i w_{i,T}}{\min_i w_{i,T}}\min_i w_{i,T}^{-\epsilon/2}\right).$$

Consider the following intermediate result, which serves the purpose of determining the order or magnitude of $w_{i,T}$ - note that $\kappa\in\left(0,\frac{1}{2}\right)$

$$E\left(\frac{y_{i,t}}{1+y_{i,t}^2}\right)^2=\int_{-\infty}^{+\infty}\left(\frac{x}{1+x^2}\right)^2dP\left[y_{i,t}\leq x\right]\geq\int_{t^\kappa}^{t^{1/2}}\left(\frac{x}{1+x^2}\right)^2dP\left[y_{i,t}\leq x\right]$$

$$=\int_{t^\kappa}^{t^{1/2}}\left(\frac{x}{1+x^2}\right)^2\sqrt{\frac{2}{\pi t}}e^{-\frac{2}{t}x^2}dx+E_{i,t},$$

based on similar passages as in the proof of Lemma 7.4. Now

$$\int_{t^\kappa}^{t^{1/2}}\left(\frac{x}{1+x^2}\right)^2\sqrt{\frac{2}{\pi t}}e^{-2x^2/t}dx\geq\frac{t}{(1+t)^2}\int_{t^\kappa}^{t^{1/2}}\sqrt{\frac{2}{\pi t}}e^{-2x^2/t}dx\geq C\frac{t}{(1+t)^2},$$

for some $C>0$ when $t\geq2$. Also, the approximation error $E_{i,t}$ is bounded by (see Lemma 7.4*(ii)*) $E_{i,t}$ $\leq 2^{\nu'}m_i't c_0\int_{t^\kappa}^{t^{1/2}}\left(\frac{x}{1+x^2}\right)^2dx^{-\nu'}\leq C\frac{t^{2\kappa+1}}{(1+t^{2\kappa})^2}\int_{t^\kappa}^{t^{1/2}}dx^{-\nu'}$; hence, $\sum_{t=2}^{T}E_{i,t}<\infty$, so that it follows that $\sum_{t=2}^{T}w_{i,T}\geq C\ln T$ for all i. Therefore, by (7.25), $\sum_{t=2}^{T}E(|x_t|^{2+\epsilon}|\mathcal{H}_{t-1})=o_P(1)$, and Theorem 7.1 can be applied. Putting all together, part *(ii)* follows.

We now turn to part *(iii)*. As a preliminary result, applying Lemma 7.4*(iv)* yields that $\sum_{t=2}^{T}E\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2$ $=O(1)$; therefore, the same passages as above therefore yield $\hat{\varphi}_{N,T}-\varphi=O_P\left[\frac{1}{T}\min\left\{\frac{1}{\sqrt{N}},\sqrt{\frac{r_N'}{N^2}}\right\}\right]$. As far as the limiting distribution is concerned, when $r_N'/N=o(1)$ the asymptotics is driven by $N^{-1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}\frac{y_{i,t-1}}{1+y_{i,t-1}^2}=N^{-1/2}\sum_{i=1}^{N}Y_{i,T}$. Conditionally on $\{v_t\}_{t=1}^{T}$, the sequence $Y_{i,T}$ has mean

zero, it is independent across i and, for some $\epsilon > 0$, $\sum_{i=1}^{N} E \left| N^{-1/2} Y_{i,T} \right| \{v_t\}_{t=1}^{T} \Big|^{2+\epsilon} \leq \sum_{i=1}^{N} N^{-1-\frac{1}{2}\epsilon} E \left| e_{i,0} \right|^{2+\epsilon}$

$\leq C N^{-\frac{1}{2}\epsilon}$, which can be shown using the same logic as in the proof of (7.19), and Lemma 7.4(iv). Thus, a conditional version of Theorem 2 of Phillips and Moon (1999, p. 1070) can be applied, yielding

$N^{-1/2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \xrightarrow{\mathcal{D}} N(0, a_3)$ with $\lim_{N,T\to\infty} N^{-1} \sum_{i=1}^{N} \tau_i^2 \sum_{t=2}^{T} \left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2} \right)^2 = a_3$. Combining this with the denominator yields (3.8). $\qquad\square$

Proof of Theorem 3.4 We start by considering the case $E b_{i,t}^2 > 0$. Let

$$(7.26) \qquad c_{N,T} = \begin{cases} T^{1/2}, & \text{under the conditions of part } (i) \\ (TN^2/r_N)^{1/2}, & \text{under the conditions of part } (ii) \\ (NT)^{1/2}, & \text{under the conditions of part } (iii). \end{cases}$$

