Journal of Econometrics 179 (2014) 46–65

Contents lists available at ScienceDirect

Journal of Econometrics

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

Inference on stochastic time-varying coefficient models

L. Giraitis [a,*], G. Kapetanios [a], T. Yates [b]

[a] Queen Mary, University of London, United Kingdom
[b] University of Bristol, United Kingdom

ARTICLE INFO

Article history:
Received 19 September 2011
Received in revised form
25 March 2013
Accepted 11 October 2013
Available online 28 October 2013

JEL classification:
C10
C14

Keywords:
Time-varying coefficient models
Random coefficient models
Nonparametric estimation
Kernel estimation
Autoregressive processes

ABSTRACT

Recently, there has been considerable work on stochastic time-varying coefficient models as vehicles for modelling structural change in the macroeconomy with a focus on the estimation of the unobserved paths of random coefficient processes. The dominant estimation methods, in this context, are based on various filters, such as the Kalman filter, that are applicable when the models are cast in state space representations. This paper introduces a new class of autoregressive bounded processes that decompose a time series into a persistent random attractor, a time varying autoregressive component, and martingale difference errors. The paper examines, rigorously, alternative kernel based, nonparametric estimation approaches for such models and derives their basic properties. These estimators have long been studied in the context of deterministic structural change, but their use in the presence of stochastic time variation is novel. The proposed inference methods have desirable properties such as consistency and asymptotic normality and allow a tractable studentization. In extensive Monte Carlo and empirical studies, we find that the methods exhibit very good small sample properties and can shed light on important empirical issues such as the evolution of inflation persistence and the purchasing power parity (PPP) hypothesis.

1. Introduction

This paper introduces a new class of random time varying coefficient (RC) models for bounded non-stationary processes with AR(1) dynamics and proposes kernel-based nonparametric methods for inference on the paths of the unobserved drifting coefficient processes. RC models have been widely discussed in the last few years in applied macroeconomic time series analysis. Work has ranged across topics such as accounting for the Great Moderation, documenting changes in the effect of monetary policy shocks and in the degree of exchange rate pass-through, see e.g. Cogley and Sargent (2001, 2005b), Cogley et al. (2010), Benati (2010), Pesaran et al. (2006), Stock and Watson (1998) and Koop and Potter (2008). It is clear that RC models provide a de facto benchmark technology for analysing structural change. The breadth of the previous work means that the results of this paper have many applications. While kernel based methods form the main approach for estimating models, whose parameters change smoothly and deterministically over time, they have never been considered in the literature as potential methods for inference on RC models, which have been estimated in the context of state space model representations. While the theoretical asymptotic properties of estimating such processes via the Kalman, or related filters are unclear, we show that under very mild conditions, kernel-based estimates of random coefficient processes have very desirable properties such as consistency and asymptotic normality.

The crucial conditions that need to be satisfied to obtain our theoretical results are also commonly imposed for RC models used in applied macroeconomic analysis. These are pronounced persistence of the coefficient process (usually a random walk assumption) coupled with a restriction that the process remains bounded. We formalize these conditions, in a direct intuitive way, while noting that a variety of alternative bounding devices can be used.

The crucial issue of the choice of bandwidth that is perennially present in kernel based estimation is also addressed. We find that a simple choice of bandwidth has wide applicability and can be used irrespective of many aspects of the true nature of the coefficient processes. The latter may have both a deterministic and a stochastic time varying component thus generalizing the two existing polar paradigms. We find that kernel estimation can cope effectively with such a general model and that the choice of bandwidth can be made robust to this possibility.

Although we focus on a simple autoregressive form for the model as a vehicle to investigate our estimator of the unobserved drifting coefficient process, our results are relevant much more widely. They apply to general regression models, multivariate VAR-type models and can be extended to models that allow for

* Corresponding author. Tel.: +44 0 20 7882 8826.
E-mail address: l.giraitis@qmul.ac.uk (L. Giraitis).

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.10.009

time-varying stochastic volatility which are used widely in applied macroeconometrics.

The theoretical analysis in this paper is coupled with a extensive Monte Carlo study that addresses a number of issues arising out of our theoretical investigations. In particular, it confirms the desirable properties of the proposed estimators, identified in our theoretical analysis. For example, the theoretically optimal choice of bandwidth is also one of the best in small samples. We illustrate the usefulness of RC modelling in two applications that have received attention in previous work. The first documents changes in inflation persistence over time. The second analyses whether changes in the persistence of deviations from purchasing power parity (PPP) have occurred or not.

The rest of the paper is structured as follows. Section 1.1 discusses the existing literature and provides a framework for our contribution. Section 2 presents the model and some of its basic properties that are of use for theoretical developments. It also contains the main theoretical results on the asymptotic properties of the new estimator. Section 3 provides an extensive Monte Carlo study while Section 4 discusses the empirical application of the new inference methods to CPI inflation and real exchange rate data. Finally, Section 5 concludes. The proofs of all results are relegated to an Appendix.

1.1. Background literature

The investigation of structural change in applied econometric models has been receiving increasing attention in the literature over the past couple of decades. This development is not surprising. Assuming wrongly that the structure of a model remains fixed over time, has clear adverse implications. The first implication is inconsistency of the parameter estimates. A related implication is the fact that structural change is likely to be responsible for most major forecast failures of time invariant models.

As a result a large literature on modelling structural change has appeared. Most of the work assumes that structural changes in parametric models occur rarely and are abrupt. A number of tests for the presence of structural change of that form exist in the literature starting with the ground-breaking work of Chow (1960) who assumed knowledge of the point in time at which the structural change occurred. Other tests relax this assumption. Examples include Brown et al. (1974), Ploberger and Kramer (1992) and many others. In this context it is worth noting that little is being said about the cause of structural breaks in either statistical or economic terms. The work by Kapetanios and Tzavalis (2010) provides a possible avenue for modelling structural breaks and, thus, addresses partially this issue.

A more recent strand of the literature takes an alternative approach and allows the coefficients of parametric models to evolve randomly over time. To achieve this the parameters are assumed to be persistent stochastic processes giving rise to RC models. An early and influential example is Doan et al. (1984) who estimate an RC model on macroeconomic time series and emphasize the utility of Bayesian methods as a way to encode – amongst other things – theoretically informed views that explosive models for data ought to have very low or zero probability. Cogley and Sargent (2005b) deploy an RC model to address the question of whether it was changes in the variance of shocks, or changes in coefficients – policy or otherwise – that gave rise to the period of macroeconomic calmness in the 90s and early 2000s, dubbed the 'Great Moderation'. In this work, and research influenced by it, the authors assume a random walk process for the coefficients of the VAR model, but bound them so that at each point in time the VAR is non-explosive. For the univariate models this amounts to bounding the coefficients between -1 and $+1$. This assumption is justified on the grounds that the monetary authorities would act somehow to ensure that inflation was not explosive. A main point of Cogley and Sargent (2005b) was to respond to criticisms of earlier work (Cogley and Sargent, 2001) that had found evidence of changes in coefficients but without allowing for changes in volatilities, thus potentially biasing their findings in favour of documenting structural change in VAR coefficients. They find evidence of change in the coefficients of the inflation process despite the inclusion of time-varying volatilities. In subsequent work, Cogley et al. (2010) used the same model to investigate whether there had been significant changes in the persistence of inflation (more precisely the gap between inflation and its time varying unobserved permanent component) during the Great Moderation, using the same RC tool. Other examples of the use of this RC tool abound. Benati and Surico (2008) estimate a similar VAR model for inflation and use it to infer that the decline in the persistence of inflation is related to an increased responsiveness of interest rates to deviations of inflation from its target. Mumtaz and Surico (2009) estimate an RC model to characterize evolutions in the term structure and the correspondence of changes therein with the monetary regime. Benigno et al. (2010) estimate a VAR with random walks in the propagation coefficients involving productivity growth, real wage growth and the unemployment rate and find that increases in the variance of productivity growth have a long run effect on the level of unemployment. Researchers have also debated some of the difficulties with the approach. For example, Stock and Watson (1998) discuss how maximum likelihood implementations tend to overstate the probability that the variance of the shock to coefficients is low or zero; Koop and Potter (2008) discuss the difficulty in imposing inequality restrictions on the time-varying autoregressive coefficients, particularly in large dimensional applications and note that it can be hard to find posterior draws that satisfy such conditions.

While the above account gives a clear idea of the current state of the relevant econometric literature, the economic justification of RC models, whose parameters evolve as bounded random walk processes, merits an additional comment. On a practical level, these models are now widely used by empirical macroeconomists. The economic reason for their attractiveness is well explained in the discussion in Cogley and Sargent (2001), who pointed out that fluctuations in parameters of a reduced form economic system may result from evolving beliefs of the policymaker leading to the evolving policy rules. The evolution in beliefs itself is a potential product of the interaction of model misspecification by policymakers and the effects produced by the policy itself, in the economy. Cogley and Sargent (2001) refer to the seminal paper by Lucas (1976), who noted that the practice of macroeconometric modellers of introducing intercept corrections (discussed also in Cooley and Prescott (1973)), may ignore the risk of misspecification due to using a model detached from macroeconomic theory. In Lucas (1976), the author describes how pursuing a policy based on an initial estimation of an atheoretical model would result in time varying reduced form econometric coefficients (not just in intercepts, but more generally). The subsequent work by Sargent (2001) (and later work by Cogley and Sargent (2005a) and Sargent et al. (2006)) was an attempt to describe post war history as the result of a perpetual repetition of the mistakes (or perpetual 'learning') described in the Lucas Critique paper. In summary, the evolution of beliefs and time varying policy in a stochastic economy, may explain why reduced form VAR parameters may evolve themselves as stochastic processes over time. Clearly, the above justification of stochastically varying coefficient models is only one possibility. Equally, time variation may arise because of evolving cultural norms or behaviours, of any agent in the economy, providing a number of other possible avenues for the motivation of RC AR modelling discussed in our work.

A particular issue with the use of such models is the relative difficulty involved in estimating them. As the focus of the analysis is quite often the inference of the time series of the time-varying coefficients, models are usually cast in state space form and estimated using variants of the Kalman filter. More recently, the addition of various new features in these models, such as time-varying variances for the error terms, has meant that the Kalman filter may not be sufficient and a variety of techniques, quite often of a Bayesian flavour, have been used for such inference.

Yet another strand of the vast structural change literature assumes that regression coefficients change but in a smooth deterministic way. Such modelling attempts have a long pedigree in statistics starting with the work of Priestley (1965). Priestley's paper suggested that processes may have time-varying spectral densities which change slowly over time. The context of such modelling is nonparametric and has, more recently, been followed up by Robinson (1989, 1991), Dahlhaus (1997), and others, some of whom refer to such processes as locally stationary processes. We will refer to such parametric models as deterministic time-varying coefficient (DTVC) models. A disadvantage of such an approach is that the change of deterministic coefficients cannot be modelled or, for that matter, forecasted. Both of these are theoretically possible with RC. However, an important assumption underlying DTVC models is that coefficients change slowly. As a result forecasting may be carried out by assuming that the coefficients remain at their end-of-observed-sample value. The above approach while popular in statistics has not really been influential in applied macroeconometric analysis where, as mentioned above, RC models dominate. Kapetanios and Yates (2008) is an exception, using DTVC models to revisit the study of the evolution of inflation persistence, in Cogley and Sargent (2005b). Finally, it is worth noting the work of Muller and Watson (2008) and Muller and Petalas (2010) who also examine structural change and consider both deterministic and stochastic versions for the time-varying parameters.

While both approaches can be used for the same modelling purposes, the underlying models have very distinct properties and have been analyzed in very distinct contexts. As we noted in the introduction, this paper uses the kernel approach to carry out inference on RC models.

2. The model and its basic properties

2.1. The model

In this section we introduce a class of autoregressive models driven by a random drifting autoregressive parameter ρ_t, that evolves as a non-stationary process, standardized to take values in the interval $(-1, 1)$. We also allow for a random drifting intercept term in the model.

Such an autoregressive model aims to replicate patterns of evolution of autoregressive coefficients that are relevant for the modelling of the evolution of macroeconomic variables such as inflation. Such AR models have been extensively discussed in the recent macroeconometric literature, see e.g. Cogley and Sargent (2005b) and Benati (2010). Our objective is to develop a suitable statistical model that allows estimation and inference.

The limit theory for stationary autoregressive models with non-random coefficients is well understood. For AR models with time-invariant coefficients it was developed by Anderson (1959) and Lai and Wei (2010). Phillips (1987), Chan and Wei (1987), Phillips and Magdalinos (2007) and Andrews and Guggenberger (2008) extended it to AR(1) models that are local to unity. A class of a locally stationary processes that includes AR processes with deterministic time-varying coefficients was introduced by Dahlhaus (1997). Estimation of such processes was discussed in Dahlhaus and Giraitis (1998). In this paper, we develop an AR(1) model with *random coefficients*, which encompasses stationary and locally stationary AR(1) models. The simplest case of a random coefficient process is a driftless random walk.

We consider the AR(1) models

$$y_t = \rho_{t-1} y_{t-1} + u_t, \tag{2.1}$$

$$y_t = \alpha_t + \rho_{t-1} y_{t-1} + u_t, \quad t = 1, 2, \ldots, \tag{2.2}$$

with a drifting random coefficient ρ_t, a random intercept α_t and initialization y_0, where $\{u_t\}$ is a stationary ergodic martingale difference sequence (m.d.s.) with respect to some natural filtration $\mathcal{F}_t$ and ρ_t, α_t are $\mathcal{F}_t$ measurable, i.e. $E[u_{t+1}|\mathcal{F}_t] = 0$ and $E[\rho_t|\mathcal{F}_t] = \rho_t$. For example, in (2.1), one can set $\mathcal{F}_t = \sigma\{u_j, \rho_j, \ j \le t\}$.

The literature on locally stationary AR(1) processes assumes that coefficients μ_t and ρ_t are smooth deterministic functions. Then, y_t behaves locally as a stationary process, which has different theoretical properties compared to AR processes with random coefficients. Moreover, the model (2.2) contains an additional parameter of interest, a random persistent attractor, see Section 2.3. Specification of ρ_t requires additional structural assumptions. In applied literature, it is often assumed that ρ_t is a rescaled random walk which is a stringent restriction.

In this paper we assume that ρ_t is given by

$$\rho_t = \rho \frac{a_t}{\max\limits_{0 \le k \le t} |a_k|}, \quad t \ge 0, \tag{2.3}$$

where the stochastic process a_t determines the random drift, and $\rho \in (0, 1)$,[1] restricts ρ_t away from the boundary points -1 and 1. Both a_t and ρ are unknown, and $\rho_t \in [-\rho, \rho] \subset (-1, 1)$.

We split $a_t = \{a_t - Ea_t\} + Ea_t$ into a random part $\{a_t - Ea_t\}$ and the non-random mean Ea_t. We shall assume that ρ_t combines deterministic and random components. The most novel case is $Ea_t = 0$.

To enable inference about ρ_t, we need the following assumptions on ρ_t, y_0, u_t and a_t.

Assumption 2.1. (i) The random coefficients $\rho_t, \ a_t, \ t = 0, \ldots, n$ are $\mathcal{F}_t$ measurable; $Ea_0^4 < \infty, Ey_0^4 < \infty$ and $Eu_1^4 < \infty$.

(ii) The process $v_t := \{a_t - Ea_t\} - \{a_{t-1} - Ea_{t-1}\}, t = 1, \ldots, n$ is stationary with zero mean and finite variance.

Part (ii) implies that a non-stationary process a_t with $Ea_t = 0$ is a partial sum of shocks v_j:

$$a_t = a_0 + v_1 + \cdots + v_t.$$

The popular empirical choice of a_t is a driftless random walk with i.i.d. first differences, v_t, see, e.g., Cogley and Sargent (2005b). In addition, if v_1 has $2 + \delta$ finite moments, then the process $a_{[\tau n]}, \ 0 \le \tau \le 1$ converges weakly in Skorokhod space $D[0, 1]$ to a standard Brownian motion B_τ:

$$n^{-1/2} a_{[\tau n]} \Rightarrow_{D[0,1]} \sigma_v^2 B_\tau, \quad 0 \le \tau \le 1.$$

In this paper, v_t's are allowed to be dependent. The only assumption on a_t is the weak convergence of a renormalized process $a_{[\tau n]}$ to some non-degenerate limit process, which may differ from the standard Brownian motion B_τ, and may be even non-Gaussian.

Assumption 2.2. There exists $\gamma \in (0, 1)$ such that

$$n^{-\gamma} a_{[\tau n]} \Rightarrow_{D[0,1]} W_\tau + g(\tau), \quad 0 \le \tau \le 1, \tag{2.4}$$

[1] The results of this paper remain valid also for negative $\rho \in (-1, 0)$ in (2.3).

where $(W_\tau,\ 0 \le \tau \le 1)$ is zero mean random process with finite variance, W_1 has continuous probability distribution, and $g(\tau)$ is a deterministic continuous bounded function. Moreover,

$$\begin{aligned} &n^{-\gamma}(a_{[\tau n]} - Ea_{[\tau n]}) \Rightarrow_{D[0,1]} W_\tau, \\ &n^{-\gamma} Ea_{[\tau n]} \to g(\tau), \quad 0 \le \tau \le 1, \\ &|Ea_t - Ea_{t+k}| \le Ck^\gamma, \quad 1 \le k < t. \end{aligned} \tag{2.5}$$

Remark 2.1. Assumption 2.2 is satisfied by the sum process $a_t = v_1 + \cdots + v_t$ where

$$v_j = \sum_{k=0}^{\infty} \nu_k \zeta_{j-k}, \quad j \ge 0, \qquad \sum_{k=0}^{\infty} \nu_k^2 < \infty, \tag{2.6}$$

is a linear process with stationary ergodic m.d. innovations ζ_k, $E\zeta_1^2 < \infty$, under a minimal additional condition that

$$\mathrm{Var}(a_n) = \mathrm{Var}\left(\sum_{j=1}^{n} v_j\right) \sim Cn^{2\gamma}, \quad \text{for some } \gamma \in (0,1). \tag{2.7}$$

In Theorem 3.1 of Abadir et al. (forthcoming) it is shown that (2.6) and (2.7) imply the weak convergence $n^{-\gamma} a_{[\tau n]} \Rightarrow_{D[0,1]} cW_\tau$, $(c > 0)$, to a standardized fractional Brownian motion W_τ, as long as $E|\zeta_1|^p < \infty$ for some $p > \max(1/\gamma, 2)$. Assumptions (2.6) and (2.7) are satisfied by stationary ARMA(p,q) processes with $\gamma = 1/2$ and by ARFIMA(p,d,q), $|d| < 1/2$ processes with $\gamma = (1/2) + d$. Stationary seasonal long memory GARMA(p,d,q) processes v_j, whose spectral density has an infinite peak at a frequency $\omega \ne 0$, also satisfy (2.6) and (2.7) with $\gamma = 1/2$, see Section 7.2.2 of Giraitis et al. (2012).

Remark 2.2. The above setup and assumptions are designed to guarantee persistence and boundedness, which are the two main properties of the stochastic coefficient process ρ_t and the intercept α_t. It is worth briefly commenting on how they lead to our results. Boundedness is essential in avoiding explosive behaviour for y_t while persistence is needed to enable estimation of the unobserved ρ_t and α_t by local averaging. These properties are repeatedly employed in the proofs. Their use is particularly apparent in Lemma A.1 which is the main building block of the proofs of our major results in Sections 2.2–2.4.

Under Assumption 2.2, the coefficient process ρ_t, as n increases, behaves as a rescaled limit process W_τ of (2.5):

$$\begin{aligned} &\rho_{[n\tau]} \to_D \rho W_\tau^{(b)}, \quad \forall \tau \in (0,1), \\ &W_\tau^{(b)} := \big(W_\tau + g(\tau)\big) / \sup_{0 \le x \le \tau} |W_x + g(x)|. \end{aligned}$$

In particular, W_τ can be standard or fractional Brownian motion. Then $\rho_{[n\tau]}$ evolves around $\rho g(\tau)$, and can take any value in the interval $[-\rho, \rho]$. Below $\to_D$ and $\to_p$ denote convergence in distribution and probability, respectively, $=_D$ indicates equality of distributions, whereas $[x]$ denotes the integer part of a real number x.

Example 2.1. A typical example of a process a_t, satisfying Assumption 2.2 with the parameter $0 < \gamma < 1$, is

$$a_t = \{v_1 + \cdots + v_t\} + t^\gamma \times \frac{h_1 + \cdots + h_t}{t}, \tag{2.8}$$

where v_j's are stationary zero mean r.v.'s and h_j are non-random numbers such that $\max_j |h_j| < \infty$. It has the stochastic part $z_t := a_t - Ea_t = v_1 + \cdots + v_t$, and the mean $Ea_t = t^\gamma g_t$ where $g_t = (h_1 + \cdots + h_t)/t$, which satisfies $|Ea_t - Ea_{t+k}| = |t^\gamma g_k - (t+k)^\gamma g_{t+k}| \le Ck^\gamma$, for $1 \le k \le t$. Then, a_t satisfies Assumption 2.2, if $n^{-\gamma} z_{[\tau n]} \Rightarrow_{D[0,1]} W_\tau$, $0 \le \tau \le 1$, and $n^{-\gamma} Ea_{[\tau n]} = n^{-\gamma}[\tau n]^\gamma g([\tau n]/n) \to \tau^\gamma g(\tau)$.

Such a_t has asymptotically non-diminishing random and deterministic components. For weakly dependent v_j's, one sets $\gamma = 1/2$ and $a_t = z_t + t^{1/2} g_t$. This setting also allows to generate non-random coefficients ρ_t used in modelling of locally stationary processes.

Parametric random coefficient. Example (2.8) suggests a simple parametric model for an AR(1) random coefficient,

$$\rho_t = \rho \frac{c \sum_{k=1}^{t} u_k + t^{1/2}}{\max_{1 \le j \le t} \left| c \sum_{k=1}^{j} u_k + j^{1/2} \right|}, \quad t \ge 1, \ \rho_0 = \rho, \tag{2.9}$$

driven by the same m.d. noise u_t as in AR model (2.1), with parameters ρ and c. If $c = 0$, then $\rho_t \equiv \rho$; if $c \to \infty$, then $\rho_t = \rho(\sum_{k=1}^{t} u_k)/(\max_{1 \le j \le t} |\sum_{k=1}^{j} u_k|)$ becomes purely random, while for a finite $c > 0$, ρ_t combines random and deterministic patterns.

Remark 2.3. To restrict ρ_t in the interval $[-\rho, \rho]$, we use the normalization $\rho_t = \rho\, a_t / \max_{0 \le k \le t} |a_k|$. The normalization $\rho_t = \rho\, a_t / \max_{0 \le k \le n} |a_k|$ could also be used, and would simplify technical derivations but at the expense of an assumption of independence between ρ_t and u_t. Another popular implicit standardization used in the applied macroeconometric literature is

$$\rho_t = \begin{cases} a_{t-1} + v_t, & \text{if } |a_{t-1} + v_t| \le \rho \\ \rho, & \text{otherwise.} \end{cases}$$

The question of how best to restrict ρ_t is open. Usually in the macroeconometric literature the restriction is based on computational convenience without discussing the properties of the resulting model, and what is the best way to restrict the process ρ_t from an economic point.

Remark 2.4. The paper narrows its interest to the AR(1) time-varying random framework and identifies conditions that allow rigorous inference on it. It shows that kernel estimation and inference extends to coefficients composed of time varying random and deterministic parts. Such a finding is neither intuitively obvious nor has a trivial formal justification. Establishing the AR(1) framework opens the possibility for a general inference theory for AR and VAR models that may possess time-varying variances. To illustrate a flavour of such extensions we briefly outline some frameworks used in macroeconomic applications and ways in which these can be adapted to our setting.

(1) Time varying *AR*(p) model

$$y_t = \sum_{i=1}^{p} \rho_{t-1,i} y_{t-i} + u_t, \quad t \ge 1,$$

can be defined using the bounding condition

$$\rho_{t,i} = \rho \frac{a_{t,i}}{\max_{0 \le k \le t} \sum_{i=1}^{p} |a_{k,i}|}, \quad t \ge 1, \tag{2.10}$$

where $0 < \rho < 1$, and each $\{a_{t,i}\}$, $i = 1, \ldots, p$ are independent versions of the a_t process used above. Under these restrictions the maximum absolute eigenvalue of the matrix

$$A_t = \begin{pmatrix} \rho_{t,1} & \rho_{t,2} & \cdots & \rho_{t,p} \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{2.11}$$

or its spectral norm $\|A_t\|_{sp}$, are bounded above by one, for all t. This requirement on the companion matrix in (2.11) plays a similar role to standardization of a_t in (2.3) and is reminiscent of the standard stationarity condition for fixed coefficient $AR(p)$ models.

(2) Time Varying $VAR(1)$ model is given by

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \quad t \geq 1,$$

where $\boldsymbol{y}_t$ is an m-dimensional vector and $\boldsymbol{\Psi}_{t-1} = [\psi_{t-1,ij},\ i,j = 1,\ldots,m]$. The necessary bounding can be implemented by defining, similarly to (2.10),

$$\psi_{t-1,ij} = \rho \frac{a_{t-1,ij}}{\max\limits_{1\leq i\leq t}\sum_{j=1}^{m}|a_{t-1,ij}|}, \quad t \geq 1,$$

where $0 < \rho < 1$, and $a_{t,ij} = a_{t-1,ij} + v_{t,ij}$, where $v_{t,ij}$ are zero mean m.d. sequences with finite variance. This ensures that the maximum eigenvalue of $\boldsymbol{\Psi}_{t-1}$ is bounded from above by one in absolute value. A third extension concerns modelling the conditional variance of the error term u_t of an AR model via time varying persistent processes that can be estimated using kernel estimation methods. The latter extensions show the great scope for adapting the suggested framework to the needs of empirical researchers in applied macroeconometrics, and are current topics of research by the authors.

2.2. AR(1) model with no intercept

In this section we consider the AR(1) model y_t, (2.1), with no intercept.

2.2.1. Basic properties of y_t

In this subsection we investigate the structure of y_t and the properties of its covariance function. To write y_t as a moving average of the noise u_j, define the (random) weights

$$c_{t,0} = 1, \quad c_{t,j} := \rho_{t-1}\cdots\rho_{t-j}, \quad 1 \leq j \leq t.$$

Note that

$$|c_{t,j}| \leq \rho^j, \quad 1 \leq j \leq t. \tag{2.12}$$

The next theorem describes the basic properties of an AR(1) process y_t, $t = 1,\ldots,n$, (2.1), with no intercept.

Theorem 2.1. *Under Assumption 2.1, the process y_t of (2.1) has the following properties.*

(i) *y_t can be written as*

$$y_t = \sum_{j=0}^{t-1} c_{t,j}u_{t-j} + c_{t,t}y_0, \quad t \geq 1. \tag{2.13}$$

(ii) *The second and fourth moments satisfy*

$$\begin{aligned} Ey_t^2 &\leq 2(1-\rho)^{-2}(\sigma_u^2 + Ey_0^2), \\ Ey_t^4 &\leq 4(1-\rho)^{-4}(Eu_1^4 + Ey_0^4). \end{aligned} \tag{2.14}$$

The next theorem shows that y_t can be approximated by a truncated AR(1) process with an AR coefficient ρ_t,

$$z_t(\rho_t) := \sum_{k=0}^{t-1} \rho_t^k u_{t-k}, \quad t \geq 1, \tag{2.15}$$

and establishes the properties of the autocovariance $\mathrm{Cov}(y_{t+k}, y_t)$, as $t \to \infty$.

Theorem 2.2. *Suppose Assumptions 2.1 and 2.2 are satisfied. Then, as $t \to \infty$,*

$$y_t = z_t(\rho_t) + o_p(1), \quad Ey_t \to 0. \tag{2.16}$$

In addition, if ρ_t's, u_t's and y_0 are mutually independent, then

$$E[y_t y_{t+k}] = \sigma_u^2 E[\rho_t^k(1-\rho_t^2)^{-1}] + o(1), \quad t \to \infty,\ \forall k \geq 0, \tag{2.17}$$

$$|\mathrm{Cov}(y_t, y_{t+k})| \leq \rho^k(1-\rho^2)^{-1}(\sigma_u^2 + Ey_0^2), \quad \forall t \geq 1,\ k \geq 0. \tag{2.18}$$

2.2.2. Estimation and inference

In this section we construct a feasible estimation procedure for a path of the drifting coefficient process ρ_t, based on observations $y_1,\ldots,y_n$ of an AR(1) model (2.1) with no intercept. The proposed estimate of ρ_t can be written as a weighted sample autocorrelation at lag 1. Under Assumptions 2.1 and 2.2, it is consistent and asymptotically normally distributed. Computation of standard errors is straightforward and accommodates a martingale difference noise u_t. The method allows the construction of pointwise confidence intervals for the drifting coefficient ρ_t under minimal restrictions.

Let $H = H_n$ be a sequence of integers such that

$$H \to \infty, \qquad H = o(n). \tag{2.19}$$

The parameter ρ_t can be estimated by the moving window estimator

$$\hat{\rho}_{n,t} := \frac{\sum_{k=t-H}^{t+H} y_k y_{k-1}}{\sum_{k=t-H}^{t+H} y_{k-1}^2},$$

which is a local sample correlation of y_t's at lag 1, based on $2H+1$ observations $y_{t-H},\ldots,y_{t+H}$. This estimate belongs to a general class of kernel estimators considered in this paper. We will analyse properties of estimates

$$\hat{\rho}_{n,t} := \frac{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_k y_{k-1}}{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_{k-1}^2}, \tag{2.20}$$

where $K(x) \geq 0$, $x \in \mathbb{R}$ is a continuous bounded function (kernel) with a bounded first derivative such that $\int K(x)dx = 1$,

$$\begin{aligned} &K(x) = O(e^{-cx^2}), \quad \exists c > 0, \\ &|(d/dx)K(x)| = O(|x|^{-2}), \quad x \to \infty. \end{aligned} \tag{2.21}$$

Examples of K include

$K(x) = (1/2)I(|x| \leq 1)$, flat kernel,

$K(x) = (3/4)(1-x^2)I(|x| \leq 1)$, Epanechnikov kernel,

$K(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, Gaussian kernel.

The flat and Epanechnikov kernels have a finite support, whereas Gaussian kernel has an infinite support. The above kernels satisfy (2.21).

Next we discuss consistency and the asymptotic normality of the estimator $\hat{\rho}_{n,t}$ of (2.20). Denote for $1 \leq t, k \leq n$,

$$b_{tk} := K\left(\frac{t-k}{H}\right), \qquad \widehat{\sigma}_{Y,t}^2 = \sum_{k=1}^{n} b_{tk} y_{k-1}^2, \qquad \widehat{\sigma}_{Yu,t}^2 = \sum_{k=1}^{n} b_{tk}^2 y_{k-1}^2 u_k^2. \tag{2.22}$$

To establish the asymptotic normality of $\hat{\rho}_{n,t}$, we will need the conditional variance $V_j := E[u_j^2 | u_{j-1}, u_{j-2}, \ldots]$ of the stationary noise u_t to be bounded away from 0:

$$V_1 = E[u_1^2 | u_0, u_{-1}, \ldots] \geq c > 0, \quad \exists c > 0. \tag{2.23}$$

This assumption is not restrictive and satisfied, for example, by ARCH type white noises. Clearly, it holds when u_t is an i.i.d. process.

Notation: we set $\bar{H} = H$, if the kernel K has finite support, and $\bar{H} = H \log^{1/2} H$, if the kernel K has infinite support.

Theorem 2.3. *Let $y_1, \ldots, y_n$ be a sample from the AR(1) model with no intercept, (2.1), and $t = [n\tau]$, where $0 < \tau < 1$ is fixed. Assume that Assumptions 2.1 and 2.2 hold with some $\gamma \in (0, 1)$, and H and K satisfy (2.19) and (2.21), respectively.*

(i) *Then,*

$$\hat{\rho}_{n,t} - \rho_t = O_p\big((\bar{H}/n)^{\gamma} + H^{-1/2}\big). \tag{2.24}$$

(ii) *In addition, if u_t satisfies (2.23) and H is such that $(\bar{H}/n)^{\gamma} = o(H^{-1/2})$, then*

$$\frac{\widehat{\sigma}^2_{Y,t}}{\widehat{\sigma}_{Yu,t}} \Big(\hat{\rho}_{n,t} - \rho_t\Big) \to_D N(0, 1). \tag{2.25}$$

In particular, for $\gamma \geq 1/2$, (2.25) holds, if $H = o(n^{1/2}/\log^{1/4} n)$.

Observe that studentization in (2.25), adjusted to accommodate a martingale difference noise u_t, is different from the one used in the i.i.d. case.

Corollary 2.1 implies that the random normalization $\widehat{\sigma}^2_{Y,t}/\widehat{\sigma}_{Yu,t}$ in the normal approximation (2.25) yields the $\sqrt{H}$ rate of convergence.

Corollary 2.1. *Under assumptions of Theorem 2.3(ii),*

$$\frac{\widehat{\sigma}^2_{Y,t}}{\widehat{\sigma}_{Yu,t}} = H^{1/2} c_t (1 + o_p(1)),$$

where $c_1 \leq c_t \leq c_2$, for some finite constants $c_1, c_2 > 0$.

The above corollary is a direct implication of the results of Lemma A.2. The next corollary shows that studentization in (2.25) becomes operational using residuals $\hat{u}_j = y_j - \hat{\rho}_t y_{j-1}$. Let $\widehat{\sigma}^2_{Y\hat{u},t} := \sum_{j=1}^n b_{tj}^2 y_{j-1}^2 \widehat{u}_j^2$.

Corollary 2.2. *Under assumptions of Theorem 2.3(ii),*

$$\frac{\widehat{\sigma}^2_{Y,t}}{\widehat{\sigma}_{Y\hat{u},t}} \Big(\hat{\rho}_{n,t} - \rho_t\Big) \to_D N(0, 1).$$

Remark 2.5. The consistency of the estimator $\hat{\rho}_{n,t}$ is guaranteed by the persistence of the process ρ_t, or by stochastic or deterministic trending behaviour of the process a_t. The main restriction on a_j is the weak convergence $n^{-\gamma} a_{[\tau n]} \to_{D[0,1]} W_\tau + g(\tau)$ for some $0 < \gamma < 1$ where γ does not have to be known. The main example for a_t is a fractionally integrated process $a_t \sim I(d)$, $1/2 < d < 3/2$, discussed in Remark 2.1. It satisfies Assumption 2.2 with $\gamma = d - 1/2$. When γ is close to 0, the pattern of trending behaviour of a_t and the consistency rate in (2.24) deteriorate. For a stationary process a_t, the above estimation of ρ_t is not consistent. It is practical to choose $H = o(n^{1/2})$, leading to an asymptotic normality (2.25) for $1/2 \leq \gamma < 1$. In particular, $\gamma = 1/2$ corresponds to a unit root process $a_t \sim I(1)$.

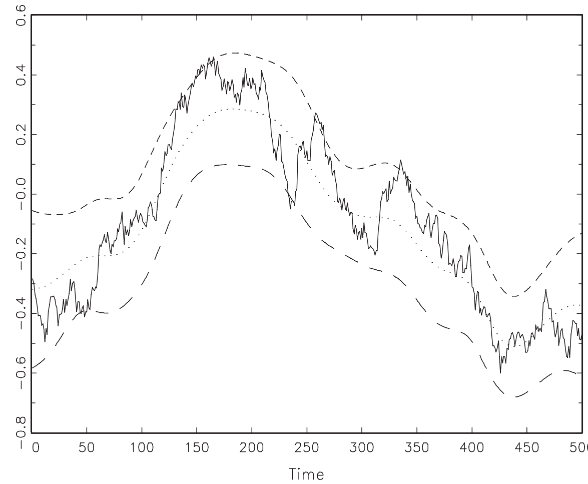

Fig. 1. Realizations of ρ_t, $\widehat{\rho}_{n,t}$ and 90% confidence intervals for ρ_t for the normal kernel.

To give an idea of the nature of the pointwise confidence intervals, implied by Theorem 2.3, we include Fig. 1 showing a realization of ρ_t based on a random walk model for a sample size of 500, its estimate $\hat{\rho}_{n,t}$ based on a normal kernel and a bandwidth $H = \sqrt{n}$, and 90% confidence intervals. The process ρ_t is well tracked and the point-wise confidence band contains ρ_t most of the time (for 92.8 of t's).

We complete this section describing properties of the weighted sample mean $\bar{y}_t := (\sum_{j=1}^n b_{tj} y_j)/(\sum_{k=1}^n b_{tk})$ of an AR(1) model with no intercept. By Theorem 2.2, $Ey_t \to 0$, as $t \to \infty$. We will show consistency of the estimate $\bar{y}_t \to_p 0$ and establish its normal approximation. We also obtain for $\bar{y}_t$ the martingale approximation (2.26) that resembles the well-known result for stationary linear processes, based on Beveridge–Nelson decomposition, see Phillips and Solo (1992).

Let $B_{1t} := \sum_{k=1}^n b_{tk}$, $B_{2t}^2 = \sum_{k=1}^n b_{tk}^2$, and $\bar{u}_t = B_{1t}^{-1} \sum_{j=1}^n b_{tj} u_j$.

Proposition 2.1. *Let y_t satisfy assumptions of Theorem 2.3(i), and $\bar{H} = o(n)$. Then,*

$$\bar{y}_t = (1 - \rho_t)^{-1} \bar{u}_t + O_p\big((\bar{H}/n)^{\gamma} + H^{-1}\big),$$
$$\bar{y}_t = O_p\big((\bar{H}/n)^{\gamma} + H^{-1/2}\big), \tag{2.26}$$

$$\frac{B_{1t}}{B_{2t}} \frac{1 - \rho_t}{\sigma_u} \bar{y}_t \to_D N(0, 1), \quad \text{if } (\bar{H}/n)^{\gamma} = o(H^{-1/2}). \tag{2.27}$$

2.3. AR(1) process with a persistent random attractor

We showed that an AR(1) model y_t with no intercept has asymptotically negligible mean, $Ey_t \to 0$, as $t \to \infty$, see (2.16). Next we extend this model by adding a persistent term μ_t, which in fixed coefficient AR models plays the role of the mean.

We decompose $y_t = \mu_t + (y_t - \mu_t)$ into a persistent (random) term μ_t, which we refer to as the *attractor* and a dynamic component $y_t - \mu_t$, that evolves as an AR(1) process (2.1):

$$y_t - \mu_t = \rho_{t-1}(y_{t-1} - \mu_{t-1}) + u_t, \quad t \geq 1. \tag{2.28}$$

This model can also be written as an AR(1) process $y_t = \alpha_t + \rho_{t-1} y_{t-1} + u_t$ with the intercept $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$. As seen below, although the attractor μ_t can be estimated, in general, it cannot be interpreted as the mean Ey_t.

To make decomposition (2.28) meaningful, a random process μ_t has to be persistent, so that both μ_t and ρ_t can be extracted from the data $y_1, \dots, y_n$. We assume that μ_t is bounded in the sense that $\max_j E\mu_j^2 < \infty$. To evaluate μ_t, ρ_t and α_t, we use estimates

$$\bar{y}_t = \frac{\sum_{j=1}^n b_{tj} y_j}{\sum_{j=1}^n b_{tj}}, \qquad \hat{\rho}_{n,t} := \frac{\sum_{k=1}^n b_{tk}(y_k - \bar{y}_t)(y_{k-1} - \bar{y}_t)}{\sum_{k=1}^n b_{tk}(y_{k-1} - \bar{y}_t)^2}, \qquad (2.29)$$
$$\widehat{\alpha}_t = \bar{y}_t - \widehat{\rho}_{n,t}\bar{y}_t.$$

Remark 2.6. The process y_t of (2.28) is *bounded*, i.e. $\max_j Ey_j^2 < \infty$. The latter implies

$$\max_j P(|y_j| \geq c) \leq c^{-2} \max_j Ey_j^2 \to 0, \quad c \to \infty.$$

To verify boundedness, notice, that $Ey_j^2 \leq 2E\mu_j^2 + 2E(y_j - \mu_j)^2$. According to (2.28), $y_t - \mu_t$ follows AR(1) model with no intercept, and therefore by (2.14), $\max_j E(y_t - \mu_t)^2 < \infty$. Hence, $\max_j Ey_j^2 < \infty$.

The next assumption describes a class of processes μ_t allowing estimation of μ_t and ρ_t.

Assumption 2.3. The attractor μ_t is $\mathcal{F}_t$ measurable, $\max_j E\mu_j^4 < \infty$, and satisfies either (i) or (ii) for $t \geq 1$, $1 \leq h < t/2$ with some $\beta \in (0, 1]$ and $C < \infty$.

(i) $E(\mu_t - \mu_{t+k})^2 \leq C(k/t)^{2\beta}$, $1 \leq k \leq h$.

(ii) $\mu_t - \mu_{t+k} = m(t, k) + \tilde{m}(t, k)$, where $Em^2(t, k) \leq C(k/t)^{2\beta}$, $1 \leq k \leq h$, and $\max_{1 \leq k \leq h} |\tilde{m}(t, k)| = O_p\big((h/t)^\beta + h^{-1}\big)$.

Example 2.2. A typical example of μ_t, satisfying Assumption 2.3(i), is

$$\mu_t = t^{-\beta}(v_1 + \dots + v_t) + t^{-1}(h_1 + \dots + h_t), \quad (0 < \beta \leq 1), \qquad (2.30)$$

where v_j's are stationary zero mean r.v.'s such that $E(v_1 + \dots + v_k)^2 \leq Ck^{2\beta}$, and h_j's are non-random numbers, $\max_j |h_j| < \infty$. It covers the case of a deterministic constant mean $\mu_t = t^{-1}(\mu + \dots + \mu) = \mu$, a time varying mean $\mu_t = g(t/n)$, and the case of a purely random attractor $\mu_t = t^{-1/2}\sum_{j=1}^t v_j$, where v_j's are i.i.d. random variables. An attractor, μ_t, satisfying Assumption 2.3(ii), arises in an AR(1) model with intercept, see Section 2.4. Assumption 2.3(i) is a subcase of (ii).

The next theorem establishes consistency rates and asymptotic normality of estimates of the parameters μ_t, ρ_t and α_t. It assumes $\beta \geq \gamma$, which means that the attractor μ_t is more persistent than ρ_t. We use notation $\widehat{\sigma}^2_{Y',t} = \sum_{k=1}^n b_{tk} y'^2_{k-1}$ and $\widehat{\sigma}^2_{Y'u,t} = \sum_{k=1}^n b^2_{tk} y'^2_{k-1} u_k^2$ where $y'_k = y_k - \mu_k$, $k \geq 1$. B_{1t} and B_{2t} are the same as in Proposition 2.1. Below notation $a_n \asymp b_n$ means that $a_n/b_n = O_p(1)$ and $b_n/a_n = O_p(1)$.