Theorem 3.2 yields $c_{N,T}(\hat{\varphi}_{N,T} - \varphi) = O_P(1)$. Using (7.26) we get

$$z_{i,t} z_{j,t} = \left[(\varphi - \hat{\varphi}_{N,T}) y_{i,t-1} + b_{i,t} y_{i,t-1} + e_{i,t} + \gamma_i v_t \right] \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \times$$

$$\left[(\varphi - \hat{\varphi}_{N,T}) y_{j,t-1} + b_{j,t} y_{j,t-1} + e_{j,t} + \gamma_j v_t \right] \frac{y_{j,t-1}}{1+y_{j,t-1}^2}$$

and therefore

$$\begin{aligned}
U_{N,T} &= (\varphi - \hat{\varphi}_{N,T})^2 \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}^2}{1+y_{j,t-1}^2} + 2(\varphi - \hat{\varphi}_{N,T}) \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{j,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}^2}{1+y_{j,t-1}^2} \\
&+ 2(\varphi - \hat{\varphi}_{N,T}) \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{j,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} + 2(\varphi - \hat{\varphi}_{N,T}) \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_j v_t \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} \\
&+ \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{i,t} b_{j,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}^2}{1+y_{j,t-1}^2} + 2 \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{i,t} e_{j,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} \\
&+ 2 \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{i,t} \gamma_j v_t \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} + \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{i,t} e_{j,t} \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} \\
&+ 2 \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{i,t} \gamma_j v_t \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} + \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} v_t^2 \gamma_i \gamma_j \frac{y_{i,t-1}}{1+y_{i,t-1}^2} \frac{y_{j,t-1}}{1+y_{j,t-1}^2} \\
&= D_{N,T}(1) + \ldots + D_{N,T}(10).
\end{aligned}$$

By (7.26), we have that $(NT)^{-2} c_{N,T}^2 |D_{N,T}(1)| = O_P(T^{-1})$, and

$$\frac{c_{N,T}^2}{N^2 T^2} |D_{N,T}(2)| = O_P(1) \frac{c_{N,T}}{N^2 T^2} \left| \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{j,t} \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2} \frac{y_{j,t-1}^2}{1+y_{j,t-1}^2} \right|.$$

We have

$$E\left(\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{j,t}\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}^2}{1+y_{j,t-1}^2}\right)^2$$

$$=\sum_{t,s=2}^{T}\sum_{i,k=1}^{N}\sum_{j,\ell=1}^{N}Eb_{j,t}b_{\ell,s}\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}^2}{1+y_{j,t-1}^2}\frac{y_{k,s-1}^2}{1+y_{k,s-1}^2}\frac{y_{\ell,s-1}^2}{1+y_{\ell,s-1}^2}$$

$$=\sum_{t=2}^{T}\sum_{i,k=1}^{N}\sum_{j=1}^{N}\sigma_j^2 E\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}^2}{1+y_{j,t-1}^2}\frac{y_{k,t-1}^2}{1+y_{k,t-1}^2}\frac{y_{j,t-1}^2}{1+y_{j,t-1}^2}=O(N^3T).$$

Hence, by Chebyshev's inequality, $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(2)|=O_P(1)N^{3/2}T^{1/2}\,(NT)^{-2}\,c_{N,T}=o_P(1)$. Similarly, $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(3)|=o_P(1)$. Using again (7.26) we conclude

$$\frac{c_{N,T}^2}{N^2T^2}|D_{N,T}(4)|=O_P(1)\frac{c_{N,T}}{N^2T^2}\left|\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_j v_t\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right|,$$

and by Assumptions 1(i)-(iii) we have

$$E\left(\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_j v_t\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right)^2$$

$$=\sum_{t,s=2}^{T}\sum_{i,k=1}^{N}\sum_{j,\ell=1}^{N}E\left[\gamma_j\gamma_\ell v_t v_s\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{k,s-1}^2}{1+y_{k,s-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{\ell,s-1}}{1+y_{\ell,s-1}^2}\right]$$

$$=\sum_{t=2}^{T}\sum_{i,k=1}^{N}\sum_{j,\ell=1}^{N}E\left[\gamma_j\gamma_\ell\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{k,t-1}^2}{1+y_{k,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{\ell,t-1}}{1+y_{\ell,t-1}^2}\right]$$

$$\leq N^2\sum_{t=2}^{T}\sum_{j,\ell=1}^{N}E\left|\gamma_j\gamma_\ell\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{\ell,t-1}}{1+y_{\ell,t-1}^2}\right|.$$