Theorem 2.4. *Let $y_1, \dots, y_n$ be a sample of AR(1) model with an attractor, (2.28), and $t = [n\tau]$, where $0 < \tau < 1$ is fixed. Assume that H and K satisfy (2.19) and (2.21), a_t satisfies Assumptions 2.1 and 2.2 with parameter $0 < \gamma < 1$, and μ_t satisfies Assumption 2.3 with parameter $\beta \geq \gamma$.*

(i) *Then, with $\kappa_n := (\bar{H}/n)^\gamma + H^{-1/2}$, $\bar{H} = o(n)$,*

$$\bar{y}_t - \mu_t = O_p(\kappa_n), \qquad \widehat{\rho}_{n,t} - \rho_t = O_p(\kappa_n), \qquad (2.31)$$
$$\widehat{\alpha}_{n,t} - \alpha_t = O_p(\kappa_n).$$

(ii) *Moreover, if $(\bar{H}/n)^\gamma = o(H^{-1/2})$, (2.23) holds and $E|u_1|^{4+\delta} < \infty$ for some $\delta > 0$, then*

$$\frac{B_{1t}}{B_{2t}}\frac{(1-\rho_t)}{\sigma_u}\big(\bar{y}_t - \mu_t\big) \to_D N(0, 1),$$
$$\frac{\widehat{\sigma}^2_{Y',t}}{\widehat{\sigma}_{Y'u,t}}\Big(\widehat{\rho}_{n,t} - \rho_t\Big) \to_D N(0, 1), \qquad (2.32)$$
$$\frac{B_{1t}}{B_{3t}}\Big(\widehat{\alpha}_{n,t} - \alpha_t\Big) \to_D N(0, 1),$$
$$B^2_{3t} := \sum_{j=1}^n b^2_{nj}\Big(1 - \mu_t \frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}} y'_{j-1}\Big)^2 u_j^2.$$

In addition, $B_{1t}/B_{2t} \asymp H^{1/2}$ and $B_{1t}/B_{3t} \asymp H^{1/2}$.

Studentization in (2.32) becomes operational by replacing u_k's with the residuals $\hat{u}_k = y_k - \hat{\rho}_t y_{k-1} - \widehat{\alpha}_t$, $k \geq 1$. Set $\widehat{\sigma}^2_{\hat{Y}\hat{u},t} := \sum_{k=1}^n b^2_{tk}\hat{y}^2_{k-1}\hat{u}^2_k$ and $\widehat{\sigma}^2_{\hat{Y},t} := \sum_{k=1}^n b_{tk}\hat{y}^2_{k-1}$, where $\hat{y}_k = y_k - \bar{y}_t$.

Corollary 2.3. *Let assumptions of Theorem 2.4(ii) hold. Then*

$$\left|\frac{1 - \widehat{\rho}_{n,t}}{1 + \widehat{\rho}_{n,t}}\right|^{1/2} \frac{B^{3/2}_{1t}}{B_{2t}\,\widehat{\sigma}_{\hat{Y},t}}(\bar{y}_t - \mu_t) \to_D N(0, 1),$$
$$\frac{\widehat{\sigma}^2_{\hat{Y},t}}{\widehat{\sigma}_{\hat{Y}\hat{u},t}}\Big(\hat{\rho}_{n,t} - \rho_t\Big) \to_D N(0, 1), \qquad (2.33)$$
$$\frac{B_{1t}}{\widehat{B}_{3t}}\Big(\widehat{\alpha}_{n,t} - \alpha_t\Big) \to_D N(0, 1),$$
$$\widehat{B}^2_{3t} := \sum_{j=1}^n b^2_{nj}\Big(1 - \bar{y}_t \frac{B_{1t}}{\widehat{\sigma}^2_{\hat{Y},t}}(y_{j-1} - \bar{y}_t)\Big)^2 \hat{u}^2_j.$$

We showed, that an AR(1) process y_t defined by an attractor μ_t and an AR(1) dynamics ρ_t, is bounded and allows extraction of μ_t and ρ_t together with their confidence bands. The results of this section present a useful tool for inference, analysis and modelling of dynamics of bounded non-stationary processes.

To justify the attractor terminology, we include Fig. 2 which shows the plots of y_t and μ_t, based on bounded random walk models for ρ_t and μ_t, for a sample size of 500, the plot of the estimate $\bar{y}_t$ obtained using a normal kernel and a bandwidth $H = \sqrt{n}$, and 90% confidence intervals for μ_t. It shows that the process μ_t is well tracked and the point-wise confidence band contains the true process most of the time (for 85.4% of t's).

2.4. AR(1) model with intercept

Next we discuss an AR model

$$y_t = \alpha_t + \rho_{t-1}y_{t-1} + u_t, \quad t \geq 1, \qquad (2.34)$$

with a persistent intercept α_t, where ρ_t and u_t are as in (2.1). Similarly as in (2.13), y_t can be written as an AR(1) model

$$y_t = \rho_{t-1}y_{t-1} + \{\alpha_t + u_t\} = \left\{\sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i}\right\} + \left\{\sum_{i=0}^{t-1} c_{t,i}u_{t-i} + c_{t,t}y_0\right\} =: \mu_t + y'_t, \qquad (2.35)$$

with attractor μ_t, where y'_t is an AR(1) process with no intercept, $y'_t = \rho_{t-1}y'_{t-1} + u_t$, $t \geq 1$, $y'_0 = y_0$.

Estimation of this model reduces to that of a model with a persistent attractor, discussed in Section 2.3. The following assumption describes a class of permissible processes α_t allowing inference.

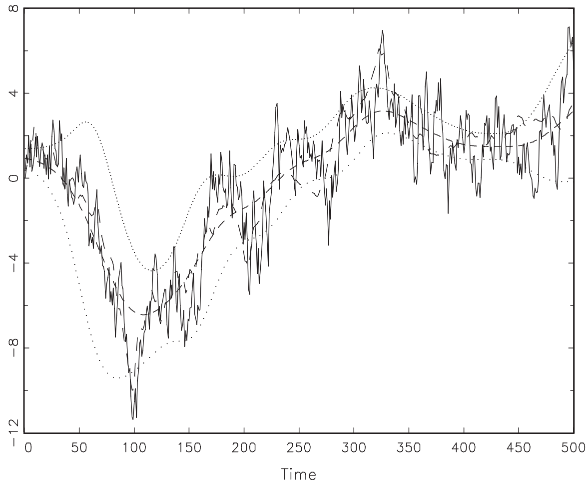

Fig. 2. Realization of $\mu_t, \bar{y}_t, y_t$ and 90% confidence intervals for μ_t using the normal kernel.

Assumption 2.4. The process α_t is $\mathcal{F}_t$ measurable, $\max_j E\alpha_j^4 < \infty$, and for some $\beta \in (0, 1]$, $E(\alpha_t - \alpha_{t+k})^2 \le C(k/t)^{2\beta}$, $t \ge 1$, $1 \le k < t/2$.

A standard example of a process α_t satisfying Assumption 2.4 is provided in (2.30). Under Assumption 2.4, the corresponding attractor μ_t in (2.35) satisfies Assumption 2.3:

Proposition 2.2. *Suppose that y_t is an AR(1) process (2.34), where ρ_t is generated by a_t that satisfies Assumptions 2.1 and 2.2 with parameter γ, and α_t satisfies Assumption 2.4 with parameter $\beta \ge \gamma$. Then μ_t in (2.35) satisfies Assumption 2.3(ii) with $\beta = \gamma$.*

Moreover, as $t \to \infty$,

$$\mu_t = (1 - \rho_t)^{-1}\alpha_t + o_p(1). \quad (2.36)$$

Since μ_t satisfies Assumption 2.3, estimates $\widehat{\alpha}_{n,t}$, $\widehat{\rho}_{n,t}$ and $\bar{y}_t$ of parameters α_t, ρ_t and μ_t of this model have properties described in Theorem 2.4 and Corollary 2.3.

Relations between the attractor $\mu_t = \sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i}$ and the intercept $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$ indicate that persistence in μ_t generates persistence in the intercept α_t and vice versa.

3. Monte Carlo study

In the following Monte Carlo simulation we study the small sample performance of the kernel estimator of a random AR(1) coefficient process, for the sample size $n = 50, 100, 200, 400, 800, 1000$. In the first set of simulations we generate data by AR(1) model (2.1) with no intercept

$$y_t = \rho_{t-1}y_{t-1} + u_t, \quad t \ge 1$$

using restriction $\rho_t = \rho a_t / \max_{0\le j\le t} |a_j|$, which bounds ρ_t between $-\rho$ and ρ. We set $\rho = 0.9$. The differences $a_t - a_{t-1} = v_t$ are modelled by stationary AR(1) and long memory ARFIMA processes.

To estimate ρ_t, we use a two-sided normal kernel estimator. The bandwidth H is set to take values n^α, $\alpha = 0.2, 0.4, 0.5, 0.6, 0.8$. The value $\alpha = 0.5$ corresponds to the closest value to the optimal bandwidth, minimizing the mean square error $E(\hat{\rho}_{n,t} - \rho_t)^2$ in pointwise estimation. The global performance of the estimator is evaluated by the average value of the mean squared error, $MSE := n^{-1}\sum_{t=1}^{n}(\hat{\rho}_{n,t} - \rho_t)^2$, computed using 1000 Monte-Carlo replications.

Table A.1 reports the average MSE and 90% coverage probabilities (CP) for the normal kernel estimate when v_t follows a short memory AR(1) model $v_t = \phi v_{t-1} + \varepsilon_t$, where ϕ is set to take values 0, 0.2, 0.5, 0.9, and ε_t is a standard normal i.i.d. noise. Here, a_t is an $I(1)$ (unit root) process, and satisfies (2.4) with $\gamma = 1/2$. It is evident that, for $\phi = 0$ and "optimal" bandwidth $H = n^{0.5}$, the average MSE falls substantially with the sample size. This bandwidth choice is best in terms of MSE. For coverage probabilities we observe that a slightly lower bandwidth value of $H = n^{0.4}$ is best.

The presence of short memory dependence in v_t does not seem to affect the estimator adversely. If anything, the performance of the estimator improves slightly as v_t becomes more persistent.

Table A.2 reports the average MSE and 90% coverage probabilities of a normal kernel estimation of the model y_t, when v_t is a stationary long memory ARFIMA process $(1 - L)^{d-1}v_t = \varepsilon_t$ and ε_t is the standard normal i.i.d. noise. The parameter d is set to take values $d = 0.51, 0.75, 1.25, 1.49$. The process a_t is a non-stationary integrated $I(d)$ process, satisfying assumption (2.4) with parameter $\gamma = d - 1/2$, taking values $\gamma = 0.01, 0.25, 0.75, 0.99$, and the persistence of a_t increases with d. We clearly see the familiar patterns observed for a short memory v_t, whereby larger values of d and γ, lead to stronger persistence in a_t and improved quality of estimation and inference for ρ_t, as suggested by the theory.

The simulation analysis above was focused on the AR(1) model y_t with i.i.d. errors u_t and no intercept. In addition, our theory allows for a model

$$y_t = \alpha_t + \rho_{t-1}y_{t-1} + u_t, \quad t \ge 1$$

with time varying intercept α_t, and for a general martingale difference noise u_t. The second set of simulations illustrates the small sample properties of estimators $\hat{\alpha}_{n,t}$, $\hat{\rho}_{n,t}$ and $\bar{y}_t$, under martingale difference noise u_t. The same set of bandwidths and the sample sizes is used as before, and ρ_t is defined as before setting $\rho = 0.9$. We set the time varying intercept to be a bounded random walk $\alpha_t = t^{-1/2}\sum_{j=1}^{t}\eta_j$, generated by another standard normal i.i.d. noise η_j. To analyse the impact of heteroscedasticity of the noise u_t on estimation of ρ_t, we consider two heteroscedastic specifications of u_t:

(a) $GARCH(1, 1)$ m.d. noise $u_t = \sigma_t\varepsilon_t$, where $\sigma_t^2 = 1 + 0.25u_{t-1}^2 + 0.25\sigma_{t-1}^2$,

(b) stochastic volatility m.d. noise $u_t = \exp(h_{t-1})\varepsilon_t$, $h_t = 0.7h_{t-1} + \varepsilon_t$.

In (a) and (b), ε_t is set to be a normal i.i.d. noise and the resulting u_t process is normalized to have unit variance. Table A.3 reports the average MSE and 90% coverage probabilities of the joint estimation of the parameters α_t, ρ_t and μ_t of the model y_t, using the normal kernel. A consideration of Table A.3 suggests that the estimator of ρ_t is not affected by the presence of the intercept, and the estimators of α_t and μ_t follow similar patterns to that of ρ_t, albeit with different absolute levels for the MSEs. It is evident, that heteroscedasticity may increase MSE and, therefore, needs to be accounted for in the studentization, see e.g. (2.25), but in general, it does not affect estimation, as suggested by our theory. The coverage probabilities suggest that the estimated confidence intervals are satisfactory for $\hat{\rho}_{n,t}$, $\bar{y}_t$ and $\hat{\alpha}_{n,t}$.

4. Empirical application

In this section we use the kernel estimator to provide new evidence for two debates that have attracted considerable attention in empirical macroeconomics. These debates relate to the time-varying persistence of inflation and the validity of the PPP hypothesis.

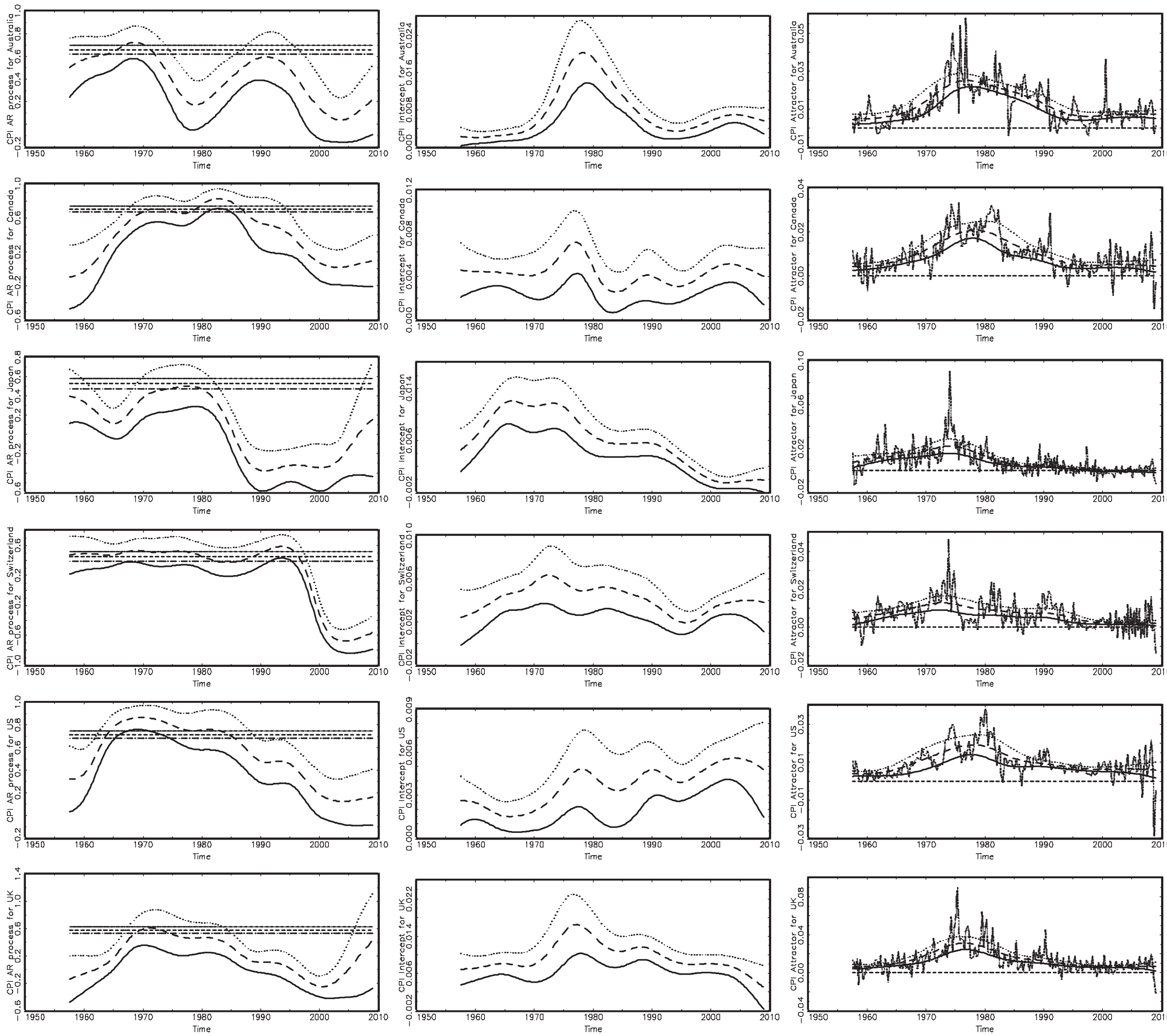

Fig. 3. Time-Varying AR coefficient, intercept, attractor and 90% confidence bands in AR(1) model, fitted to CPI inflation data using a normal kernel for 6 countries: Australia, Canada, Japan, Switzerland, US and UK. The AR coefficient panels also report the estimate of an AR parameter in a fixed coefficient AR(1) model together with its 90% confidence bands. CPI inflation data are also presented in the third column.

4.1. Data and setup

Our CPI inflation dataset is made up of 6 countries: Australia, Canada, Japan, Switzerland, US and UK. The real exchange rate (RER) dataset is made up of 6 countries where the US dollar is the base currency (and so obviously is itself excluded): Australia, Canada, Japan, Norway Switzerland and UK. The data span is 1957Q1 to 2009Q1. All data are obtained from the IMF (International Financial Statistics (IFS)). We construct the bilateral real exchange rate q against the ith currency at time t as $q_{i,t} = s_{i,t} + p_{j,t} - p_{i,t}$, where $s_{i,t}$ is the corresponding nominal exchange rate (ith currency units per one unit of the jth currency), $p_{j,t}$ the price level (CPI) in the jth country, and $p_{i,t}$ the price level of the ith country. That is, a rise in $q_{i,t}$ implies a real appreciation of the jth country's currency against the ith country's currency.

We fit an AR(1) model with a time varying autoregressive coefficient and an intercept term which is allowed to vary over time as well. Parameters are estimated using the normal and flat kernel estimators presented in Section 2 but having obtained similar results for both kernels, we choose to report results only for the normal kernel due to space considerations. A bandwidth H equal to $n^{1/2}$ is used as suggested by theory. Results are reported pictorially in Figs. 3 and 4. Fig. 3 relates to CPI inflation and Fig. 4 to real exchange rates. They report the estimated time-varying AR coefficient, intercept term, attractor and the standard time-invariant AR(1) coefficient together with their 90% point-wise confidence bands.

4.2. Empirical results

The empirical results presented in Figs. 3 and 4 can help provide answers to two important empirical topics: the origin of the persistence of inflation and the real exchange rate. We will examine each issue in turn.

4.2.1. Inflation persistence

Our first application examines whether inflation persistence has changed over time. As noted above, (Cogley et al., 2010) document using an RC model that inflation gap persistence rose during the Great Inflation of the 1970s, then fell in the 1980s. Benati (2010) presents similar findings using different techniques:

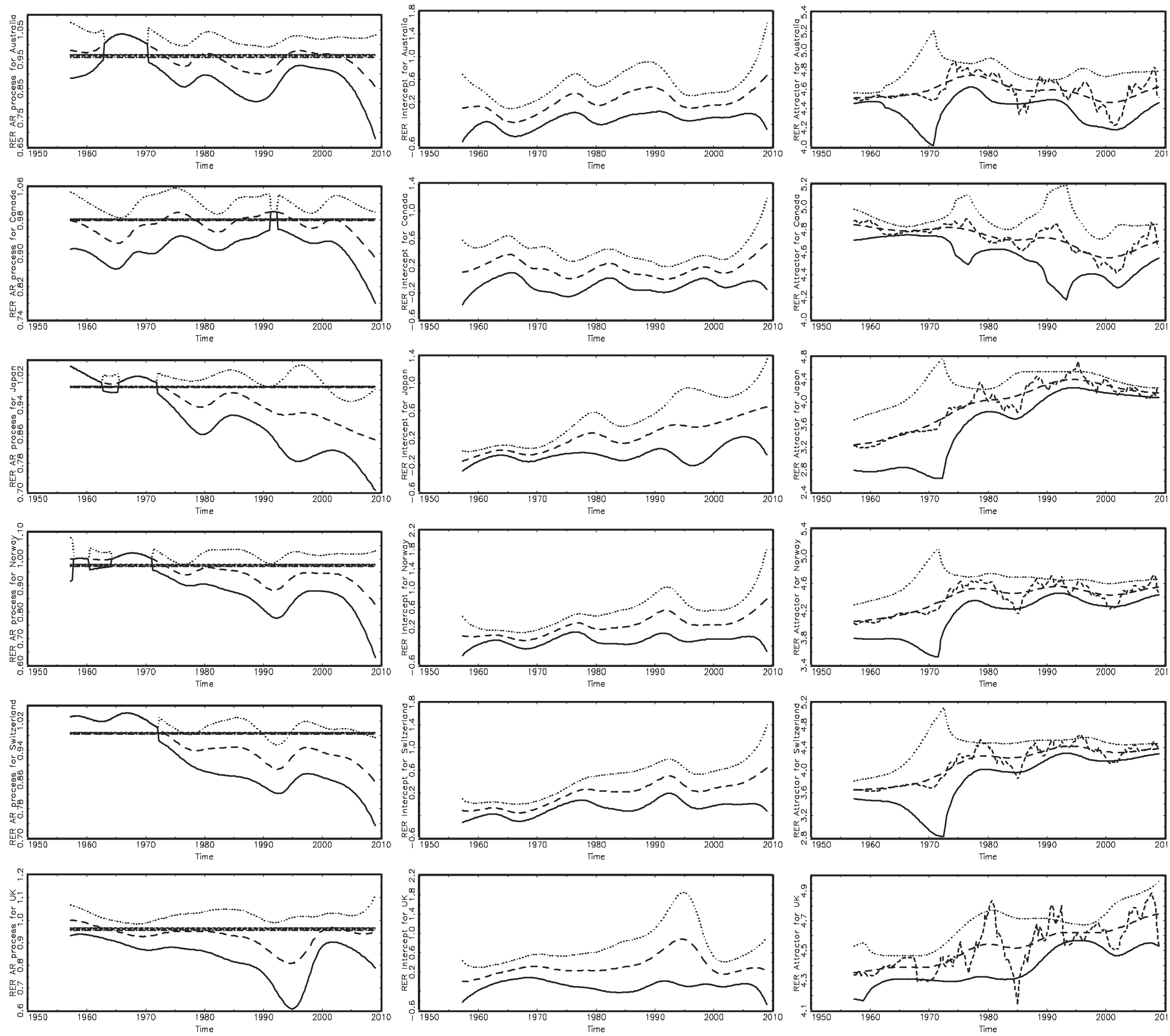

Fig. 4. Time-Varying AR coefficient, intercept, attractor and 90% confidence bands in AR(1) model, fitted to real exchange rates using a normal kernel for 6 countries: Australia, Canada, Japan, Switzerland, US and UK. The AR coefficient panels also report the estimate of an AR parameter in a fixed coefficient AR(1) model together with its 90% confidence bands. Real exchange rate data are also presented in the third column.

sub-sample estimates of a fixed-coefficient univariate model for inflation, and of a DSGE model that encodes inflation persistence into price-setting.