Following the arguments in the proof of Theorem 3.2(i) we get

$$\sum_{t=2}^{T}\sum_{j,\ell=1}^{N}E\left|\gamma_j\gamma_\ell\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{\ell,t-1}}{1+y_{\ell,t-1}^2}\right|=\sum_{t=2}^{T}\sum_{j\in C(1)}\sum_{\ell\in C(1)}E\left|\gamma_j\gamma_\ell\frac{\bar{y}_{j,t-1}}{1+\bar{y}_{j,t-1}^2}\frac{\bar{y}_{\ell,t-1}}{1+\bar{y}_{\ell,t-1}^2}\right|(1+o(1))$$

which is $O(Tr_N)$ on account of (3.11). Hence we conclude $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(4)|=O_P(1)\,(NT)^{-2}\,c_{N,T}^2$ $NT^{1/2}r_N^{1/2}=o_P(1)$. By independence and Assumption 1(ii)(b), we have that $ED_{N,T}(6)=0$ and $ED_{N,T}^2(6)=4\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^2\tau_j^2 E\left(\frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right)^2=O(N^2T)$, and therefore by Chebyshev's inequality $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(6)|=O_P(1)\,(NT)^{-2}\,c_{N,T}^2\,NT^{1/2}=o_P(1)$. Repeating the arguments above we also get $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(7)|=o_P(1)$ and $(NT)^{-2}\,c_{N,T}^2\,|D_{N,T}(9)|=o_P(1)$. This means that the leading terms are $D_{N,T}(5)$, $D_{N,T}(8)$ and $D_{N,T}(10)$. Repeating the arguments used in the proof

of Theorem 3.2 we obtain that

$$
\begin{aligned}
D_{N,T}(5) &= T\left[\sum_{i\in C(1)} \sigma_i^2 E\left(\frac{y_{i,0}^2}{1+y_{i,0}^2}\right)^2 + \sum_{i\in C(2)\cup C(3)} \sigma_i^2\right](1+o_P(1))\\
D_{N,T}(8) &= T(1+o_P(1))\sum_{i\in C(1)} \tau_i^2 E\left(\frac{y_{i,0}}{1+y_{i,0}^2}\right)^2\\
D_{N,T}(10) &= T(1+o_P(1))\sum_{i\in C(1)}\sum_{j\in C(1)} E\left[\gamma_i\gamma_j \frac{\bar{y}_{i,0}}{1+\bar{y}_{i,0}^2}\frac{\bar{y}_{j,0}}{1+\bar{y}_{j,0}^2}\right].
\end{aligned}
$$

Thus we get

$$
\frac{B_{N,T}}{U_{N,T}^{1/2}}\frac{1}{c_{N,T}} \xrightarrow{P}
\begin{cases}
a_0/\sqrt{a_1}, & \text{under the conditions of part } (i)\\[2mm]
a_0, & \text{under the conditions of part } (ii)\\[2mm]
a_0/\sqrt{a_2}, & \text{under the conditions of part } (iii),
\end{cases}
$$

and therefore the result follows from Theorem 3.2.

Consider now the case $b_{i,t}=0$. The only terms that matter are $D_{N,T}(1)$, $D_{N,T}(3)$, $D_{N,T}(4)$, $D_{N,T}(8)$, $D_{N,T}(9)$ and $D_{N,T}(10)$. Define $\tilde{c}_{N,T}$ so that $\tilde{c}_{N,T}(\hat{\varphi}_{N,T}-\varphi)=O_P(1)$; further, let $w_T = \max_{1\le i\le N} w_{i,T}$, where $w_{i,T}$ is defined in (7.24).

From the above, it follows immediately that $(NT)^{-2}\,\tilde{c}_{N,T}^2\,|D_{N,T}(1)| = O_P(T^{-1})$. Also, consider

$$
E\left[\sum_{t=2}^T\left(\sum_{i=1}^N \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\right)\sum_{j=1}^N e_{j,t}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right]^2 = E\left[\sum_{t=2}^T\left(\sum_{i=1}^N \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\right)^2\sum_{j,k=1}^N e_{j,t}e_{k,t}\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{k,t-1}}{1+y_{k,t-1}^2}\right]
$$

$$
\le N^2 E\left[\sum_{t=2}^T\sum_{j=1}^N e_{j,t}^2\left(\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right)^2\right] = O\left(N^3 w_T\right),
$$

so that $(NT)^{-2}\,\tilde{c}_{N,T}^2\,|D_{N,T}(3)| = o_P(1)$. The same passages as in the proof of (7.22) yields

$$
E\left[\sum_{t=2}^T\left(\sum_{i=1}^N \frac{y_{i,t-1}^2}{1+y_{i,t-1}^2}\right)\sum_{j=1}^N \gamma_j v_t\frac{y_{j,t-1}}{1+y_{j,t-1}^2}\right]^2 \le N^2 E\left[\sum_{j,k=1}^N |\gamma_j\gamma_k|\left|\sum_{t=2}^T \frac{y_{j,t-1}}{1+y_{j,t-1}^2}\frac{y_{k,t-1}}{1+y_{k,t-1}^2}\right|\right] = O\left(N^2 r_N' w_T\right),
$$

whence $(NT)^{-2}\,\tilde{c}_{N,T}^2\,|D_{N,T}(4)| = o_P(1)$. The same conclusion (and very similar passages) can be drawn for $D_{N,T}(9)$. Finally, as far as $D_{N,T}(8)$ is concerned, $E\sum_{t=2}^T\left(\sum_{i=1}^N e_{i,t}\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2 = \sum_{i=1}^N \tau_i^2 E\sum_{t=2}^T\left(\frac{y_{i,t-1}}{1+y_{i,t-1}^2}\right)^2 = O(Nw_T)$, which implies that $(NT)^{-2}\,\tilde{c}_{N,T}^2\,|D_{N,T}(8)| = O_P(1)$, thus being the dominating term when $\min\left\{\sqrt{N}, \sqrt{\frac{N^2}{r_N'}}\right\} = \sqrt{N}$. Also, considering $D_{N,T}(10)$, the same passages as before yield $D_{N,T}(10) = O\left(w_T r_N'\right)$, so that $D_{N,T}(10)$ is $O_P(1)$ when $\min\left\{\sqrt{N}, \sqrt{\frac{N^2}{r_N'}}\right\} = \sqrt{\frac{N^2}{r_N'}}$. $\qquad\square$

References

[1] Arellano, M., Bond, S.: Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**(1991), 277–97.

[2] Aue, A., Hormann, S. Horváth, L., Reimherr, M.: Break detection in the covariance structure of multivariate time series models. *Annals of Statistics* **37**(2009), 4046–4087.

[3] Aue, A., Horváth, L., Steinebach, J.: Estimation in random coefficient autoregressive models. *Journal of Time Series Analysis* **27**(2006), 61–76.

[4] Bai, J.: Panel data models with interactive fixed effects. *Econometrica*, **77**(2009), 1229–1279.

[5] Baltagi, B.H., Bresson, G., Griffin, J.M., Pirotte, A.: Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption. *Empirical Economics*, **28**(2003), 795–811.

[6] Baltagi, B.H., Bresson, G., Pirotte, A.: Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators. some empirical evidence from US electricity and natural-gas consumption. *Economics Letters*, **76**(2002), 375–82.

[7] Baltagi, B.H., Griffin, J.M.: Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* **77**(1997), 303–27.

[8] Baltagi, B.H., Griffin, J.M., Xiong, W.: To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand. *The Review of Economics and Statistics* **82**(2000), 117–26.

[9] Banerjee, A., Chevilon, G., Kratz, M.: Detecting and predicting rational asset price bubbles. Mimeo, 2012.

[10] Berkes, I., Horváth, L., Ling, S.: Estimation in nonstationary random coefficient autoregressive models. *Journal of Time Series Analysis* **30**(2009), 395–416.

[11] Breitung, J., Pesaran, M.H.: Unit roots and cointegration in panels. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data* (3rd ed.), Springer-Verlag, 2008.

[12] Chan, N.H., Li, D., Peng, L.: Toward a unified interval estimation of autoregressions. *Econometric Theory* **28**(2012), 705–717.

[13] Charemza, W.W., Deadman, D.F.: Speculative bubbles with stochastic explosive roots: the failure of unit root testing. *Journal of Empirical Finance* **2**(1995), 153–163.

[14] Chudik, A., Pesaran, M.H.: Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics* **188**(2015), 393–420.

[15] Davidson, J.: The Central Limit Theorem for globally nonstationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case. *Econometric Theory* **9**(1993), 402–412.

[16] Feige, E.L., Swamy, P.A.V.P.: A random coefficient model of the demand of liquid assets. *Journal of Money, Credit and Banking* **6**(1974), 241–252.

[17] Granger, C.W.J., Swanson, N.R.: An introduction to stochastic unit root processes. *Journal of Econometrics* **80**(1997), 35–62.