Establishing whether inflation persistence has changed over time can help shed light on its causes. The more it is observed to have changed, the less it is likely that this persistence is a product of hard-wired features of price-setting like those described by Christiano et al. (2005) and Smets and Wouters (2003) and the more likely it is that this persistence reflects changes in the monetary regime. Benati (2010) takes this view, inferring from the fact that both structural DSGE and time-series estimates of inflation persistence are highly variable across monetary regimes, that inflation persistence has its origin in the nature of monetary policy and not price-setting.

Fig. 3 records our results. The left hand column reports estimates of the time varying AR coefficient (with horizontal lines depicting the whole-sample, fixed-coefficient counterparts and associated confidence intervals); the middle column shows estimates of a time-varying intercept; and the third column shows estimates of the 'attractor'. Overall, it is quite clear that inflation persistence has varied considerably, and, once confidence bands are taken into account, statistically significantly, over time. The assumption of a fixed autoregressive coefficient is therefore judged inappropriate by the kernel estimator. One could justify the fixed AR model with sufficiently strong a priori views about the invariance of the economic sources of persistence. However, on the contrary, there is every reason to suspect from the perspective of economic theory, and prior work, that some of the economic sources of persistence have changed. We pick out a couple of examples that connect with previous work: inflation persistence in the US is estimated to have risen from about 0.4 at the start of the sample, to a peak of around 0.8 in 1970, falling thereafter, and steadily, to around 0.2 in the 2000s. A similar picture emerges in the UK, the one contrasting finding being that after falling back to a low point of almost −0.2 in 2000, inflation persistence is estimated to have risen rapidly through the subsequent decade. This overall pattern is not shared by every country. For example, inflation persistence in Australia seems to cycle around 0.4, with a slight downward trend; in Switzerland, inflation persistence is stable around 0.6 until 1995 or so when it falls rapidly to an average in the 2000s of −0.6. However, the basic fact, that inflation persistence shows significant time variation, is common to all countries.

Our findings add to the evidence casting doubt on the argument that this persistence has its route in nominal or real frictions, since it seems to stretch the plausibility of those models that they can be viewed as time-varying. Instead, this time-variation suggests that persistence is more likely to have its origins in monetary policy. In several countries persistence ends the sample lower than it began it, and the conjecture that this might have been caused by monetary policy accords well with other work that has documented changes in the institutions and philosophy governing monetary policy. During this time, there have been many changes in the institutions governing monetary policy, including the spread of independent central banks, the adoption of inflation or other similar targets. These facts are documented by, amongst others (Segalotto et al., 2006). Moreover, there arose the widespread acceptance of the doctrine that inflation is caused by and can be tamed by monetary policy, and that unemployment cannot be permanently held down by loose monetary policy, a doctrine that was not at all universal at the start of the sample period, see, for example, Nelson (2005).

4.2.2. Persistence of deviations from PPP

Our second application considers the debate surrounding the persistence in deviations of relative prices from purchasing power parity (PPP). A vast literature has focused on this problem, so we motivate our analysis with only a few examples. The survey by Rogoff (1996) adduces the essential finding in many papers that deviations from PPP take a very long time to die out. We note selectively the work of Frankel and Rose (1996), Papell (1997), Papell and Prodan (2006), Papell and Theodoridis (1998, 2001), Chortareas et al. (2002) and Chortareas and Kapetanios (2009). One reason that persistent deviations from PPP can occur is because of nominal rigidities. But Chari et al. (2002) note that this persistence in the data – they report an autoregressive coefficient of around 0.8 for 8 US bilateral real exchange rates – is greater than can be plausibly accounted for by nominal stickiness in traded goods prices. Benigno (2002) offers another explanation, illustrating how the persistence of the real exchange rate is in part a function of the degree of interest rate inertia in monetary policy. Imbs et al. (2005) and Chen and Engel (2005) have debated whether real exchange rate persistence is a function of aggregation bias, discussing differences between the persistence of the aggregate and its subcomponents. A final possibility is that the dynamics of PPP are affected by Balassa–Samuelson effects. When non-traded goods like labour or land are in short supply, productivity improvements in the traded sector bid up non-traded prices (higher incomes in the traded sector translate to increased demand for non-traded services) and hence the real exchange rate.

Our results are shown in Fig. 4, using the same format as the inflation persistence charts. With the exception of the UK, real exchange rate persistence seems to be lower at the end of the sample than at the beginning, with the caveat that the movements in general, are smaller than for inflation persistence, and in particular smaller relative to the confidence interval around the estimate for each period. We discern clear upward movements in the time varying constant in Japan, Norway, Switzerland and Australia. With the exception of Switzerland, these are plausibly connected with permanent Balassa–Samuelson innovations. For Canada, Norway and Australia upward movements would be associated with the increase in the productivity of their traded sectors with the discovery and/or increase in demand for raw material exports (from, e.g., the emerging Asian economies like China). In the case of Japan this would correspond to the widely studied increase in productivity in their manufacturing export sector following the second world war. As we have already noted, the possibility that real exchange rate persistence is a function solely of nominal rigidities was already strongly at odds with plausible sticky price models. Our findings, of time-varying persistence, emphasize this.

5. Concluding remarks

This paper has proposed a new class of time-varying coefficient models, allowing the decomposition of a non-stationary time series into a persistent random attractor and a process with time-varying autoregressive dynamics. The paper suggests a kernel approach for the estimation and inference of the unobserved time-varying coefficients and provides a rigorous theoretical analysis of its properties.

The proposed estimation approach has desirable properties such as consistency and studentized asymptotic normality under very weak conditions. The potential of our theoretical findings has been supported by an extensive Monte Carlo study and illustrated by some informative empirical findings relating to CPI inflation persistence and the PPP hypothesis. In particular, we have uncovered evidence in support of the PPP hypothesis for the recent past. Our findings suggest that estimating coefficient processes via kernels is robust to a number of aspects of the nature of the unobserved process such as whether it is deterministic or stochastic and to the exact specification of the process. The theoretical properties of the kernel estimator are to be contrasted with the lack of knowledge about the properties of state-space estimates of RC models which display pathologies that our approach avoids, as documented in Stock and Watson (1998) and Koop and Potter (2008).

One further extremely attractive aspect of the new estimator relates to its relative computational tractability. Estimation of RC models using standard methods, including Bayesian estimation, is extremely computationally demanding. The computational demands, associated with the use of kernel type estimates, are modest, with the estimation of even moderately large multivariate models being completed almost instantly.

At this point it might be worth summarizing a possible course of action for empirical researchers faced with the task of modelling time-variation in macroeconomic time series. It is reasonable to assume that researchers do not know whether the true coefficient process is random or not. In the absence of such information and given the theoretical findings in this paper, there is a sound case in favour of adopting a kernel estimator, since this estimator is valid both for deterministic and stochastic coefficient processes. This case is strengthened by our Monte Carlo evidence which shows that the estimator works well in small samples.

Acknowledgements

Giraitis' research is supported by the ESRC grant RES062230790. We would like to thank Peter Phillips, Donald Andrews and participants at seminars in Yale University, the Bank of England and Queen Mary, University of London, for their useful remarks. We thank the Co-Editor and two anonymous referees for constructive comments and suggestions that led to substantial improvements in the paper.

Appendix

A.1. Proof of the main results

In this subsection we prove Theorems 2.1–2.4, Corollaries 2.1 and 2.3, and Propositions 2.1 and 2.2. In the sequel, we use repeatedly the following properties of b_{nj}'s, valid under (2.21):

$$B_{1t} := \sum_{k=1}^{n} b_{tk} \sim H \int K(x)dx = H,$$
$$B_{2t}^2 := \sum_{k=1}^{n} b_{tk}^2 \sim H \int K^2(x)dx =: H\beta_K, \tag{A.1}$$

where $\beta_K = \int K^2(x)dx$.

Table A.1
MSE and 90% Coverage Probability results for $\widehat{\rho}_{n,t}$ for the normal kernel.

ϕ	H	MSE						Coverage Probability, $t=[n/2]$					
		n											
		50	100	200	400	800	1000	50	100	200	400	800	1000
	$n^{0.2}$	0.10	0.08	0.07	0.06	0.05	0.05	0.79	0.79	0.79	0.82	0.84	0.83
	$n^{0.4}$	0.05	0.04	0.03	0.02	0.02	0.01	0.84	0.84	0.84	0.84	0.84	0.85
0	$n^{0.5}$	0.04	0.02	0.02	0.01	0.01	0.01	0.84	0.85	0.88	0.83	0.79	0.77
	$n^{0.6}$	0.03	0.02	0.01	0.01	0.01	0.01	0.86	0.83	0.79	0.74	0.67	0.61
	$n^{0.8}$	0.02	0.02	0.02	0.02	0.03	0.03	0.86	0.76	0.65	0.42	0.31	0.26
	$n^{0.2}$	0.10	0.08	0.07	0.06	0.05	0.05	0.77	0.78	0.80	0.83	0.83	0.83
	$n^{0.4}$	0.05	0.03	0.03	0.02	0.02	0.02	0.83	0.86	0.86	0.82	0.85	0.84
0.2	$n^{0.5}$	0.04	0.02	0.02	0.01	0.01	0.01	0.85	0.84	0.86	0.82	0.78	0.78
	$n^{0.6}$	0.03	0.02	0.01	0.01	0.01	0.01	0.86	0.82	0.78	0.73	0.65	0.64
	$n^{0.8}$	0.02	0.02	0.02	0.02	0.03	0.04	0.83	0.79	0.63	0.45	0.30	0.27
	$n^{0.2}$	0.10	0.08	0.07	0.06	0.05	0.05	0.76	0.79	0.79	0.83	0.81	0.82
	$n^{0.4}$	0.05	0.03	0.03	0.02	0.02	0.01	0.79	0.84	0.86	0.85	0.86	0.85
0.5	$n^{0.5}$	0.04	0.02	0.02	0.01	0.01	0.01	0.84	0.85	0.84	0.81	0.78	0.79
	$n^{0.6}$	0.03	0.02	0.01	0.01	0.01	0.01	0.87	0.82	0.81	0.73	0.64	0.63
	$n^{0.8}$	0.02	0.02	0.02	0.02	0.03	0.04	0.83	0.78	0.62	0.45	0.30	0.24
	$n^{0.2}$	0.09	0.08	0.07	0.06	0.05	0.05	0.78	0.77	0.79	0.82	0.81	0.81
	$n^{0.4}$	0.05	0.03	0.02	0.02	0.01	0.01	0.84	0.85	0.87	0.84	0.86	0.85
0.9	$n^{0.5}$	0.03	0.02	0.02	0.01	0.01	0.01	0.84	0.86	0.87	0.85	0.81	0.81
	$n^{0.6}$	0.02	0.02	0.01	0.01	0.01	0.01	0.87	0.84	0.85	0.77	0.67	0.65
	$n^{0.8}$	0.02	0.02	0.02	0.02	0.03	0.04	0.88	0.79	0.65	0.46	0.28	0.24

Notes: The model is $y_t=\rho_t y_{t-1}+u_t$, $u_t\sim$ i.i.d., $\rho_t=0.9a_t/\max_{0\le j\le t}|a_j|$, $a_t-a_{t-1}=v_t$ follows AR(1) model with parameter ϕ.

Proof of Theorem 2.1. (i) Eq. (2.13) follows using recursions

$$\begin{aligned} y_t &= \rho_{t-1}y_{t-1}+u_t=\rho_{t-1}(\rho_{t-2}y_{t-2}+u_{t-1})+u_t \\ &= \rho_{t-1}\rho_{t-2}\rho_{t-3}y_{t-3}+\rho_{t-1}\rho_{t-2}u_{t-2}+\rho_{t-1}u_{t-1}+u_t \\ &= \rho_{t-1}\cdots\rho_0 y_0+\rho_{t-1}\cdots\rho_1 u_1+\cdots+\rho_{t-1}u_{t-1}+u_t \\ &= c_{t,t}y_0+c_{t,t-1}u_1+c_{t,t-2}u_2+\cdots+c_{t,1}u_{t-1}+c_{t,0}u_t. \end{aligned}$$

The above recursion also implies

$$y_t=\sum_{j=0}^{k}c_{t,j}u_{t-j}+c_{t,k+1}y_{t-k-1},\qquad 1\le k\le t-1. \tag{A.2}$$

(ii) To prove (2.14), use $|c_{t,j}|\le\rho^j$, $E|u_ju_k|\le Eu_1^2$, and (2.13), to obtain

$$\begin{aligned} Ey_t^2 &= E\left(\sum_{j=0}^{t-1}c_{t,j}u_{t-j}+c_{t,t}y_0\right)^2\le 2E\left(\sum_{j=0}^{t-1}\rho^j|u_{t-j}|\right)^2+2\rho^{2t}Ey_0^2 \\ &\le 2\left(\left(\sum_{j=0}^{t-1}\rho^j\right)^2+\rho^{2t}\right)(\sigma_u^2+Ey_0^2) \\ &\le 2(1-\rho)^{-2}(\sigma_u^2+Ey_0^2). \end{aligned}$$

Similarly, since $E|u_{j_1}\cdots u_{j_4}|\le(Eu_{j_1}^4\cdots Eu_{j_4}^4)^{1/4}=Eu_1^4$,

$$\begin{aligned} Ey_t^4 &\le E\left(\sum_{j=0}^{t-1}\rho^j|u_{t-j}|+\rho^t y_0\right)^4\le 4\left(\left(\sum_{j=0}^{t-1}\rho^j\right)^4Eu_1^4+\rho^{4t}Ey_0^4\right) \\ &\le 4(1-\rho)^{-4}(Eu_1^4+Ey_0^4).\qquad\square \end{aligned}$$

Proof of Theorem 2.2. The first equality in (2.16) holds, because (A.20) of Lemma A.1 implies $E(y_t-z_t(\rho_t))^2=o(1)$, as $t\to\infty$. To show the second claim of (2.16), setting $h=\log t$, write $y_t=z_t(\rho_{t-h})+\{y_t-z_t(\rho_{t-h})\}$. Because ρ_{t-h} is $\mathcal{F}_{t-k-1}$, $k<h$ measurable, then $E[\rho_{t-h}^k u_{t-k}]=E[\rho_{t-h}^k E[u_{t-k}|\mathcal{F}_{t-k-1}]]=0$ for $k<h$, and therefore $|Ez_t(\rho_{t-h})|=|E\sum_{k=0}^{t-1}\rho_{t-h}^k u_{t-k}|=|E\sum_{k=h}^{t-1}\rho_{t-h}^k u_{t-k}|\le\sum_{k=h}^{t-1}\rho^k E|u_{t-k}|=O(\rho^h)\to 0$. Since by (A.20), $E(y_t-z_t(\rho_{t-h}))^2=o(1)$, this proves $Ey_t\to 0$, as $t\to\infty$.

Proof of (2.17). Let $k\ge 0$. It suffices to show that, as $t\to\infty$,

$$\begin{aligned} &E\big[y_{t+k}y_t-z_t(\rho_t)z_{t+k}(\rho_t)\big]\to 0, \\ &E\big[z_t(\rho_t)z_{t+k}(\rho_t)\big]-\sigma_u^2E[\rho_t^k(1-\rho_t^2)^{-1}\big]\to 0. \end{aligned} \tag{A.3}$$

By (A.20), $E(y_t-z_t(\rho_t))^2=o(1)$ and $E(y_{t+k}-z_{t+k}(\rho_t))^2=o(1)$. By (2.14), $Ey_t^2<\infty$ and $Ez_{t+k}^2<\infty$. Thus, by the Cauchy inequality, $E|y_{t+k}y_t-z_t(\rho_t)z_{t+k}(\rho_t)|\le E|(y_{t+k}-z_{t+k}(\rho_t))y_t|+E|z_{t+k}(\rho_t)(y_t-z_t(\rho_t))|\le C\{E(y_{t+k}-z_{t+k}(\rho_t))^2\}^{1/2}+C\{E(y_t-z_t(\rho_t))^2\}^{1/2}=o(1)$, which verifies the first claim in (A.3). To show the second claim, use the independence between ρ_j's and u_j's, to obtain $E[z_t(\rho_t)z_{t+k}(\rho_t)]=\sigma_u^2E[\sum_{i=0}^{t-1}\rho_t^{k+2i}]\to\sigma_u^2E[\rho_t^k\sum_{i=0}^{\infty}\rho_t^{2i}]=\sigma_u^2E[\rho_t^k(1-\rho_t^2)^{-1}]$, as $t\to\infty$, which proves (2.17).

Proof of (2.18). By (A.2), and mutual independence of ρ_t's, u_t's and y_0,

$$\begin{aligned} Ey_t^2 &= E\left(\sum_{j=0}^{t-1}c_{t,j}u_{t-j}+c_{t,t}y_0\right)^2=E\left(\sum_{j=0}^{t-1}c_{t,j}u_{t-j}\right)^2+E\big(c_{t,t}y_0\big)^2 \\ &\le\sum_{j=0}^{t-1}E[c_{t,j}^2]\sigma_u^2+c_{t,t}^2Ey_0^2\le(\sigma_u^2+Ey_0^2)\sum_{j=0}^{\infty}\rho^{2j} \\ &\le(\sigma_u^2+Ey_0^2)(1-\rho^2)^{-1}. \end{aligned} \tag{A.4}$$

Next, by (A.2) and the fact that $E[u_{t+i}y_t]=E[y_tE[u_{t+i}|\mathcal{F}_t]]=0$, $i\ge 1$,

$$\begin{aligned} \mathrm{Cov}(y_{t+k},y_t) &= \mathrm{Cov}\left(\sum_{j=0}^{k-1}c_{t+k,j}u_{t+k-j}+c_{t+k,k}y_t,\ y_t\right) \\ &= \mathrm{Cov}\left(c_{t+k,k}y_t,\ y_t\right)\le(E[c_{t+k,k}^2y_t^2])^{1/2}(Ey_t^2)^{1/2} \\ &\le\rho^kEy_t^2,\quad k\ge 1, \end{aligned}$$

which together with (A.4) implies (2.18). $\square$

Proof of Theorem 2.3. (i) By $y_j=\rho_{j-1}y_{j-1}+u_j$,

$$\begin{aligned} \widehat{\rho}_{n,t}-\rho_t &= \widehat{\sigma}_{Y,t}^{-2}\left[\sum_{j=1}^{n}b_{tj}y_{j-1}u_j+\sum_{j=1}^{n}b_{tj}(\rho_{j-1}-\rho_t)y_{j-1}^2\right] \\ &=: \widehat{\sigma}_{Y,t}^{-2}\big[S_{Yu,t}+r_{n,t}\big]. \end{aligned} \tag{A.5}$$

Table A.2
MSE and 90% Coverage Probability results for $\widehat{\rho}_{n,t}$ for the normal kernel.

d	H	MSE						Coverage Probability, $t=[n/2]$					
		n											
		50	100	200	400	800	1000	50	100	200	400	800	1000
0.51	$n^{0.2}$	0.11	0.10	0.09	0.08	0.08	0.08	0.75	0.75	0.78	0.78	0.78	0.76
	$n^{0.4}$	0.07	0.06	0.06	0.05	0.05	0.04	0.75	0.74	0.72	0.70	0.67	0.66
	$n^{0.5}$	0.06	0.05	0.05	0.04	0.04	0.04	0.75	0.72	0.68	0.63	0.57	0.54
	$n^{0.6}$	0.05	0.05	0.05	0.04	0.04	0.04	0.70	0.64	0.58	0.48	0.46	0.40
	$n^{0.8}$	0.05	0.04	0.04	0.05	0.05	0.05	0.68	0.54	0.45	0.33	0.25	0.22
0.75	$n^{0.2}$	0.10	0.09	0.08	0.07	0.06	0.06	0.80	0.79	0.77	0.79	0.81	0.81
	$n^{0.4}$	0.06	0.04	0.04	0.03	0.03	0.03	0.79	0.83	0.81	0.81	0.79	0.79
	$n^{0.5}$	0.04	0.03	0.03	0.03	0.02	0.02	0.82	0.81	0.76	0.71	0.68	0.68
	$n^{0.6}$	0.04	0.03	0.03	0.02	0.02	0.03	0.78	0.75	0.69	0.61	0.52	0.49
	$n^{0.8}$	0.03	0.03	0.03	0.03	0.04	0.04	0.79	0.64	0.50	0.36	0.28	0.21
1.25	$n^{0.2}$	0.10	0.08	0.07	0.05	0.05	0.04	0.77	0.77	0.81	0.81	0.83	0.84
	$n^{0.4}$	0.05	0.03	0.02	0.02	0.01	0.01	0.83	0.84	0.85	0.84	0.88	0.90
	$n^{0.5}$	0.03	0.02	0.01	0.01	0.01	0.01	0.83	0.85	0.86	0.87	0.87	0.86
	$n^{0.6}$	0.02	0.02	0.01	0.01	0.01	0.01	0.85	0.84	0.86	0.83	0.78	0.73
	$n^{0.8}$	0.02	0.01	0.01	0.02	0.02	0.03	0.83	0.82	0.74	0.57	0.41	0.41
1.49	$n^{0.2}$	0.09	0.08	0.06	0.05	0.04	0.04	0.77	0.79	0.80	0.83	0.83	0.84
	$n^{0.4}$	0.05	0.03	0.02	0.02	0.01	0.01	0.84	0.85	0.87	0.87	0.85	0.89
	$n^{0.5}$	0.03	0.02	0.01	0.01	0.01	0.01	0.86	0.86	0.88	0.87	0.88	0.87
	$n^{0.6}$	0.02	0.01	0.01	0.01	0.00	0.01	0.87	0.87	0.86	0.87	0.81	0.82
	$n^{0.8}$	0.02	0.01	0.01	0.01	0.02	0.02	0.87	0.85	0.79	0.66	0.56	0.51

Notes: The model is $y_t = \rho_t y_{t-1} + u_t$, $u_t \sim$ i.i.d., $\rho_t = 0.9a_t/\max_{0\le j\le t}|a_j|$, $a_t - a_{t-1} = v_t$ follows ARFIMA $(0, d-1, 0)$ model with parameter $d-1$.

By (A.22), $\widehat{\sigma}_{Y,t}^{-2} = O_p(H^{-1})$. Observe that $ES_{Yu,t}^2 = \sum_{j=1}^n b_{tj}^2 E[y_{j-1}^2 u_j^2]$. Below, we show that

$$E\sum_{j=1}^n b_{tj}^2 y_{j-1}^2 u_j^2 = O(H), \tag{A.6}$$

$$r_{n,t} = O_p\big((\bar{H}/n)^{\gamma} H + 1\big), \tag{A.7}$$

which yields $S_{Yu,t} = O_p(H^{1/2})$ and together with (A.5) proves (2.24): $\widehat{\rho}_{n,t} - \rho_t = O_p\big((\bar{H}/n)^{\gamma} + H^{-1/2}\big)$. To prove (A.6), note that by (2.14), $E[y_{j-1}^2 u_j^2] = E[y_{j-1}^4] + E[u_j^4] \le C$ for all j, which implies $E\sum_{j=1}^n b_{tj}^2 y_{j-1}^2 u_j^2 \le C\sum_{j=1}^n b_{tj}^2 \le CH$, by (A.1). To show (A.7), select $h := b\bar{H}$ such that (A.16) holds and bound $|r_{n,t}| \le \sum_{j=1}^n b_{tj}|\rho_{j-1} - \rho_t| y_{j-1}^2 \le R_{t,h}\sum_{|j-t|\le h} b_{tj} y_{j-1}^2 + 2\sum_{|j-t|>h} b_{tj} y_{j-1}^2$ where $R_{t,h} := \max_{j:\,|t-j|\le h}|\rho_j - \rho_t|$. This implies (A.7) because

$$R_{t,h} = O_p((\bar{H}/n)^{\gamma}), \qquad \sum_{|j-t|\le h} b_{tj} y_{j-1}^2 = O_p(H),$$
$$\sum_{|j-t|>h} b_{tj} y_{j-1}^2 = o_p(1). \tag{A.8}$$

By (A.20), $R_{t,h} = O_p((h/t)^{\gamma})$. Since $t \sim \tau n$ and $h = b\bar{H}$, this implies the first claim in (A.8). The second and third claims follow noting that $\max_j Ey_j^2 \le C$ by (2.14), while (A.1) and (A.16) imply $E\sum_{|j-t|\le h} b_{tj} y_{j-1}^2 \le CH$ and $E\sum_{|j-t|>h} b_{tj} y_{j-1}^2 \le C\sum_{|j-t|>h} b_{tj} = o(1)$.

(ii) For H such that $(\bar{H}/n)^{\gamma} = o(H^{-1/2})$, by (A.7) $r_{n,t} = o_p(H^{1/2})$, while by Lemma A.2, $\widehat{\sigma}_{Yu,t}^{-1} = O_p(H^{-1/2})$. Hence, $\widehat{\sigma}_{Yu,t}^{-1} r_{n,t} = o_p(1)$, and by (A.5),

$$\frac{\widehat{\sigma}_{Y,t}^2}{\widehat{\sigma}_{Yu,t}}(\widehat{\rho}_{n,t} - \rho_t) = \frac{S_{Yu,t}}{\widehat{\sigma}_{Yu,t}} + \frac{r_{n,t}}{\widehat{\sigma}_{Yu,t}} = \frac{S_{Yu,t}}{\widehat{\sigma}_{Yu,t}} + o_p(1).$$

Since $\widehat{\sigma}_{Yu,t}^{-1} S_{Yu,t} \to N(0,1)$ by (A.38) of Lemma A.4, this completes the proof of (2.25). □

Proof of Corollary 2.2. In view of the asymptotic normality result in (2.25), it remains to show that $\widehat{\sigma}_{Y\widehat{u},t}^2/\widehat{\sigma}_{Yu,t}^2 \to_p 1$ which follows from $|\widehat{\sigma}_{Y\widehat{u},t}^2/\widehat{\sigma}_{Yu,t}^2 - 1| = \widehat{\sigma}_{Yu,t}^{-2}|\widehat{\sigma}_{Y\widehat{u},t}^2 - \widehat{\sigma}_{Yu,t}^2| = o_p(1)$, because $\widehat{\sigma}_{Yu,t}^{-2} = O_p(H^{-1})$ by Lemma A.2, and

$$\widehat{\sigma}_{Y\widehat{u},t}^2 - \widehat{\sigma}_{Yu,t}^2 = o_p(H). \tag{A.9}$$

To verify (A.9), use $\widehat{u}_j^2 - u_j^2 = (\widehat{u}_j - u_j)^2 + 2(\widehat{u}_j - u_j)u_j$ and the Cauchy inequality, to obtain

$$|\widehat{\sigma}_{Y\widehat{u},t}^2 - \widehat{\sigma}_{Yu,t}^2| \le \sum_{j=1}^n b_{tj}^2 y_{j-1}^2 |\widehat{u}_j^2 - u_j^2| \le q_n^2 + 2q_n\widehat{\sigma}_{Yu,t},$$

where $q_n^2 := \sum_{j=1}^n b_{tj}^2 y_{j-1}^2(\widehat{u}_j - u_j)^2$ and $\widehat{\sigma}_{Yu,t}$ is as in (2.22). By Lemma A.2, $\widehat{\sigma}_{Yu,t} = O_p(H^{1/2})$. Hence, to prove (A.9), it suffices to show that

$$q_n^2 = o_p(H). \tag{A.10}$$

Notice, $|\widehat{u}_j - u_j| = |y_j - \hat{\rho}_{n,t}y_{j-1} - u_j| = |(\rho_{j-1} - \hat{\rho}_t)y_{j-1}| \le (|\rho_{j-1} - \rho_t| + |\rho_t - \hat{\rho}_{n,t}|)|y_{j-1}|$. Then $y_{j-1}^2(\widehat{u}_j - u_j)^2 \le 2y_{j-1}^4\{(\rho_{j-1} - \rho_t)^2 + (\rho_t - \hat{\rho}_{n,t})^2\}$, and

$$q_n^2 \le 2\sum_{j=1}^n b_{tj}^2 y_{j-1}^4(\rho_{j-1} - \rho_t)^2 + 2(\rho_t - \hat{\rho}_{n,t})^2\sum_{j=1}^n b_{tj}^2 y_{j-1}^4$$
$$=: 2(q_{n,1} + q_{n,2}). \tag{A.11}$$

To bound $q_{n,1}$, note that $|\rho_j| \le \rho$, $\max_j b_{tj} < \infty$ and, by (2.14), $\max_j Ey_j^4 < \infty$. Hence, by the same argument as in the proof of (A.7), it follows $q_{n,1} \le C\sum_{j=1}^n b_{tj}|\rho_t - \rho_t|y_{j-1}^4 = O_p\big((\bar{H}/n)^{\gamma}H + 1\big) = o_p(H)$.

To bound $q_{n,2}$, note that by (2.24), $\rho_t - \hat{\rho}_{n,t} = o_p(1)$, while $\sum_{j=1}^n b_{tj}^2 Ey_{j-1}^4 \le C\sum_{j=1}^n b_{tj}^2 = O(H)$, by (A.1). This implies $q_{n,2} = o_p(H)$, completing the proof of (A.10). □

Proof of Proposition 2.1. The first claim in (2.26) in shown in (A.47) of Lemma A.5. It implies the second claim noting that $E\bar{u}_t^2 = \sigma_u^2 B_{2t}^{-1} = O(H^{-1})$, by (A.1).

In addition, if H is such that $(\bar{H}/n)^{\gamma} = o(H^{-1/2})$, then (A.47) implies $(1 - \rho_t)\bar{y}_t = \bar{u}_t + o_p\big(H^{-1/2}\big)$, which together with (A.48) proves (2.27). □

Table A.3
MSE and 90% Coverage Probability (CP) results for $\widehat{\rho}_{n,t}$, $\widehat{\alpha}_{n,t}$ and $\bar{y}_t$ for the normal kernel.

H	$u_t \sim$ i.i.d.			$u_t \sim$ GARCH			$u_t \sim$ stoch. vol.		
	n								
	400	800	1000	400	800	1000	400	800	1000
MSE $\hat{\rho}_t$									
$n^{0.2}$	0.07	0.06	0.06	0.08	0.07	0.07	0.09	0.08	0.07
$n^{0.4}$	0.02	0.02	0.02	0.04	0.03	0.03	0.06	0.05	0.05
$n^{0.5}$	0.02	0.02	0.02	0.03	0.03	0.03	0.06	0.05	0.05
$n^{0.6}$	0.03	0.03	0.03	0.04	0.04	0.04	0.06	0.06	0.06
$n^{0.8}$	0.06	0.08	0.08	0.07	0.08	0.09	0.08	0.09	0.10
MSE $\hat{\alpha}_t$									
$n^{0.2}$	0.71	0.67	0.56	0.78	0.64	0.67	0.56	0.50	0.501
$n^{0.4}$	0.15	0.11	0.11	0.16	0.14	0.13	0.23	0.19	0.20
$n^{0.5}$	0.15	0.13	0.14	0.17	0.17	0.16	0.23	0.18	0.20
$n^{0.6}$	0.19	0.19	0.20	0.21	0.20	0.21	0.25	0.25	0.24
$n^{0.8}$	0.37	0.36	0.38	0.38	0.39	0.36	0.38	0.36	0.41
MSE $\bar{y}_t$									
$n^{0.2}$	0.44	0.42	0.40	0.44	0.44	0.45	0.44	0.42	0.41
$n^{0.4}$	0.37	0.32	0.32	0.33	0.32	0.31	0.33	0.29	0.32
$n^{0.5}$	0.52	0.47	0.47	0.47	0.48	0.49	0.55	0.42	0.45
$n^{0.6}$	0.80	0.78	0.82	0.72	0.69	0.80	0.72	0.86	0.79
$n^{0.8}$	1.81	1.95	1.99	2.09	2.10	1.71	1.84	1.76	2.10
CP's $\hat{\rho}_t, t = [n/2]$									
$n^{0.2}$	0.79	0.78	0.79	0.78	0.77	0.78	0.77	0.76	0.75
$n^{0.4}$	0.80	0.81	0.80	0.77	0.81	0.79	0.66	0.70	0.66
$n^{0.5}$	0.74	0.68	0.67	0.71	0.70	0.69	0.59	0.62	0.60
$n^{0.6}$	0.52	0.47	0.45	0.59	0.56	0.53	0.51	0.49	0.49
$n^{0.8}$	0.21	0.16	0.14	0.31	0.25	0.23	0.35	0.29	0.27
CP's $\hat{\alpha}_t, t = [n/2]$									
$n^{0.2}$	0.79	0.81	0.80	0.79	0.76	0.78	0.78	0.77	0.78
$n^{0.4}$	0.81	0.80	0.82	0.78	0.80	0.79	0.65	0.69	0.70
$n^{0.5}$	0.71	0.67	0.65	0.68	0.67	0.64	0.54	0.59	0.59
$n^{0.6}$	0.46	0.43	0.41	0.52	0.47	0.43	0.46	0.42	0.42
$n^{0.8}$	0.18	0.15	0.14	0.19	0.16	0.14	0.20	0.18	0.19
CP's $\bar{y}_t, t = [n/2]$									
$n^{0.2}$	0.74	0.77	0.80	0.79	0.81	0.80	0.79	0.79	0.82
$n^{0.4}$	0.79	0.80	0.79	0.77	0.80	0.80	0.75	0.77	0.77
$n^{0.5}$	0.69	0.72	0.66	0.69	0.65	0.66	0.71	0.66	0.64
$n^{0.6}$	0.53	0.53	0.49	0.55	0.52	0.48	0.57	0.50	0.50
$n^{0.8}$	0.33	0.27	0.25	0.36	0.24	0.25	0.33	0.26	0.26

Notes: The model is $y_t = \alpha_t + \rho_t y_{t-1} + u_t$ with u_t i.i.d, GARCH(1,1) as in (a) or stochastic volatility noise as in (b) normalized to have unit variance; $\alpha_t = t^{-1/2} z_t$, where z_t is a random walk, $z_t - z_{t-1} \sim$ i.i.d.; and $\rho_t = 0.9 a_t / \max_{0 \le j \le t} |a_j|$, where a_t is a random walk, $a_t - a_{t-1} \sim$ i.i.d..

Proof of Theorem 2.4. By definition, $y'_j = y_j - \mu_j, j \ge 1$ follows the AR(1) model (2.1). First, we prove the claims about $\bar{y}_t$. Note that

$$\bar{y}_t - \mu_t = B_{1t}^{-1} \sum_{j=1}^n b_{tj}(y_j - \mu_t) = B_{1t}^{-1} \sum_{j=1}^n b_{tj}(\mu_j - \mu_t) + B_{1t}^{-1} \sum_{j=1}^n b_{tj} y'_j =: r_{1t} + \bar{y}'_t.$$

By (A.1), $B_{1t} \sim H$, which together with (A.52) of Lemma A.6 implies $r_{1t} = O_p\big((\bar{H}/n)^\beta + H^{-1}\big)$. Since by (2.26), $\bar{y}'_t = O_p\big((\bar{H}/n)^\gamma + H^{-1/2}\big)$, and $\beta \ge \gamma$, this proves (2.31). In addition, if $(\bar{H}/n)^\gamma = o(H^{-1/2})$, then $r_{1t} = o_p(H^{-1/2})$, so that $\bar{y}_t - \mu_t = \bar{y}'_t + o_p(H^{-1/2})$, which together with (2.27) proves (2.32).

To prove the claims about $\widehat{\rho}_{n,t}$, set $\hat{y}_j := y_j - \bar{y}_t$, $S_{\widehat{Y}\widehat{Y},t} := \sum_{j=1}^n b_{tj} \hat{y}_j \hat{y}_{j-1}$, $S_{Y'Y',t} := \sum_{j=1}^n b_{tj} y'_j y'_{j-1}$, $\hat{\sigma}^2_{\hat{Y},t} := \sum_{j=1}^n b_{tj} \hat{y}^2_{j-1}$ and $\hat{\sigma}^2_{Y',t} := \sum_{j=1}^n b_{tj} y'^2_{j-1}$. By definition (2.29), $\widehat{\rho}_{n,t} = S_{\hat{Y}\hat{Y},t}/\widehat{\sigma}^2_{\hat{Y},t}$. Observe that $\widetilde{\rho}_{n,t} := S_{Y'Y',t}/\widehat{\sigma}^2_{Y',t}$ is the estimator of the parameter ρ_t of an AR(1) model with no intercept.

Hence, by (A.53) of Lemma A.6, with $\xi_n := O_p\big((\bar{H}/n)^\gamma H + 1\big)$,

$$\begin{aligned} |\widehat{\rho}_{n,t} - \widetilde{\rho}_{n,t}| &= \left| \frac{S_{\widehat{Y}\widehat{Y},t}}{\widehat{\sigma}^2_{\hat{Y},t}} - \frac{S_{Y'Y',t}}{\widehat{\sigma}^2_{Y',t}} \right| = \left| \frac{S_{Y'Y',t} + \xi_n}{\widehat{\sigma}^2_{Y',t} + \xi_n} - \frac{S_{Y'Y',t}}{\widehat{\sigma}^2_{Y',t}} \right| \\ &\le \frac{|\xi_n|\{|S_{Y'Y',t}/\widehat{\sigma}^2_{Y',t}| + 1\}}{\widehat{\sigma}^2_{Y',t} + \xi_n} \\ &= O_p(|\xi_n| H^{-1}) = \delta_n, \end{aligned} \tag{A.12}$$

where $\delta_n = O_p\big((\bar{H}/n)^\gamma + H^{-1}\big)$, since $|\widetilde{\rho}_{n,t}| = |S_{Y'Y',t}/\widehat{\sigma}^2_{Y',t}| = O_p(1)$ by (2.24), and $\widehat{\sigma}^2_{Y',t} \ge cH$ for some $c > 0$ by (A.22).

Hence, $\hat{\rho}_{n,t} = \widetilde{\rho}_{n,t} + \delta_n$. Since $\widetilde{\rho}_{n,t}$ satisfies (2.24), this together with Corollary 2.1 implies (2.31), while for $(\bar{H}/n)^\gamma = o(H^{-1/2})$, $\widetilde{\rho}_{n,t}$ satisfies (2.25) which implies asymptotic normality (2.32).

Finally, we discuss the estimate $\widehat{\alpha}_{n,t} = \bar{y}_t(1 - \widehat{\rho}_{n,t})$ of $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$. Note that

$$\bar{y}_t - \mu_t = \bar{y}'_t + B_{1t}^{-1} \sum_{j=1}^n b_{tj}(\mu_j - \mu_t) = \bar{u}_t(1 - \rho_t)^{-1} + \delta_n,$$

in view of (2.26), (A.52), and assumption $\beta \geq \gamma$. Moreover, by (A.12), $\widehat{\rho}_{n,t} - \widetilde{\rho}_{n,t} = \delta_n$, while by (A.48), $\bar{u}_t = O_p(H^{-1/2})$, and, by (2.32), $\widehat{\rho}_{n,t} - \rho_t = O_p(H^{-1/2}) + \delta_n$. Hence,

$$\widehat{\alpha}_{n,t} = \Big(\mu_t + \bar{u}_t(1-\rho_t)^{-1} + \delta_n\Big)(\{1-\rho_t\} + \{\rho_t - \widehat{\rho}_{n,t}\})$$
$$= \mu_t(1-\rho_t) + \bar{u}_t + \mu_t(\rho_t - \widehat{\rho}_{n,t}) + \delta_n$$
$$= \mu_t(1-\rho_t) + \bar{u}_t + \mu_t(\rho_t - \widetilde{\rho}_{n,t}) + \delta_n. \quad (A.13)$$

Assumption 2.3 and (A.20) straightforwardly imply that $\alpha_t - \mu_t(1-\rho_t) = \mu_t\rho_t - \mu_{t-1}\rho_{t-1} = \delta_n$. Hence $\widehat{\alpha}_{n,t} - \alpha_t = Z_{n,t} + \delta_n$, where $Z_{n,t} = \bar{u}_t + \mu_t(\rho_t - \widetilde{\rho}_{n,t})$. Clearly, the above bounds imply $Z_{n,t} = O_p(H^{-1/2}) + \delta_n$, which proves (2.31). In addition, if $(\bar{H}/n)^\gamma = o(H^{-1/2})$, then $\delta_n = o_p(H^{-1/2})$. Since $(B_{1t}/B_{3t})Z_{n,t} \to_D N(0,1)$ by (A.39), and $B_{1t}/B_{3t} = O_p(H^{1/2})$ by (A.1) and (A.23), this implies (2.32) for $\widehat{\alpha}_{n,t}$ and completes the proof. □

Proof of Corollary 2.3. To prove the first claim, use Lemmas A.2 and A.6 to obtain $\widehat{\sigma}^2_{\hat{Y},t} = \widehat{\sigma}^2_{Y',t} + o_p(H) = Hv^2_{1,t} + o_p(H)$, where $v^2_{1,t} = \sigma^2_u(1-\rho^2_t)^{-1}$. This together with $\widehat{\rho}_{n,t} - \rho_t = o_p(1)$ of (2.31) and $B^{1/2}_{1t} \sim H^{1/2}$ of (A.1) implies

$$\left|\frac{1-\hat{\rho}_{n,t}}{1+\hat{\rho}_{n,t}}\right|^{1/2} \frac{B^{3/2}_{1t}}{B_{2t}\widehat{\sigma}_{\hat{Y},t}} = \left|\frac{1-\rho_t}{1+\rho_t}\right|^{1/2} \frac{B_{1t}(1-\rho^2_t)^{1/2}}{B_{2t}\sigma_u}(1+o_p(1))$$
$$= \frac{B_{1t}}{B_{2t}}\frac{(1-\rho_t)}{\sigma_u}(1+o_p(1)),$$

which, in view of asymptotic normality (2.32), proves the first claim in (2.33).

To prove the second and third claims in (2.33), in view of (2.32), it suffices to show that

$$(\widehat{\sigma}^2_{\hat{Y},t}/\widehat{\sigma}_{\hat{Y}\hat{u},t})/(\widehat{\sigma}^2_{Y',t}/\widehat{\sigma}_{Y'u,t}) \to_p 1, \qquad \hat{B}^2_{3t}/B^2_{3t} \to_p 1. \quad (A.14)$$

To check the first claim, note that by Lemma A.6, $\widehat{\sigma}^2_{\hat{Y},t} = \widehat{\sigma}^2_{Y',t} + o_p(H)$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u},t} = \widehat{\sigma}^2_{Y'u,t} + o_p(H)$, while by (A.22) $\widehat{\sigma}^{-2}_{Y',t} = O_p(H^{-1})$ and $\widehat{\sigma}^{-2}_{Y'u,t} = O_p(H^{-1})$. Therefore, $\widehat{\sigma}^2_{\hat{Y},t}/\widehat{\sigma}^2_{Y',t} \to_p 1$ and $\widehat{\sigma}_{\hat{Y}\hat{u},t}/\widehat{\sigma}_{Y'u,t} \to_p 1$, which implies the first result of (A.14).