[18] Hill, J.B., Peng, L.: Unified interval estimation for random coefficient autoregressive models. *Journal of Time Series Analysis* **35**(2014), 282–297.

[19] Horváth, L., Hušková, M.: Change–point detection in panel data. *Journal of Time Series Analysis* **33**(2012), 631–648.

[20] Horváth, L., Trapani, L. : Supplement to "Statistical inference in a random coefficient panel model". Mimeo, 2016.

[21] Hsiao, C.: Some estimation methods for a random coefficient model. *Econometrica* **43**(1975), 305–325.

[22] Hsiao, C., Pesaran, M.H.: Random Coefficient Panel Data Models. Cambridge Working Papers in Economics 0434, Faculty of Economics, University of Cambridge, 2004.

[23] Janečková, H., Prášková, Z.: CWLS and ML estimates in a heteroscedastic RCA(1) model. *Statistics & Decisions*, **22**(2004), 245–259.

[24] Juselius, K., Mladenovic, Z.: High inflation, hyperinflation and explosive roots: the case of Yugoslavia. Discussion Papers 02-23, University of Copenhagen. Department of Economics (2002).

[25] Koul, H. L., Schick, A.: Adaptive estimation in a random coefficient autoregressive model. *Annals of Statistics*, **24**(1996), 1025–1052.

[26] Moon, H.R, Weidner, M.: Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, **83**(2015), 1543–1579.

[27] Ng, S.: A simple test for nonstationarity in mixed panels. *Journal of Business and Economic Statistics* **26**(2008), 113–126.

[28] Nicholls, D.F., Quinn, B.G.: *Random Coefficient Autoregressive Models: An Introduction*. Springer–Verlag, New York, 1983.

[29] Onatski, A.: Asymptotics of the Principal Components estimator of large factor models with weakly influential factors. *Journal of Econometrics* **168**(2012), 244–258.

[30] Pesaran, M.H.: Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**(2006), 967–1012.

[31] Petrov, V.V.: *Limit Theorems of Probability Theory*, Clarendon Press, Oxford, U.K., 1995.

[32] Phillips, P.C.B.: Time series regression with a unit root. *Econometrica*, **55**(1987), 277–301.

[33] Phillips, P.C.B., Moon, H. R.: Linear regression limit theory for nonstationary panel data. *Econometrica* **67**(1999), 1057–1112.

[34] Phillips, P.C.B., Y. Wu, Yu, J.: Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? *International Economic Review* **52**(2011), 201–26.

[35] Schick, A.: $\sqrt{N}$-consistent estimation in a random coefficient autoregressive model. *Australian Journal of Statistics*, **38**(1996), 155–160.

[36] Song, M.: Asymptotic theory for dynamic heterogeneous panels with cross sectional dependence and its applications. Mimeo, 2013.

[37] Swamy, P.A.V.B.: Efficient inference in a random coefficient regression model. *Econometrica* **38**(1970), 311–323.

[38] Wang, X., Yu, J.: Limit theory for an explosive autoregressive process. *Economics Letters* **126**(2015), 176–180.

[39] Westerlund, J., Larsson, R.: Testing for a unit root in a random coefficient panel data model. *Journal of Econometrics* **167**(2012), 254–273.