To prove the second result, note that $|\widehat{B}^2_{3t}/B^2_{3t} - 1| = B^{-2}_{3t}|\widehat{B}^2_{3t} - B^2_{3t}| = o_p(1)$, since $B^{-2}_{3t} = O_p(H^{-1})$ by Lemma A.2, and as shown below, $\widehat{B}^2_{3t} - B^2_{3t} = o_p(H)$. Indeed, setting $\widehat{\sigma}^2_{Y'uu,t} := \sum_{j=1}^n b^2_{tj}y'_{j-1}u^2_j$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u},t} := \sum_{j=1}^n b^2_{tj}\hat{y}_{j-1}\hat{u}^2_j$, from definition of B_{3t} and $\widehat{B}_{3t}$ we obtain

$$B^2_{3t} - \widehat{B}^2_{3t} = -2\mu_t\frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}}\widehat{\sigma}^2_{Y'uu,t} + 2\bar{y}_t\frac{B_{1t}}{\widehat{\sigma}^2_{\hat{Y},t}}\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u},t}$$
$$+ \mu^2_t\frac{B^2_{1t}}{\widehat{\sigma}^4_{Y',t}}\widehat{\sigma}^2_{Y'u,t} - \bar{y}^2_t\frac{B^2_{1t}}{\widehat{\sigma}^4_{\hat{Y},t}}\widehat{\sigma}^2_{\hat{Y}\hat{u},t}. \quad (A.15)$$

Combining the above properties of $\widehat{\sigma}^2_{Y',t}$, $\widehat{\sigma}^2_{\hat{Y},t}$, $\widehat{\sigma}^2_{Y'u,t}$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u},t}$ with $\bar{y}_t = \mu_t + o_p(1)$ of (2.31) and $\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u},t} = \widehat{\sigma}^2_{Y'uu,t} + o_p(H) = O_p(H)$ of Lemma A.6, and noting that $E\mu^2_t < \infty$ implies $\mu_t = O_p(1)$, we obtain

$$\bar{y}_t\frac{B_{1t}}{\widehat{\sigma}^2_{\hat{Y},t}}\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u}} = \mu_t\frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}}\widehat{\sigma}^2_{Y'uu,t} + o_p(H),$$
$$\bar{y}^2_t\frac{B^2_{1t}}{\widehat{\sigma}^4_{\hat{Y},t}}\widehat{\sigma}^2_{\hat{Y}\hat{u}} = \mu^2_t\frac{B^2_{1t}}{\widehat{\sigma}^4_{Y',t}}\widehat{\sigma}^2_{Y'u,t} + o_p(H),$$

which yields $B^2_{3t} - \widehat{B}^2_{3t} = o_p(H)$, completing the proof of (A.14). □

Proof of Proposition 2.2. Let $1 \leq k \leq h < t/2$. Write, by (2.35),

$$\mu_{t+k} - \mu_t = \sum_{i=0}^{t+k-1} c_{t+k,i}\alpha_{t+k-i} - \sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i} = m(t,k) + \tilde{m}(t,k),$$

where $\tilde{m}(t,k) := \sum_{i=0}^{t-1}(c_{t+j,i} - c_{t,i})\alpha_{t-i}$, and $m(t,k) = \mu_{t+k} - \mu_t - \tilde{m}(t,k)$. To verify Assumption 2.3(ii), it remains to show that $Em^2(t,k) \leq C|k/t|^{2\gamma}$, $1 \leq k \leq h$, and $\max_{1\leq k\leq h}|\tilde{m}(t,k)| \leq C\big((h/t)^\gamma + h^{-1}\big)$. By Assumption 2.4, $\max_i E|\alpha_i| < \infty$ and $E|\alpha_{t+k-i} - a_{t-i}| \leq C(k/(t-i))^\beta$ for $i \leq t/2$. Since $|c_{t,i}| \leq \rho^i$, then

$$E|m(t,k)| \leq CE\left(\sum_{i=t}^{t+k-1}|c_{t+k,i}||\alpha_{t+k-i}| + \sum_{i=1}^{t-1}|c_{t+k,i}||\alpha_{t+k-i} - a_{t-i}|\right)$$
$$\leq C\left(\sum_{i=t/2}^{\infty}\rho^i + \sum_{i=1}^{t/2}\rho^i(k/(t-i))^\beta\right)$$
$$\leq C(\rho^{[t/2]} + |k/t|^\beta) \leq C|k/t|^\beta.$$

On the other hand, $\tilde{m}(t,j) = \sum_{i=1}^h[\cdots] + \sum_{i=h+1}^{t-1}[\cdots] =: s_{nj} + r_{nj}$. By (A.17), for $i = 1,\ldots,h$, $|c_{t+j,i} - c^i_t| \leq 3i\rho^{i-1}R_{t,h}$, where $R_{t,h} = \max_{j:|j-t|\leq h}|\rho_t - \rho_j| = O_p((h/t)^\gamma)$ by (A.20). Hence, $|s_{nj}| \leq CR_{t,h}\sum_{i=1}^h\rho^{i-1}i|\alpha_{t-i}| = O_p((h/t)^\gamma)$, because $E\sum_{i=1}^h\rho^{i-1}i|\alpha_{t-i}| \leq C\sum_{i=1}^\infty\rho^{i-1}i < \infty$. Finally, $|r_{nj}| \leq 2\sum_{i=h+1}^{t-1}\rho^i|\alpha_{t-i}|$ for all $j \leq h$. Since $E\sum_{i=h+1}^{t-1}\rho^i|\alpha_{t-i}| \leq C\rho^h = O(h^{-1})$, this implies $\max_{j=1,\ldots,h}|r_{nj}| = O_p(h^{-1})$, which completes the verification of Assumption 2.3(ii) with $\beta = \gamma$.

Proof of (2.36). By (2.35), $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$. Thus, $|\mu_t - (1-\rho_t)^{-1}\alpha_t| \leq (1-\rho_t)^{-1}|(1-\rho_t)\mu_t - \alpha_t| \leq C|\rho_t\mu_t - \rho_{t-1}\mu_{t-1}| \leq C(|\rho_t - \rho_{t-1}||\mu_t| + \rho|\mu_t - \mu_{t-1}|) = o_p(1)$, because $\rho_t - \rho_{t-1} = o_p(1)$ by (A.20) and $\mu_t - \mu_{t-1} = (\mu_t - \mu_{t+h}) + (\mu_{t+h} - \mu_{t-1}) = o_p(1)$ by Assumption 2.3(ii) verified for μ_t above. This proves (2.36) and completes the proof of proposition. □

A.2. Auxiliary results

This section contains auxiliary results used to prove the main theorems.

Recall notation: $\bar{H} := H$ when K has finite support, and $\bar{H} = H\log^{1/2}H$ when K has infinite support. We will repeatedly use the following property of the kernel weights b_{tj}: there exists $b > 0$ such that, as $H \to \infty$,

$$\sum_{1\leq j\leq n,\ |t-j|\geq b\bar{H}} b_{tj} = o(1). \quad (A.16)$$

Indeed, if kernel K has finite support, then $K(x) = 0$, $|x| \geq x_0$ for some finite $x_0 > 0$ and (A.16) holds with $b > x_0$. If K has infinite support, use b such that $b^2c \geq 1$, where c is the same as in $K(x) = O(e^{-cx^2})$ in (2.21). Then, $\sum_{|t-j|\geq b\bar{H}} b_{tj} \leq C\int_{b\bar{H}}^\infty e^{-c(x/H)^2}dx \leq (b\bar{H})^{-1}\int_{b\bar{H}}^\infty e^{-c(x/H)^2}xdx \leq CH\bar{H}^{-1} = o(1)$.

We shall use the notation $R_{t,h} := \max_{j:|j-t|\leq h}|\rho_t - \rho_j|$. In the following lemma, y_t, u_t and $c_{t,j}$ are as in Theorem 2.1.

Lemma A.1. (i) *For $t \geq 1$, and $1 \leq j, s \leq t-1$,*

$$|c_{t,j} - \rho^j_t| \leq \rho^{j-1}j\max_{k=1,\ldots,j}|\rho_t - \rho_{t-k}|,$$
$$|\rho^j_t - \rho^j_s| \leq \rho^{j-1}j|\rho_t - \rho_{ts}|. \quad (A.17)$$

(ii) *For* $1 \le t_0 \le t, j \ge 1$,

$$|\rho_{t+j} - \rho_t| \le 2\rho |a_{t_0}|^{-1} \max_{k=1,\dots,j} |a_{t+k} - a_t|. \tag{A.18}$$

(iii) *For* $t \ge 1$, $1 \le h < t/2$ *and* j *such that* $|t - j| \le h$,

$$|y_j - z_j(\rho_t)| \le (R_{t,2h} + h^{-1}) Y_{j,h}, \quad EY^4_{j,h} \le C, \tag{A.19}$$

where random variables $Y_{j,h}$ *do not depend on* t, *and* C *does not depend on* t, h *and* j.

(iv) *Under Assumptions* 2.1 *and* 2.2, *as* $t \to \infty$, $h \to \infty$ *and* $h = o(t)$,

$$R_{t,h} = O_p\big((h/t)^\gamma\big), \quad \max_{j:|t-j|\le h} E(y_j - z_j(\rho_t))^2 = o(1). \tag{A.20}$$

Proof. (i) Notice that $|a_1 \cdots a_j - b_1 \cdots b_j| = |(a_1 - b_1)a_2 \cdots a_k + b_1(a_2 - b_2)a_3 \cdots a_j + b_1 \cdots b_{j-1}(a_j - b_j)| \le j \max_{i=1,\dots,j} |a_i - b_i| a^{j-1}$, if $|a_i| \le a$ and $|b_i| \le a$. Thus, for $1 \le j \le t-1$,

$$|c_{t,j} - \rho_t^j| = |\rho_{t-1} \dots \rho_{t-j} - \rho_t^j| \le j\rho^{j-1} \max_{i=1,\dots,j} |\rho_t - \rho_{t-i}|,$$

which proves the first claim of (A.17), while the second claim follows by the same argument.

(ii) To prove (A.18), denote $m_t := \max_{0 \le s \le t} |a_s|$. Then, $m_{t+j} \ge m_t \ge |a_{t_0}|$, $m_t \ge |a_t|$, and

$$\begin{aligned} |\rho_{t+j} - \rho_t| &= \rho \left| \frac{a_{t+j}}{m_{t+j}} - \frac{a_t}{m_t} \right| \\ &\le \rho \left(\frac{|a_{t+j} - a_t|}{m_{t+j}} + |a_t| \frac{|m_{t+j} - m_t|}{m_{t+j} m_t} \right) \\ &\le \rho |a_{t_0}|^{-1} (|a_{t+j} - a_t| + |m_{t+j} - m_t|). \end{aligned}$$

We show that

$$|m_{t+j} - m_t| \le \max_{k=1,\dots,j} |a_{t+k} - a_t|, \tag{A.21}$$

which completes the proof of (A.18). Let $m_{t+j} = |a_{j_*}|$. If $j_* \le t$, then $m_{t+j} - m_t = 0$ and (A.21) holds. If $t < j_* \le t + j$, then $m_{t+j} \ge m_t \ge |a_t|$, and $m_{t+j} - m_t \le |a_{j_*}| - |a_t|$, and (A.21) holds.

(iii) *Proof of* (A.19). By (2.13) and (2.15), $|y_j - z_j(\rho_t)| = \big|\sum_{k=1}^{j-1}(c_{j,k} - \rho_t^k)u_{j-k} + c_{j,j}y_0\big|$, where by (2.12) and (2.3), $|c_{j,k}| \le \rho^k$ and $|\rho_t| \le \rho$. For $|j - t| \le h$, by (A.17), $|c_{j,k} - \rho_t^k| \le |c_{j,k} - \rho_j^k| + |\rho_j^k - \rho_t^k| \le \rho^{k-1} k \{\max_{i=1,\dots,k} |\rho_j - \rho_{j-i}| + |\rho_j - \rho_t|\} \le 3\rho^{k-1} k R_{t,2h}$, while $|c_{j,j}| \le \rho^j \le \rho^h \le (h|\log \rho|)^{-1}$, since $\rho < 1$ and $j \ge h$. Hence,

$$\begin{aligned} |y_j - z_j(\rho_t)| &= \sum_{k=1}^{h-1} |c_{j,k} - \rho_t^k||u_{j-k}| + 2\sum_{k=h}^{j-1} \rho^k |u_{j-k}| + \rho^h |y_0| \\ &\le (R_{t,2h} + h^{-1}) C Y_{j,h}, \end{aligned}$$

where $Y_{j,h} = \sum_{k=1}^{h-1} 3\rho^{k-1} k |u_{j-k}| + 2\sum_{k=h}^{j-1} \rho^{k-h} |u_{j-k}| + |y_0|$. Since $0 < \rho < 1$, $Eu_1^4 < \infty$ and $Ey_0^4 < \infty$, then by the same argument as in the proof of (2.14), $EY^4_{j,h} \le C\big((3\sum_{k=1}^{\infty} \rho^{k-1} k)^4 + (2\sum_{k=0}^{\infty} \rho^k)^4 + Ey_0^4\big) < \infty$. This proves (A.19).

(iv) To prove the first claim in (A.20), let $t_0 := t - h$. Then by $|\rho_t - \rho_j| \le |\rho_t - \rho_{t_0}| + |\rho_j - \rho_{t_0}|$ and (A.18),

$$\begin{aligned} R_{t,h} &\le 2 \max_{k=1,\dots,2h} |\rho_{t_0+k} - \rho_{t_0}| \le 4\rho |a_{t_0}|^{-1} i_n, \\ i_n &:= \max_{k=1,\dots,2h} |a_{t_0+k} - a_{t_0}|. \end{aligned}$$

We will show that $|a_{t_0}|^{-1} = O_p(t^{-\gamma})$ and $i_n = O_p(h^\gamma)$ which implies $R_{t,h} = O_p((h/t)^\gamma)$. Firstly, $|a_{t_0}|^{-1} = O_p(t^{-\gamma})$ holds since by Assumption 2.2 $t_0^{-\gamma} a_{t_0} \to_D W_1 + g(1)$ as $t_0 = t - h \to \infty$, where W_1 has continuous distribution. To bound i_n, let $a'_j := a_j - Ea_j$ and $\tilde{i}_n := \max_{k=1,\dots,2h} |a'_{t_0+k} - a'_{t_0}|$. Then $i_n \le \tilde{i}_n + \max_{k=1,\dots,2h} |Ea_{t_0+k} - Ea_{t_0}| \le \tilde{i}_n + Ch^\gamma$, because $|Ea_{t_0+k} - Ea_{t_0}| \le Ck^\gamma$ by (2.5). The stationarity of v_i's in Assumption 2.1(ii) implies that $\tilde{i}_n =_D \max_{k=1,\dots,2h} |a'_k - a'_0|$. Thus, by the weak convergence (2.5), as $h \to \infty$,

$$(2h)^{-\gamma} \tilde{i}_n =_D \sup_{0 \le \tau \le 1} (2h)^{-\gamma} |a'_{[2\tau h]} - a'_0| \to_D \sup_{0 \le \tau \le 1} |W_\tau| = O_p(1).$$

Hence, $\tilde{i}_n = O(h^\gamma)$, $i_n = O_p(h^\gamma)$ which completes the proof of the first claim of (A.20).

To show the second claim in (A.20), use (A.19) and the Cauchy inequality, to obtain for $|j - t| \le h$, $E(y_j - z_j(\rho_t))^2 \le E[(R_{t,2h} + h^{-1})^2 Y^2_{j,h}] \le C(E(R_{t,2h} + h^{-1})^4)^{1/2}$, where $C < \infty$ does not depend on t, j, h. Because $h^{-1} \to 0$, and $|\rho_t| \le \rho$ implies $R_{t,2h} \le 2\rho$, it remains to show that $ER_{t,2h} \to 0$. Note that for any $\epsilon > 0$, $ER_{t,2h} \le \epsilon P(|R_{t,2h}| < \epsilon) + 2P(|R_{t,2h}| \ge \epsilon) \le 2\epsilon$, as $t \to \infty$, because by the first part of (A.20), $R_{t,2h} = o_p(1)$. This completes the proof of (A.20) and the lemma. □

Denote $v^2_{1,t} = \sigma_u^2 (1 - \rho_t^2)^{-1}$, $v^2_{2,t} = \beta_K E[U_0^2(a) u_1^2]|_{a=\rho_t}$ and $v^2_{3,t} := \beta_K E[(1 - \delta U_0(a))^2 u_1^2]|_{a=\rho_t,\, \delta=\delta_t}$, where $U_0(a) = \sum_{k=0}^{\infty} a^k u_{-k}$ and $\delta_t := \mu_t (1 - \rho_t^2)/\sigma_u^2$. Observe that $v^2_{2,t} = \beta_K \sum_{i,s=0}^{\infty} \rho_t^{i+s} E[u_{-i} u_{-s} u_1^2]$. Let B_{3t} be as in (2.32) of Theorem 2.4. In the lemma below, $c_1, c_2 > 0$ denote some positive finite constants.

Lemma A.2. *The following holds for* $\widehat{\sigma}^2_{Y,t}$ *in Theorem* 2.3(i), *for* $\widehat{\sigma}^2_{Yu,t}$ *in Theorem* 2.3(ii) *and for* $B^2_{3,t}$ *in Theorem* 2.4(ii):

$$\widehat{\sigma}^2_{Y,t}/(Hv^2_{1,t}) \to_p 1, \quad \widehat{\sigma}^2_{Yu,t}/(Hv^2_{2,t}) \to_p 1, \tag{A.22}$$

$$B^2_{3,t}/(Hv^2_{3,t}) \to_p 1, \tag{A.23}$$

where $c_1 \le v^2_{s,t} \le c_2$, $s = 1, 2$ *and* $c_1 \le v^2_{3,t} = O_p(1)$.

Proof. First we verify the claims about $v^2_{1,t}$, $v^2_{2,t}$ and $v^2_{3,t}$. The claim $c_1 \le v^2_{1,t} \le c_2$ holds, because $|\rho_t| \le \rho$ implies $\sigma_u^2 \le v^2_{1,t} \le \sigma_u^2 (1 - \rho^2)^{-1}$. The claim $v^2_{2,t} \le c_2$ follows by applying in $v^2_{2,t} = \beta_K E[U_0^2(a) u_1^2]|_{a=\rho_t}$ the bound $E[U_0^2(a) u_1^2] \le E[U_0^4(a)] + Eu_1^4 \le (1 - \rho)^{-4} Eu_1^4 + Eu_1^4 < \infty$, valid for $|a| \le \rho$. The claim $v^2_{2,t} \ge c_1$ follows by noting that by (2.23), $V_1 := E[u_1^2 | u_0, \cdots] \ge c > 0$, which implies $E[U_0^2(a) u_1^2] = E[U_0^2(a) V_1] \ge cE[U_0^2(a)] \ge c\sigma_u^2 > 0$.

To verify the claim that $v^2_{3,t} = O_p(1)$, use the bound $E[(1 - \delta U_0(a))^2 u_1^2] \le 4(1 + \delta^4 E[U_0^4(a)] + Eu_1^4) \le 4(1 + \delta^4 (1 - \rho)^{-4} Eu_1^4 + Eu_1^4) < \infty$ for $0 \le a \le \rho$, and note that $\delta_t = O_p(1)$, because $E|\delta_t| \le E|\mu_t| 2\sigma_u^{-2} < \infty$. To show that $v^2_{3,t} \ge c_1$, note that $V_1 = E[u_1^2 | u_0, \dots] \ge c > 0$ implies $E[(1 - \delta U_0(a))^2 u_1^2] = E[(1 - \delta U_0(a))^2 V_1] \ge cE[(1 - \delta U_0(a))^2] = c + \delta^2 EU_0^2(a) \ge c$.

Proof of (A.22). We verify the second claim, (the first claim can be verified similarly). Set $t_0 = t - 2h$ where $h = b\bar{H}$ with b as in (A.16). We shall approximate $\widehat{\sigma}^2_{Yu,t}$ by

$$\widehat{\sigma}^{2(apr)}_{Yu,t} := \sum_{j:|t-j|\le h} b^2_{tj} z^2_{t_0,j-1} u_j^2, \quad z_{t_0,j} := \sum_{k=0}^{h} \rho_{t_0}^k u_{j-k}. \tag{A.24}$$

We show below that

$$\widehat{\sigma}^2_{Yu,t} - \widehat{\sigma}^{2(apr)}_{Yu,t} = o_p(H), \tag{A.25}$$

$$\widehat{\sigma}^{2(apr)}_{Yu,t}/(Hv^2_{2,t_0}) \to_p 1, \quad v^2_{2,t_0}/v^2_{2,t} \to_p 1. \tag{A.26}$$

Together with $v_{2,t} \ge c_1 > 0$ this implies (A.22): $(Hv^2_{2,t})^{-1} \widehat{\sigma}^2_{Yu,t} = (Hv^2_{2,t})^{-1} (\widehat{\sigma}^2_{Yu,t} - \widehat{\sigma}^{2(apr)}_{Yu,t}) + (Hv^2_{2,t})^{-1} \widehat{\sigma}^{2(apr)}_{Yu,t} \to_p 1$.

Proof of (A.25). The claim follows from the bound $|\widehat{\sigma}^2_{Yu,t} - \widehat{\sigma}^{2(apr)}_{Yu,t}| \le \sum_{j:|t-j|\ge h} b_{tj}^2 y_{j-1}^2 u_j^2 + \sum_{j:|t-j|<h} b_{tj}^2 |y_{j-1}^2 - z_{t_0,j-1}^2| u_j^2$, observing that

$$E \sum_{j:|t-j|\ge h} b_{tj}^2 y_{j-1}^2 u_j^2 = o_p(1), \tag{A.27}$$

$$E \sum_{j:|t-j|<h} b_{tj}^2 |y_{j-1}^2 - z_{t_0,j-1}^2| u_j^2 = o(H). \tag{A.28}$$

The result (A.27) follows from (A.1) since $E[y_{j-1}^2 u_j^2] \le Ey_{j-1}^4 + Eu_j^4 \le C$ for all j by (2.14). To prove (A.28), use $\sum_{j=1}^n b_{tj}^2 = O(H)$ of (A.1) and

$$\max_{j:\,|j-t|<h} E|y_{j-1}^2 - z_{t_0,j-1}^2| u_j^2 = o(1). \tag{A.29}$$

In turn, to verify (A.29), first we show that

$$\max_{j:|j-t|<h} E(y_{j-1}^2 - z_{t_0,j-1}^2)^2 \to 0. \tag{A.30}$$

By the same argument as in the proof of (A.19) it follows that

$$|y_j - z_{t_0,j}| \le (R_{t,2h} + h^{-1}) Y_{j,h}, \quad EY_{j,h}^2 \le C, \quad |j-t| \le h \tag{A.31}$$

where random variables $Y_{j,h}$ do not depend on t, and C does not depend on t, h and j. Thus, (A.30) follows by the same argument as in the proof of the second claim in (A.20). Now, to show (A.29), set $u_{j,1} := u_j I(|u_j| \ge L)$, $L > 0$. Then $|y_{j-1}^2 - z_{t_0,j-1}^2| u_j^2 \le |y_{j-1}^2 - z_{t_0,j-1}^2|(u_{j,1}^2 + L^2)$. By the Cauchy inequality and stationarity of u_t, $E|y_{j-1}^2 - z_{t_0,j-1}^2| u_{j,1}^2 \le (E(y_{j-1}^2 - z_{t_0,j-1}^2)^2)^{1/2}(Eu_{j,1}^4)^{1/2} \le C(Eu_{1,1}^4)^{1/2} \to 0$, as $L \to \infty$ uniformly in j, because $Ey_j^4 \le C$ and $Ez_{t_0,j}^4 \le C$ for all j, h and t, see (2.14). On the other hand, $E|y_{j-1}^2 - z_{t_0,j-1}^2| = E|(y_{j-1} - z_{t_0,j-1})^2 + 2z_{t_0,j-1}(y_{j-1} - z_{t_0,j-1})| \le E(y_{j-1} - z_{t_0,j-1})^2 + 2(E(y_{j-1} - z_{t_0,j-1})^2)^{1/2}(Ez_{t_0,j-1}^2)^{1/2} \to 0$ uniformly in $|j-t| \le h$ because of (A.30). This proves (A.29).

Proof of (A.26). Let $e_{t_0}^2 := \sum_{i,s=0}^h \rho_{t_0}^{i+s} \sum_{j:|t-j|\le h} b_{tj}^2 E[u_{j-1-i}u_{j-1-s} u_j^2]$. To prove the first result of (A.26), it suffices to show

$$e_{t_0}^2 = Hv_{2,t_0}^2(1 + o_p(1)), \quad \widehat{\sigma}^{2(apr)}_{Yu,t} - e_{t_0}^2 = o_p(H). \tag{A.32}$$

First, as $h \to \infty$, $\sum_{j:|t-j|\le h} b_{tj}^2 \sim H\beta_K$, by (A.16) and (A.1), while $|\sum_{s=h+1}^\infty \rho_{t_0}^{2s}| \le \sum_{s=h+1}^\infty \rho^{2s} \to 0$, and by stationarity $E[u_{j-1-i} u_{j-1-s} u_j^2] = E[u_{-i}u_{-s}u_1^2]$. Therefore,

$$e_{t_0}^2 = \sum_{i,s=0}^\infty \rho_{t_0}^{i+s} E[u_{-i}u_{-s}u_1^2] H\beta_K(1 + o_p(1)) = Hv_{2,t_0}^2(1 + o_p(1)).$$

To prove the second claim in (A.32), bound

$$H^{-1}|\widehat{\sigma}^{2(apr)}_{Yu,t} - e_{t_0}^2| = \Big|\sum_{i,s=0}^h \rho_{t_0}^{i+s} T_{n,is}\Big| \le \sum_{i,s=0}^h \rho^{i+s}|T_{n,is}|,$$

$$T_{n,is} := H^{-1} \sum_{j:|t-j|\le h} b_{tj}^2 x_j,$$

where $x_j := u_{j-1-i}u_{j-1-s}u_j^2 - E[u_{j-1-i}u_{j-1-s}u_j^2]$. We will show that

$$\max_{i,s} E|T_{n,is}| \le C; \quad T_{n,is} \to_p 0, \quad \forall s, t, \tag{A.33}$$

which implies $H^{-1}|\widehat{\sigma}^{2(apr)}_{Yu,t} - e_{t_0}^2| \to_p 0$, since then for any fixed L, $\sum_{i,s=0}^L \rho^{i+s}|T_{n,is}| \to_p 0$, while $E\sum_{\max(i,s)>L}^h \rho^{i+s}|T_{n,is}| \le C\sum_{i=L}^\infty \rho^i \to 0$, as $L \to \infty$.

The first claim in (A.33) is valid because $E|T_{n,is}| \le H^{-1}\sum_{j:|t-j|\le h} b_{tj}^2 2Eu_1^4 \le C$ by (A.1). To verify $T_{n,is} \to_p 0$ for fixed i, s we shall use Lemma A.3. Write $T_{n,is} = \sum_{j=1}^n z_{nj}x_j$, where $z_{nj} := H^{-1}b_{tj}^2 I(|j-t| \le h)$. Note that x_j is a stationary ergodic process, because u_j is stationary ergodic, see Theorem 3.5.8 in Stout (1974). In addition, $Ex_1 = 0$, $E|x_1| < \infty$ and $T_{n,is} = 0$. Moreover, the z_{nj}'s satisfy the assumptions of Lemma A.3 with $\nu_n = 1$. Indeed, $\sum_{j=1}^n |z_{nj}| \le H^{-1}\sum_{j=1}^n b_{tj}^2 = O(1)$, by (A.1), while $\sum_{j=1}^n |z_{nj} - z_{n,j-1}| \le H^{-1}(\sum_{j:|t-j|\le h} |b_{nj}^2 - b_{n,j-1}^2| + b_{n,t-h}^2 + b_{n,t+h}^2) = o(1)$, because of (2.21). Hence, by Lemma A.3, $T_{n,is} = o_p(\nu_n) = o_p(1)$. This completes the proof of (A.33).

Proof of the second claim in (A.26). We have $|v_{2,t}^2/v_{2,t_0}^2 - 1| \le |v_{2,t}^2 - v_{2,t_0}^2|/v_{2,t_0}^2 = o_p(1)$ because $v_{2,t_0}^2 \ge c_1 > 0$ and $v_{2,t}^2 - v_{2,t_0}^2 \to_p 0$. To show the latter, recall $v_{2,t}^2 = \beta_K \sum_{i,s=0}^\infty \rho_t^{i+s} E[u_{-i}u_{-s} u_1^2]$ and $E|u_{-i}u_{-s}u_1^2| \le Eu_1^4 < \infty$. Then by (A.17), $|v_{2,t}^2 - v_{2,t_0}^2| \le C\sum_{i,s=0}^\infty |\rho_{t_0}^{i+s} - \rho_t^{i+s}| \le C|\rho_{t_0} - \rho_t| \sum_{i,s=1}^\infty (i+s)\rho^{i+s-1} \le C|\rho_{t_0} - \rho_t| = o_p(1)$, because $|\rho_{t_0} - \rho_t| \le R_{t,2h} = O_p((h/t)^\gamma) = o_p(1)$ by (A.20).

Proof of (A.23). Observe that $y_j' = y_j - \mu_j$ is an AR(1) process (2.28) with no intercept. Let $g_t := \mu_t B_{1t}/\widehat{\sigma}^2_{Y',t}$. Then $B_{3t}^2 = \sum_{j=1}^n b_{tj}^2(1 - g_t y_{j-1}')^2 u_j^2$. We shall approximate B_{3t}^2 by $B_{3t}^{2\,(apr)} := \sum_{|j-t|\le h} b_{tj}^2(1 - \tilde{g}_{t_0} z_{t_0,j-1})^2 u_j^2$, where t_0 and $z_{t_0,j}$ are as in (A.24) and $\tilde{g}_{t_0} = \mu_{t_0} B_{1t_0}/(Hv_{1,t_0}^2)$. We shall prove that

$$B_{3t}^2 - B_{3t}^{2\,(apr)} = o_p(H), \tag{A.34}$$

$$B_{3t}^{2\,(apr)}/(Hv_{3,t_0}^2) \to_p 1, \quad v_{3,t_0}^2/v_{3,t}^2 \to_p 1, \tag{A.35}$$

which implies (A.23): $(Hv_{3,t}^2)^{-1}B_{3t}^2 = (Hv_{3,t}^2)^{-1}(B_{3t}^2 - B_{3t}^{2\,(apr)}) + (Hv_{3,t}^2)^{-1}B_{3t}^{2\,(apr)} \to_p 1$ because $v_{3,t} \ge c_1 > 0$.

Claim (A.34) follows using a similar argument as in the proof of (A.25) combined with

$$g_t - \tilde{g}_{t_0} = o_p(1), \quad \tilde{g}_{t_0} = O_p(1). \tag{A.36}$$

To verify (A.36), note that

$$|g_t - \tilde{g}_{t_0}| \le B_{1t}\{|(\mu_t - \mu_{t_0})\widehat{\sigma}^{-2}_{Y',t}| + |\mu_{t_0}|(|\widehat{\sigma}^{-2}_{Y',t} - \widehat{\sigma}^{-2}_{Y',t_0}| + |\widehat{\sigma}^{-2}_{Y',t_0} - (Hv_{1,t_0}^2)^{-1}|)\} = o_p(1).$$

The latter holds because $B_{1t} = O(H)$ by (A.1), $\mu_t - \mu_{t_0} = O_p((h/n)^\beta + h^{-1}) = o_p(1)$ and $\mu_{t_0} = O_p(1)$ by Assumption 2.3, whereas by (A.22) $\widehat{\sigma}^{-2}_{Y',t} = O_p(H^{-1})$, $\widehat{\sigma}^{-2}_{Y',t_0} - (Hv_{1,t_0}^2)^{-1} = o_p(H)$ and

$$\widehat{\sigma}^{-2}_{Y,t} - \widehat{\sigma}^{-2}_{Y,t_0} = o_p(H^{-1}). \tag{A.37}$$

To verify (A.37), note that an approximation, similar to (A.25), implies $\widehat{\sigma}^2_{Y,t} - \widehat{\sigma}^2_{Y,t_0} = o_p(H)$, while by (A.22), $\widehat{\sigma}^{-2}_{Y,t} = O_p(H^{-1})$ and $\widehat{\sigma}^{-2}_{Y,t_0} = O_p(H^{-1})$. Hence, $|\widehat{\sigma}^{-2}_{Y,t} - \widehat{\sigma}^{-2}_{Y,t_0}| = \sigma^{-2}_{Y,t}\sigma^{-2}_{Y,t_0}|\widehat{\sigma}^2_{Y,t} - \widehat{\sigma}^2_{Y,t_0}| = o_p(H^{-1})$. The above bounds also yield $|\tilde{g}_{t_0}| = B_{1t}|\mu_{t_0}|(Hv_{1,t_0}^2)^{-1} = O_p(1)$, proving (A.36).

Claims in (A.35) follow by arguing as in the proof of (A.26) and using (A.36). This completes the proof of the lemma. □

We state for convenience the following result, which is shown in Lemma 4.7 of Dalla et al. (2012).

Lemma A.3. *Let $T_n = \sum_{j\in\mathbb{Z}} z_{nj}x_j$, where $\{x_j\}$ is a stationary ergodic sequence, $E|x_1| < \infty$, and z_{nj} are real numbers such that for some $\nu_n < \infty$, $\sum_{j\in\mathbb{Z}} |z_{nj}| = O(\nu_n)$ and $\sum_{j\in\mathbb{Z}} |z_{nj} - z_{n,j-1}| = o(\nu_n)$, as $n \to \infty$. Then $E|T_n - ET_n| = o(\nu_n)$.*

Next we establish asymptotic normality of the sum $S_{Yu,t} = \sum_{j=1}^n b_{tj}y_{j-1}u_j$ appearing in (A.5), and of $Z_{n,t} = \bar{u}_t + \mu_t(\rho_t - \widetilde{\rho}_{n,t})$, used in (A.13). Recall definition (2.32) of B_{3t}.

Lemma A.4. *The following holds for $S_{Yu,t}$ in Theorem 2.3(ii) and for $Z_{n,t}$ in Theorem 2.4(ii):*

$$\widehat{\sigma}_{Yu,t}^{-1} S_{Yu,t} \to_D N(0,1), \tag{A.38}$$

$$B_{3t}^{-1} B_{1t} Z_{n,t} \to_D N(0,1). \tag{A.39}$$

Proof. To prove (A.38), set $h = b\bar{H}$ with b as in (A.16), and $t_0 = t - 2h$. We shall approximate $S_{Yu,t}$ by $S_{Yu,t}^{(apr)} = \sum_{|j-t|\le h} b_{tj} z_{t_0,j-1} u_j$ with $z_{t_0,j}$ as in (A.24). We will show that

$$S_{Yu,t} - S_{Yu,t}^{(apr)} = o_p(H^{1/2}), \tag{A.40}$$

$$d_n^{-1} S_{Yu,t}^{(apr)} \to_D \mathcal{N}(0,1), \quad d_n^2 := Hv_{2,t_0}^2. \tag{A.41}$$

By (A.22) and (A.26), $d_n^2/\widehat{\sigma}_{Yu,t}^2 \to_p 1$, and $\widehat{\sigma}_{Yu,t}^{-1} = O_p(H^{-1/2})$, which implies (A.38):

$$\frac{S_{Yu,t}}{\widehat{\sigma}_{Yu,t}} = \frac{S_{Yu,t} - S_{Yu,t}^{(apr)}}{\widehat{\sigma}_{Yu,t}} + \frac{S_{Yu,t}^{(apr)}}{d_n}\frac{d_n}{\widehat{\sigma}_{Yu,t}}$$
$$= o_p(1) + \frac{S_{Yu,t}^{(apr)}}{d_n}(1 + o_p(1)) \to_D \mathcal{N}(0,1).$$

Proof (A.40). Since u_j is a m.d. noise,

$$\begin{aligned} E\big(S_{Yu,t} - S_{Yu,t}^{(apr)}\big)^2 &= E\Bigg(\sum_{|j-t|\le h} b_{tj}(y_{j-1} - z_{t_0,j-1})u_j\Bigg)^2 \\ &\quad + E\Bigg(\sum_{j:|j-t|>h} b_{tj} y_{j-1} u_j\Bigg)^2 \\ &\le E\sum_{|j-t|\le h} b_{tj}^2 (y_{j-1} - z_{t_0,j-1})^2 u_j^2 \\ &\quad + E\sum_{j:|j-t|>h} b_{tj}^2 (y_{j-1}u_j)^2 = o(H) \end{aligned}$$

by the same argument as in the proof of (A.25). This proves (A.40).

Proof (A.41). Notice that d_n is $\mathcal{F}_{t_0}$ measurable and $X_{nj} := d_n^{-1} b_{tj} z_{t_0,j-1} u_j I(|j-t| \le h)$ are martingale differences with respect to filtration $\mathcal{F}_j$. By Theorem 3.2 of Hall and Heyde (1980), to prove $\sum_{j=1}^n X_{nj} \to_D \mathcal{N}(0,1)$ it suffices to verify that

$$\text{(a)} \sum_{j=1}^n X_{nj}^2 \to_p 1, \qquad \text{(b)} \max_{1\le j\le n} |X_{nj}| \to_p 0, \qquad \text{(c)}\, E\max_{1\le j\le n} X_{nj}^2 \to 0. \tag{A.42}$$

Claim (a) is shown in (A.26). To show (b), notice $d_n^{-1} \le CH^{-1/2}$ because $v_{2,t_0} \ge c > 0$ by Lemma A.2. Bound $|z_{t_0,j}| = |\sum_{k=0}^h \rho_{t_0}^k u_{j-k}| \le \zeta_j := \sum_{k=0}^\infty \rho^k |u_{j-k}|$ where ζ_j is a stationary process with $E\zeta_1^4 < \infty$. Thus, $|X_{nj}| \le CH^{-1/2} b_{nj}\theta_j \le C'H^{-1/2}\theta_j$, $\theta_j := |\zeta_{j-1}u_j|$. Hence, $EX_{nj}^2 I(|X_{nj}| \ge \varepsilon) \le CH^{-1} b_{nj}^2 E[\theta_j I(C'^2 H^{-1/2}\theta_j \ge \varepsilon)]$, and

$$\begin{aligned} P\Big(\max_{1\le j\le n} |X_{nj}| \ge \varepsilon\Big) &\le \varepsilon^{-2}\sum_{j=1}^n EX_{nj}^2 I(|X_{nj}| \ge \varepsilon) \\ &\le C\varepsilon^{-2} E[\theta_1^2 I(C'^2 H^{-1/2}\theta_1 \ge \varepsilon)] \to 0, \\ & H \to \infty,\ \forall \varepsilon > 0 \end{aligned} \tag{A.43}$$

by stationarity of θ_j, $E\theta_1^2 < \infty$ and (A.1). This proves (b) and (c).

Proof of (A.39). By definition, $B_{1t}Z_{n,t} = \sum_{j=1}^n b_{tj}u_j - g_t S_{Y'u,t} = \sum_{j=1}^n b_{tj}(1 - g_t y'_{j-1})u_j$ where $g_t = \mu_t(B_{1t}/\widehat{\sigma}_{Y',t}^2)$ and $y'_j = y_j - \mu_j$ is an AR(1) process (2.28) with no intercept. We shall approximate this sum by $Q_{n,t}^{(apr)} = \sum_{|j-t|\le h} b_{tj}(1 - \tilde{g}_{t_0} z_{t_0,j-1})u_j$, where t_0 and $z_{t_0,j}$ are as in (A.24), and $\tilde{g}_{t_0} = \mu_{t_0}B_{1t_0}/(Hv_{1,t_0}^2)$.

We show that

$$B_{1t}Z_{n,t} - Q_{n,t}^{(apr)} = o_p(H^{1/2}), \tag{A.44}$$

$$\tilde{d}_n^{-1} Q_{n,t}^{(apr)} \to_D \mathcal{N}(0,1), \quad \tilde{d}_n^2 := Hv_{3,t_0}^2. \tag{A.45}$$

By (A.23) and (A.35), $B_{13}^{-1} = O_p(H^{-1/2})$ and $\tilde{d}_n/B_{3,t} \to_p 1$, which implies (A.39):

$$\frac{B_{1t}Z_{n,t}}{B_{3,t}} = \frac{B_{1t}Z_{n,t} - Q_{n,t}^{(apr)}}{B_{3,t}} + \frac{Q_{n,t}^{(apr)}}{\tilde{d}_n}\frac{\tilde{d}_n}{B_{3,t}}$$
$$= o_p(1) + \frac{Q_{n,t}^{(apr)}}{\tilde{d}_n}(1 + o_p(1)) \to_D \mathcal{N}(0,1).$$

Proof of (A.44). Let $i_n := \sum_{|j-t|>h} b_{tj}u_j$. Then

$$|B_{1t}Z_{n,t} - Q_{n,t}^{(apr)}| \le |i_n| + |(g_t - \tilde{g}_{t_0})S_{Y'u,t} + \tilde{g}_{t_0}(S_{Y'u,t} - S_{Y'u,t}^{(apr)})|,$$

where $i_n = o_p(1)$ because $E|i_n| \le E|u_1|\sum_{j:|j-t|\ge h} b_{tj} = o(1)$ by (A.16), $S_{Yu,t} = O_p(H^{1/2})$ by (A.38) and (A.22), and $S_{Yu,t} - S_{Y'u,t}^{(apr)} = o_p(H^{1/2})$ by (A.40). Together with (A.36), this implies $|B_{1t}Z_{n,t} - Q_{n,t}^{(apr)}| = o_p(H^{1/2})$, proving (A.44).

Proof of (A.45). Write $\tilde{d}_n^{-1}Q_{n,t}^{(apr)} = \sum_{j=1}^n \nu_{nj}$ where $\nu_{nj} := \tilde{d}_n^{-1}b_{tj}(1 - \tilde{g}_{t_0}z_{t_0,j-1})u_j I(|j-t| \le h)$. Let $p_n := \log H$ and $X_{nj} := \nu_{nj}I(|\mu_{t_0}| \le p_n)$. Since $E\mu_{t_0}^2 \le C$ and $H \to \infty$, then $P(|\mu_{t_0}| \ge p_n) \to 0$. Hence, it suffices to verify that $\sum_{j=1}^n X_{nj} \to \mathcal{N}(0,1)$, which in turn holds if m.d.s. X_{nj} satisfies (a), (b) and (c) of (A.42).