φ	n	10			20			40			80		
	T	25	50	100	25	50	100	25	50	100	25	50	100
-1.5	$bias$	1.469	-2.889	0.848	0.151	-0.913	1.039	-1.587	-0.438	0.000	1.603	-0.439	-0.258
	MSE	17.59	8.391	4.418	9.497	4.099	2.124	4.508	2.135	1.043	2.291	1.052	0.534
	$size$	0.063	0.049	0.060	0.075	0.056	0.058	0.064	0.052	0.048	0.068	0.058	0.057
-1	$bias$	3.006	-1.716	1.532	1.381	-0.824	1.418	-0.981	-0.235	0.187	2.057	-0.128	-0.149
	MSE	19.28	9.030	4.575	10.09	4.290	2.226	4.835	2.247	1.106	2.482	1.137	0.561
	$size$	0.064	0.054	0.060	0.072	0.051	0.058	0.069	0.051	0.053	0.066	0.052	0.059
-0.5	$bias$	4.247	-0.432	0.765	1.249	-0.302	2.069	0.226	0.804	0.688	2.124	-0.476	-0.292
	MSE	22.85	10.88	4.977	11.25	5.110	2.535	5.663	2.702	1.307	2.933	1.358	0.658
	$size$	0.065	0.063	0.056	0.064	0.052	0.055	0.061	0.058	0.053	0.068	0.061	0.061
0	$bias$	1.162	-4.178	1.203	-0.281	-0.926	1.066	-0.223	0.127	0.369	2.032	0.000	-0.580
	MSE	25.85	12.18	5.436	13.28	5.830	2.945	6.386	3.025	1.508	3.209	1.607	0.737
	$size$	0.064	0.058	0.051	0.067	0.047	0.048	0.058	0.057	0.050	0.065	0.059	0.047
0.5	$bias$	-0.687	-4.337	0.558	-2.176	-2.070	0.634	-0.242	-1.055	0.236	2.352	-0.766	-0.473
	MSE	22.89	10.40	4.918	11.59	5.016	2.617	5.624	2.591	1.280	2.867	1.409	0.654
	$size$	0.069	0.055	0.052	0.072	0.054	0.058	0.060	0.051	0.053	0.065	0.062	0.055
1	$bias$	-2.842	-4.58	0.662	-1.759	-1.856	0.540	-0.275	-0.706	0.177	2.071	-0.184	-0.246
	MSE	19.24	9.153	4.495	10.06	4.338	2.196	4.964	2.274	1.088	2.444	1.153	0.555
	$size$	0.070	0.050	0.054	0.073	0.054	0.057	0.065	0.050	0.050	0.067	0.061	0.060
1.5	$bias$	-1.274	-3.618	0.838	-1.467	-1.485	0.678	-0.217	-0.649	0.157	1.820	-0.329	-0.322
	MSE	17.81	8.570	4.381	9.339	4.093	2.105	4.451	2.132	1.038	2.278	1.068	0.537
	$size$	0.066	0.052	0.061	0.076	0.057	0.054	0.060	0.053	0.047	0.073	0.059	0.058

Table 1. Simulation results. In each entry of the table, the figures represent, respectively: the bias of $\hat{\varphi}$ multiplied by 10^3, calculated using (5.1); the Mean Squared Error associated with $\hat{\varphi}$ multiplied by 10^3, calculated using (5.2); and the empirical rejection frequency for the null that φ is equal to its true value. Note that the number of replications is 2000, which entails that the empirical rejection frequencies have a 95% confidence interval of $[0.04, 0.06]$. As far as the design of the simulation is concerned, it is based on the specifications described in Section 5, setting $\gamma = 0$ - i.e. no cross sectional dependence.

φ		n = 10			n = 20			n = 40			n = 80		
	T	25	50	100	25	50	100	25	50	100	25	50	100
−1.5	bias	0.784	−2.416	1.171	0.291	−1.222	1.155	−0.628	−0.364	0.126	1.713	−0.276	−0.295
	MSE	17.22	8.417	4.376	9.257	4.077	2.122	4.383	2.109	1.057	2.325	1.068	0.541
	size	0.060	0.050	0.053	0.073	0.057	0.056	0.056	0.048	0.050	0.060	0.057	0.060
−1	bias	2.460	−1.387	0.924	0.126	−1.277	1.305	−0.156	−0.473	0.199	1.608	−0.210	−0.300
	MSE	18.23	8.852	4.473	10.04	4.156	2.191	4.759	2.182	1.089	2.541	1.163	0.556
	size	0.064	0.054	0.054	0.071	0.050	0.055	0.055	0.050	0.055	0.060	0.060	0.058
−0.5	bias	5.019	−0.588	0.765	0.628	−0.863	1.603	−0.520	0.756	0.539	1.188	−0.393	0.000
	MSE	20.87	9.980	4.977	10.99	4.695	2.426	5.611	2.658	1.249	3.198	1.412	0.663
	size	0.068	0.061	0.056	0.066	0.056	0.056	0.059	0.057	0.050	0.062	0.053	0.062
0	bias	2.758	−4.657	1.203	0.000	−3.622	1.376	−0.206	0.224	0.566	0.969	−0.557	0.423
	MSE	23.05	11.17	5.436	12.22	5.372	2.819	6.829	3.179	1.568	4.135	1.890	0.902
	size	0.063	0.062	0.051	0.063	0.050	0.054	0.063	0.060	0.056	0.066	0.055	0.055
0.5	bias	0.205	−4.086	0.558	−1.840	−3.105	1.902	−0.696	−0.335	0.364	1.501	−0.569	0.000
	MSE	20.92	9.788	4.918	11.09	4.798	2.438	5.717	2.604	1.254	3.246	1.459	0.677
	size	0.068	0.058	0.052	0.062	0.054	0.054	0.064	0.055	0.052	0.064	0.054	0.060
1	bias	−2.918	−3.163	0.662	0.755	−1.848	0.412	−0.171	−0.293	0.350	1.080	0.000	−0.361
	MSE	18.29	8.812	4.495	9.830	4.264	2.161	4.828	2.191	1.071	2.540	1.144	0.549
	size	0.066	0.057	0.054	0.068	0.062	0.053	0.059	0.054	0.055	0.061	0.055	0.057
1.5	bias	−1.576	−3.474	0.838	0.000	−1.801	0.870	−0.202	−0.540	0.000	1.222	−0.261	−0.367
	MSE	17.45	8.449	4.382	9.474	4.042	2.109	4.355	2.092	1.044	2.291	1.068	0.535
	size	0.065	0.050	0.061	0.058	0.063	0.054	0.054	0.045	0.051	0.064	0.057	0.058

Table 2. Simulation results. The figures in the Table have the same meaning as in Table 1; the design of the simulation is the same as in Table 1, save for γ, which here is set to $\gamma = 1$.

φ			n = 10			n = 20			n = 40			n = 80		
		T	25	50	100	25	50	100	25	50	100	25	50	100
-1.5		$bias$	-3.723	-2.699	1.094	2.439	-3.233	0.928	-0.193	-0.306	0.316	0.852	0.448	-0.344
		MSE	26.73	10.77	5.021	18.64	5.917	2.566	11.84	3.617	1.419	8.355	2.545	0.895
		$size$	0.051	0.055	0.061	0.061	0.053	0.055	0.047	0.041	0.050	0.051	0.050	0.053
-1		$bias$	-1.744	-2.736	1.199	1.597	-3.874	0.444	-0.793	-0.979	0.642	0.285	0.016	0.000
		MSE	30.52	11.99	5.480	21.19	6.649	2.799	13.85	4.172	1.580	9.953	2.979	1.019
		$size$	0.056	0.056	0.062	0.062	0.048	0.055	0.053	0.048	0.047	0.050	0.053	0.055
-0.5		$bias$	2.721	-0.529	2.528	0.552	-2.704	0.257	-0.127	-0.393	0.149	0.000	0.546	0.633
		MSE	37.58	15.68	6.716	25.53	8.486	3.720	16.88	4.859	2.194	13.27	4.251	1.526
		$size$	0.059	0.069	0.062	0.052	0.053	0.059	0.053	0.047	0.049	0.046	0.054	0.063
0		$bias$	1.626	-1.490	2.189	1.047	-3.614	-1.166	-0.778	-0.288	0.177	-0.457	0.771	0.520
		MSE	46.56	18.83	8.659	31.52	11.24	5.124	21.50	7.052	3.556	18.20	6.744	2.756
		$size$	0.065	0.055	0.054	0.057	0.046	0.052	0.053	0.050	0.052	0.056	0.051	0.050
0.5		$bias$	-2.394	-0.719	1.300	2.578	-5.186	0.590	0.301	0.171	0.527	-1.252	0.556	0.717
		MSE	38.12	15.04	6.535	25.67	8.198	3.536	16.71	5.530	2.187	13.11	4.408	1.508
		$size$	0.054	0.053	0.054	0.061	0.044	0.049	0.051	0.047	0.054	0.054	0.051	0.050
1		$bias$	-3.173	-2.164	0.493	3.147	-3.962	0.922	-1.370	-0.345	0.418	-0.293	0.222	0.000
		MSE	32.38	12.47	5.264	19.91	6.597	2.729	13.53	4.180	1.557	10.01	3.001	1.016
		$size$	0.063	0.056	0.058	0.054	0.050	0.056	0.045	0.047	0.046	0.054	0.052	0.055
1.5		$bias$	-1.143	-2.146	0.666	2.227	-3.587	0.752	-1.891	0.739	0.344	0.824	0.809	0.000
		MSE	26.83	11.09	4.999	18.15	5.851	2.531	11.52	3.580	1.442	8.485	2.505	0.897
		$size$	0.051	0.058	0.057	0.058	0.043	0.053	0.046	0.045	0.051	0.051	0.048	0.055

Table 3. Simulation results. The figures in the Table have the same meaning as in Table 1; the design of the simulation is the same as in Table 1, save for γ, which here is set to $\gamma = 10$.

n		20				40				80		
T		50		100		50		100		50		100	
γ		0	10	0	10	0	10	0	10	0	10	0	10
-1.5	$bias$	0.050	0.828	0.028	0.491	0.106	0.122	0.001	0.442	0.045	1.190	0.016	0.075
	MSE	0.055	2.138	0.013	0.521	0.028	1.938	0.006	0.431	0.014	1.665	0.003	0.388
	$size$	0.040	0.020	0.041	0.021	0.039	0.016	0.046	0.011	0.042	0.029	0.043	0.016
-1	$bias$	0.082	0.774	0.257	1.371	0.503	1.630	0.003	0.203	0.228	0.425	0.025	0.011
	MSE	0.365	8.427	0.124	2.619	0.185	5.823	0.061	0.085	0.089	1.979	0.029	0.694
	$size$	0.052	0.055	0.055	0.060	0.059	0.062	0.058	0.051	0.051	0.057	0.043	0.060
-0.5	$bias$	1.010	0.715	1.054	-3.517	0.245	-4.003	-0.340	1.484	-0.020	-0.914	0.106	4.059
	MSE	1.634	53.822	0.781	25.776	0.827	54.713	0.389	26.441	0.413	47.773	0.199	25.475
	$size$	0.058	0.055	0.048	0.044	0.058	0.062	0.047	0.050	0.054	0.048	0.054	0.052
0	$bias$	1.382	3.718	0.615	-6.979	-0.468	1.081	-0.516	3.218	-0.227	-1.777	-0.056	0.443
	MSE	1.897	66.058	0.945	34.669	0.963	66.611	0.481	32.352	0.511	65.070	0.249	32.154
	$size$	0.054	0.051	0.050	0.050	0.051	0.056	0.049	0.048	0.060	0.056	0.061	0.050
0.5	$bias$	0.637	-4.324	0.601	-10.025	0.228	-2.678	0.354	5.355	0.036	1.460	-0.361	2.213
	MSE	1.568	53.973	0.793	26.786	0.758	51.262	0.391	26.056	0.427	51.676	0.203	25.710
	$size$	0.051	0.054	0.058	0.051	0.053	0.047	0.049	0.052	0.063	0.064	0.059	0.057
1	$bias$	-0.688	-2.867	-0.119	-0.351	-0.449	-2.567	-0.085	-0.351	-0.202	-1.704	-0.066	0.288
	MSE	0.359	8.311	0.118	2.595	0.175	4.556	0.061	2.036	0.086	1.959	0.031	0.719
	$size$	0.054	0.054	0.044	0.061	0.055	0.054	0.052	0.058	0.050	0.057	0.057	0.058
1.5	$bias$	-0.178	-2.723	0.046	0.346	0.037	0.409	0.065	0.573	0.074	0.049	0.042	0.291
	MSE	0.056	2.096	0.013	0.500	0.027	1.883	0.065	0.421	0.013	1.641	0.003	0.385
	$size$	0.039	0.016	0.042	0.018	0.046	0.021	0.046	0.018	0.039	0.016	0.044	0.011

(The row index on the left is φ.)

Table 4. Simulation results for the case $b_{i,t} = 0$. The figures in the Table have the same meaning as in Table 1, and are obtained under the same design.

	GDP	M2	Equity Index	Short term rate	IP	UK House Price Index
(N, T)	$(15, 212)$	$(18, 55)$	$(18, 75)$	$(16, 80)$	$(18, 90)$	$(13, 45)$
$\tilde{\varphi}$	0.9988	1.002	1.000	0.9605	1.0007	1.0136
$\sqrt{\widehat{Var(\widehat{\varphi} - \varphi)}}$	0.00098	0.00016	0.00177	0.01841	0.00032	0.00081
95% confidence interval	$[0.9968, 1.0007]$	$[1.0016, 1.0023]$	$[0.9965, 1.0034]$	$[0.9244, 0.9965]$	$[1.0000, 1.0013]$	$[1.0090, 1.0121]$
Test for $H_0 : \varphi \leq 1$						
t-stat	-1.224	12.5	0	-2.145	2.187	13.086
p-value	$[0.889]$	$[0.000]$	$[0.500]$	$[0.984]$	$[0.014]$	$[0.000]$

Table 5. Empirical applications. The first panel of the table (columns headings: GDP, M2, Equity Index, Short Term rate and IP) contains quarterly EU macroeconomic and financial data, as described in this section. The last column contains the Nationwide house price index. The last two rows contain a t-test for the null of no explosive root - the test statistic is defined as $(\tilde{\varphi} - 1) / \sqrt{\widehat{Var(\widehat{\varphi} - \varphi)}}$.

Lajos Horváth, Department of Mathematics, University of Utah, Salt Lake City, UT 84112–0090 USA

Lorenzo Trapani, Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, UK.