Claim (a) follows from (A.35). To show (b) and (c), observe that $\tilde{d}_n^{-1} \le CH^{-1/2}$ by Lemma A.2, while $|z_{t_0,j}| \le \zeta_j$ where ζ_j is the same as in the proof of (A.41). Note that $E\zeta_1^4 < \infty$ and $|\tilde{g}_{t_0}| = |\mu_{t_0}B_{1t_0}/(Hv_{1,t_0}^2)| \le C|\mu_{t_0}|$ by (A.1) and Lemma A.2. Hence, $|X_{nj}| \le CH^{-1/2}p_n b_{nj}\theta_j \le C'H^{-1/2}p_n\theta_j$, $\theta_j = |\zeta_{j-1}u_j|$. We shall show that

$$E[p_n^2\theta_1^2 I(C'H^{-1/2}p_n\theta_1 \ge \varepsilon)] \to 0, \quad n \to \infty,\ \forall \varepsilon > 0, \tag{A.46}$$

which yields (b) and (c) by the same argument as in (A.43). By assumption, $E|u_1|^{4+\delta} < \infty$ for some $\delta > 0$. Since $|\theta_1| = |\zeta_0 u_1| \le \max(\zeta_0^2 p_n^{-2},\ p_n^2 u_1^2)$, then $p_n^2 E[\theta_1^2 I(C'H^{-1/2}p_n\theta_1 \ge \varepsilon)] \le p_n^2\big(E[\zeta_0^4 p_n^{-4}] + E[p_n^4 u_1^4 I(C'H^{-1/2}p_n^2 u_1^2 \ge \varepsilon)]\big) \le Cp_n^2(p_n^{-4} + E[p_n^4 u_1^4 (C'H^{-1/2}p_n^2 u_1^2/\varepsilon)^{\delta/2}]) \le C\big(p_n^{-2} + p_n^{6+\delta}H^{-\delta/4}E|u_1|^{4+\delta}\big) \le C(p_n^{-2} + p_n^{6+\delta}H^{-\delta/4}) \to 0$ as $H \to \infty$, proving (A.46). □

Let $\bar{y}_t$ and $\bar{u}_t$ be as in Proposition 2.1.

Lemma A.5. *Let y_t satisfy the assumptions of Theorem 2.3(i), and $\bar{H} = o(n)$. Then,*

$$\bar{y}_t = (1-\rho_t)^{-1}\bar{u}_t + O_p\big((\bar{H}/n)^\gamma + H^{-1}\big), \tag{A.47}$$

$$(B_{2t}\sigma_u)^{-1}B_{1t}\bar{u}_t \to_D N(0,1). \tag{A.48}$$

Proof. First we prove (A.47). Let $h := b\bar{H}$ be as in (A.16) and $t_0 := t - 2h$. Set $S_{y,t} := \sum_{j=1}^n b_{tj}y_j$, $S_{u,t} := \sum_{j=1}^n b_{tj}u_j$ and $S'_{u,t} := \sum_{j:|t-j|\le h} b_{tj}u_j$. Bound

$$\begin{aligned} |\bar{y}_t - (1-\rho_t)^{-1}\bar{u}_t| &= B_{1t}^{-1}|S_{y,t} - (1-\rho_t)^{-1}S_{u,t}| \\ &\le B_{1t}^{-1}\{|S_{y,t} - (1-\rho_{t_0})^{-1}S'_{u,t}| + |\{(1-\rho_{t_0})^{-1} \\ &\quad - (1-\rho_t)^{-1}\}S'_{u,t} + (1-\rho_t)^{-1}(S'_{u,t} - S_{u,t})|\} \\ &\le CH^{-1}\{|S_{y,t} - (1-\rho_{t_0})^{-1}S'_{u,t}| \\ &\quad + |(\rho_{t_0} - \rho_t)S'_{u,t}| + |S_{u,t} - S'_{u,t}|\}, \end{aligned} \tag{A.49}$$

using $B_{1t} \sim H$ and $|\rho_t| \le \rho < 1$. Observe that $\rho_{t_0} - \rho_t = O_p\big((\bar H/n)^\gamma\big)$ by (A.20), $S'_{u,t} = O_p(H^{1/2})$, because $ES'^2_{u,t} \le CB_{1t} \le H$, and $S_{u,t} - S'_{u,t} = o_p(1)$, because $E|S_{u,t} - S'_{u,t}| \le E|u_1| \sum_{j:|j-t|\ge h} b_{tj} = o(1)$ by (A.16). This, together with

$$S_{y,t} - (1-\rho_{t_0})^{-1} S'_{u,t} = O_p\big((\bar H/n)^\gamma H + 1\big) \tag{A.50}$$

and (A.49) implies (A.47). To show (A.50), let $z_{t_0,j}$ be as in (A.24). Then, $S_{y,t} - (1-\rho_{t_0})^{-1} S'_{u,t} = t_{n,1} + t_{n,2} + t_{n,3}$, where $t_{n,1} = \sum_{|t-j|\le h} b_{tj}(y_j - z_{t_0,j})$,

$$t_{n,2} := \sum_{|t-j|\le h} b_{tj} z_{t_0,j} - \left(\sum_{k=0}^{h} \rho_{t_0}^k\right) S'_t,$$

$$t_{n,3} := \sum_{|t-j|> h} b_{tj} y_j - \sum_{k=h+1}^{\infty} \rho_{t_0}^k S'_t.$$

Thus, it suffices to verify that

$$t_{n,1} = O_p\big((\bar H/n)^\gamma H + 1\big), \quad t_{n,2} = O_p(1), \quad t_{n,3} = O_p(1). \tag{A.51}$$

To bound $t_{n,1}$, note that by (A.31), $|t_{n,1}| \le C(R_{t,2h} + h^{-1}) \sum_{|t-j|\le h} b_{tj}|Y_{j,h}|$, where $E\sum_{|t-j|\le h} b_{tj}|Y_{j,h}| \le C\sum_{|t-j|\le h} b_{tj} \le CH$, and $R_{t,h} = O_p\big((h/t)^\gamma\big) = O_p\big((\bar H/n)^\gamma\big)$ by (A.20). This implies (A.51) for $t_{n,1}$.

To bound $t_{n,2}$, write

$$t_{n,2} = \sum_{j:|t-j|\le h} b_{tj} \sum_{k=0}^{h} \rho_{t_0}^k u_{j-k} - \sum_{k=0}^{h} \rho_{t_0}^k \sum_{j:|t-j|\le h} b_{tj} u_j = \sum_{k=0}^{h} \rho_{t_0}^k \theta_k,$$

where $\theta_k := \sum_{j:|t-j|\le h} b_{tj}(u_{j-k} - u_j)$. We shall show that $E|\theta_k| \le Ck$, which implies $E|t_{n,2}| \le \sum_{k=0}^{h} \rho^k E|\theta_k| \le C\sum_{k=0}^{\infty} \rho^k k < \infty$, that proves $t_{n,2} = O_p(1)$. Let $b'_{tj} := b_{tj} I(|j-t| \le h)$. Then $\theta_k = \sum_{j=2}^{n} (b'_{t,j+k} - b'_{tj}) u_j$, and

$$E|\theta_k| \le E|u_1| \sum_{j=2}^{n} |b'_{t,j+k} - b'_{tj}| \le Ck \sum_{j=2}^{n} |b'_{t,j+1} - b'_{tj}|$$
$$\le Ck \left(\sum_{|t-j|\le h} |b_{t,j+1} - b_{tj}| + b_{t,t-h} + b_{t,t+h}\right) \le Ck,$$

because b_{nj}'s are bounded and under (2.21), $\sum_{j=1}^{n} |b_{t,j+1} - b_{tj}| \le C$.

Finally, $|t_{n,3}| \le \sum_{j:|t-j|\ge h} b_{nj}|y_j| + C\rho^h S'_t$. By (2.14) and (A.16), $E\sum_{j:|t-j|\ge h} b_{nj}|y_j| = O(1)$, which together with $\rho^h = O(H^{-1})$ and $E|S'_t| \le E|u_1| \sum_{j=1}^{n} b_{nj} = O(H)$ implies $t_{n,3} = O_p(1)$.

Proof of (A.48). Write $(\sigma_u B_{2t})^{-1} B_{1t} \bar u_t = \sum_{j=1}^{n} X_{nj}$, where $X_{nj} := (\sigma_u B_{2t})^{-1} b_{tj} u_j$. Since X_{nj} is a m.d. sequence, to show the asymptotic normality (A.48), it suffices to verify conditions (a), (b) and (c) of (A.42). Observe that $E\sum_{j=1}^{n} X_{nj}^2 = 1$. Thus, the claim (a) $\sum_{j=1}^{n} X_{nj}^2 \to_p 1$ follows by the same argument as in the proof of the second claim in (A.33), while (b) and (c) can be verified arguing as in (A.43). □

In the next lemma, $S_{\hat Y \hat Y,t}$, $S_{Y'Y',t}$, $\widehat\sigma^2_{\hat Y,t}$ and $\widehat\sigma^2_{Y',t}$ are defined as in the proof of Theorem 2.4, and $\widehat\sigma^2_{\hat Y \hat u,t}$, $\widehat\sigma^2_{Y'uu,t}$, $\widehat\sigma^2_{\hat Y \hat u \hat u,t}$, $\hat y_j$ and $\hat u_j$ as in Corollary 2.3 and (A.15).

Lemma A.6. *Suppose the assumptions of Theorem 2.4(i) hold and $\bar H = o(n)$. Then,*

$$\sum_{j=1}^{n} b_{tj}(\mu_j - \mu_t) = O_p\big((\bar H/n)^\beta H + 1\big), \tag{A.52}$$

$$\begin{aligned} &S_{\hat Y \hat Y,t} - S_{Y'Y',t} = O_p((\bar H/n)^\gamma H + 1),\\ &\widehat\sigma^2_{\hat Y,t} - \widehat\sigma^2_{Y',t} = O_p((\bar H/n)^\gamma H + 1), \end{aligned} \tag{A.53}$$

$$\begin{aligned} &\widehat\sigma^2_{\hat Y \hat u,t} - \widehat\sigma^2_{Y'u,t} = o_p(H),\\ &\widehat\sigma^2_{\hat Y \hat u \hat u,t} = \widehat\sigma^2_{Y'uu,t} + o_p(H) = O_p(H). \end{aligned} \tag{A.54}$$

Proof. Since Assumption 2.3(i) is covered by Assumption 2.3(ii), it suffices to show (A.52) when μ_j satisfies Assumption 2.3(ii). Let $h = b\bar H$ be as in (A.16) and $t_0 := t - h$. Write $\sum_{j=1}^{n} b_{tj}(\mu_j - \mu_t) = \sum_{|j-t|>h}[\cdots] + \sum_{|j-t|\le h}[\cdots]$. By assumption, $\max_j E|\mu_j| < \infty$, and therefore, $E|\sum_{|j-t|>h} b_{tj}(\mu_j - \mu_t)| \le C\sum_{|j-t|>h} b_{tj} = o(1)$ by (A.16), so the first sum is $o_p(1)$. Bound $\sum_{|j-t|\le h} b_{tj}|\mu_j - \mu_t| \le 2\sum_{k=0}^{2h} b_{t,t_0+k}|\mu_{t_0+k} - \mu_{t_0}| =: s_n$. Recall $t \sim \tau n$ and $h = o(n)$. Assumption 2.3(ii) implies that $|\mu_{t_0+k} - \mu_{t_0}| \le |m(t_0,k)| + |\tilde m(t,k)|$, and by the properties of $m(t_0,k)$ and $\tilde m(t_0,k)$, $E\sum_{k=0}^{2h} b_{t,t_0+k}|m(t_0,k)| \le C(h/t)^\beta \sum_{k=0}^{2h} b_{t,t_0+k} \le O((h/t)^\beta H)$ by (A.1), while $\sum_{k=0}^{2h} b_{t,t_0+k}|\tilde m(t_0,k)| = O_p((h/t)^\beta + h^{-1}) \sum_{k=0}^{2h} b_{t,t_0+k} = O_p((h/t)^\beta H + 1)$. This yields $s_n = O_p((\bar H/t)^\beta H + 1)$, which proves (A.52).

Proof of (A.53). By (2.28), $y_j = \mu_j + y'_j$ where y'_j is an AR(1) process with no intercept. Since $\hat y_j = y_j - \bar y_t$ implies $\sum_{j=1}^{n} b_{tj}\hat y_j = 0$, then $S_{\hat Y \hat Y,t} - S_{Y'Y',t} = \sum_{j=1}^{n} b_{tj}\{\hat y_j \hat y_{j-1} - y'_j y'_{j-1}\} = \sum_{j=1}^{n} b_{tj}\{\hat y_j(y_{j-1} - \mu_t) - y'_j y'_{j-1}\}$ where $\hat y_j(y_{j-1} - \mu_t) - y'_j y'_{j-1} = \hat y_j(\mu_{j-1} - \mu_t) + (\mu_j - \bar y_t)y'_{j-1} = (\mu_{j-1} - \mu_t)\hat y_j + (\mu_j - \mu_t)y'_{j-1} + (\mu_t - \bar y_t)y'_{j-1}$. Thus,

$$S_{\hat Y \hat Y,t} - S_{Y'Y',t} = q_{n,1} + q_{n,2} + (\mu_t - \bar y_t) q_{n,3}, \tag{A.55}$$

$$q_{n,1} := \sum_{j=1}^{n} b_{tj}(\mu_{j-1} - \mu_t)\hat y_j,$$

$$q_{n,2} := \sum_{j=1}^{n} b_{tj}(\mu_j - \mu_t) y'_{j-1}, \qquad q_{n,3} := \sum_{j=1}^{n} b_{tj} y'_{j-1}.$$

By (2.14) and Assumption 2.3, $Ey'^2_j \le C$ and $Ey_j^2 = E(y'_j + \mu_j)^2 \le 2Ey'^2_j + 2E\mu_j^2 \le C$ uniformly in j. Hence, $E\hat y_j^2 \le 2Ey_j^2 + 2E\bar y_t^2 \le C$ for all j. Thus, the same argument as used in the proof of (A.52) implies $q_{n,i} = O_p\big((\bar H/n)^\beta H + 1\big)$, $i = 1, 2$. In addition, by (2.26), $\bar y'_t = O_p\big((\bar H/n)^\gamma + H^{-1/2}\big)$. Therefore, $q_{n,3} = B_{n1}\bar y'_t = O_p\big((\bar H/n)^\gamma H + H^{1/2}\big)$, while $\mu_t - \bar y_t = B_{n1}^{-1}\sum_{j=1}^{n} b_{tj}(\mu_t - \mu_j) - \bar y'_t = O_p\big((\bar H/n)^\gamma + H^{-1/2}\big)$ by (A.52), (2.26) and $\beta \ge \gamma$. Hence, $(\mu_t - \bar y_t)q_{n,3} = O_p\big((\bar H/n)^\gamma H + 1\big)$ and $S_{\hat Y \hat Y,t} - S_{Y'Y',t} = O_p\big((\bar H/n)^\gamma H + 1\big)$. This completes the proof of the first claim in (A.53). The proof of the second claim in (A.53) follows using a similar argument.

Proof of (A.54). We verify the first claim. (The second claim follows using the same argument.) By $\hat y_{j-1}^2 \hat u_j^2 - y'^2_{j-1}u_j^2 = (\hat y_{j-1}\hat u_j - y'_{j-1}u_j)^2 + 2y'_{j-1}u_j(\hat y_{j-1}\hat u_j - y'_{j-1}u_j)$ and the Cauchy inequality,

$$|\widehat\sigma^2_{\hat Y \hat u,t} - \widehat\sigma^2_{Y'u,t}| \le \sum_{j=1}^{n} b_{nj}^2 |\hat y_{j-1}^2 \hat u_j^2 - y'^2_{j-1}u_j^2| \le q_{n,1} + 2q_{n,1}^{1/2} q_{n,2}^{1/2},$$

$$q_{n,1} := \sum_{j=1}^{n} b_{nj}^2(\hat y_{j-1}\hat u_j - y'_{j-1}u_j)^2, \qquad q_{n,2} := \sum_{j=1}^{n} b_{nj}^2 y'^2_{j-1}u_j^2.$$

We will show that $q_{n,1} = o_p(H)$ and $q_{n,2} = O_p(H)$ which implies $\widehat\sigma^2_{\hat Y \hat u,t} - \widehat\sigma^2_{Y'u,t} = o_p(H)$.

To bound $q_{n,2}$, notice that y'_t is an AR(1) process (2.28) with no intercept. Hence by (2.14), $Ey'^4_j \le C$, and $Ey'^2_{j-1}u_j^2 \le Ey'^4_{j-1} + Eu_j^4 \le C$ for all j. Thus, $Eq_{n,2} \le C\sum_{j=1}^{n} b_{nj}^2 \le CH$ by (A.1), which implies $q_{n,2} = O_p(H)$.

To bound $q_{n,1}$, note that $(\hat{y}_{j-1}\hat{u}_j - y'_{j-1}u_j)^2 \le 2(\hat{y}_{j-1} - y'_{j-1})^2 u_j^2 + 2(\hat{u}_j - u_j)^2 \hat{y}_{j-1}^2$. Observe that $|\hat{u}_j - u_j| = |(\hat{y}_j - \hat{\rho}_t \hat{y}_{j-1}) - (y'_j - \rho_{j-1} y'_{j-1})| \le |\hat{y}_j - y'_j| + |\rho_{j-1}||y'_{j-1} - \hat{y}_{j-1}| + |\hat{\rho}_t - \rho_{j-1}||\hat{y}_{j-1}|$. Thus, $(\hat{y}_{j-1}\hat{u}_j - y'_{j-1}u_j)^2 \le C\{(\hat{y}_{j-1} - y'_{j-1})^2 + (\hat{y}_j - y'_j)^2\}\{u_j^2 + 1 + \hat{y}_{j-1}^2\} + (\hat{\rho}_t - \rho_{j-1})^2 \hat{y}_{j-1}^4$. To bound the term $u_j^2 + 1 + \hat{y}_{j-1}^2$ note that by (2.13) and (2.12) one has $|y'_j| \le \zeta_j + |y'_0|$ where $\zeta_j := \sum_{k=0}^{\infty} \rho^k |u_{j-k}|$ is a stationary process such that $E\zeta_1^2 < \infty$. Set $u_{j,1} := |u_j| I(|u_j| \ge L)$, $\zeta_{j,1} := |\zeta_j| I(|\zeta_j| \ge L)$ where $L > 1$. Hence, $u_j^2 \le L^2 + u_{j,1}^2$ and $\hat{y}_j^2 = (y'_j - \bar{y}_t)^2 \le 4(L^2 + \zeta_{j,1}^2 + y_0'^2 + \bar{y}_t^2)$. Then $(\hat{y}_{j-1}\hat{u}_j - y'_{j-1}u_j)^2 \le C\{(\hat{y}_{j-1} - y_{j-1})^2 + (\hat{y}_j - y_j)^2\}\{L^2 + \bar{y}_t^2 + y_0'^2 + u_{j,1}^2 + \zeta_{j,1}^2\} + (\hat{\rho}_t - \rho_{j-1})^2 \hat{y}_{j-1}^4$. Thus,

$$q_{n,1} \le C\big(r_{n,1} + (L^2 + \bar{y}_t^2 + y_0'^2) r_{n,2} + r_{n,3}\big),$$

$$r_{n,1} = \sum_{j=1}^{n} b_{nj}^2 (\hat{\rho}_t - \rho_{j-1})^2 \hat{y}_{j-1}^4,$$

$$r_{n,2} = \sum_{j=1}^{n} b_{nj}^2 \{(\hat{y}_{j-1} - y_{j-1})^2 + (\hat{y}_j - y_j)^2\}, \qquad r_{n,3} = \sum_{j=1}^{n} b_{nj}^2 \xi_{n,j},$$

where $\xi_{n,j} := \{(\hat{y}_{j-1} - y_{j-1})^2 + (\hat{y}_j - y_j)^2\}\{u_{j,1}^2 + \zeta_{j,1}^2\}$. Notice that $r_{n,1}$, $r_{n,2}$ do not depend on L and $r_{n,1} = o_p(H)$, $r_{n,2} = o_p(H)$ which follows using the same argument as in the proof of (A.10) and (A.52), respectively. For a fixed L, $(L^2 + \bar{y}_t^2 + y_0'^2) = O_p(1)$ because $E\bar{y}_t^2 = O(1)$ and $Ey_0'^2 < \infty$. In addition we show that

$$H^{-1} r_{n,3} \to_p 0, \quad n \to \infty, \ L \to \infty, \tag{A.56}$$

which, together with the above relations, implies $q_{n,1} = o_p(H)$. To bound $r_{n,3}$, notice that $E\hat{y}_j^4 = E(\mu_j + y'_j)^4 \le 4(E\mu_j^4 + Ey_j'^4) \le C$ by (2.14). Hence, by the Cauchy inequality and the stationarity of u_j and ζ_j, $E\xi_{n,j} \le C\{(Eu_{j,1}^4)^{1/2} + (E\zeta_{j,1}^4)^{1/2}\} = C((Eu_{1,1}^4)^{1/2} + (E\zeta_{1,1}^4)^{1/2}) =: \varepsilon_L \to 0$ as $L \to \infty$. Hence, $EH^{-1}|r_{n,3}| \le \varepsilon_L(H^{-1}\sum_{j=1}^{n} b_{nj}^2) \le C\varepsilon_L \to 0$ as $L \to \infty$ by (A.1), which proves (A.56). □

References

Abadir, K., Distaso, W., Giraitis, L., Koul, H.L., 2013. Asymptotic normality for weighted sums of linear processes. Econometric Theory (forthcoming).
Anderson, T.W., 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. Annals of Mathematical Statistics 30, 676–687.
Andrews, D.W.K., Guggenberger, P., 2008. Asymptotics for stationary very nearly unit root processes. Journal of Time Series Analysis 29, 203–210.
Benati, L., 2010. Investigating inflation persistence in monetary regimes. Quarterly Journal of Economics 123, 1636–1652.
Benati, L., Surico, P., 2008. Evolving US monetary policy and the decline of inflation predictability. JEEA 6, 634–646.
Benigno, G., 2002. Real exchange rate persistence and monetary policy rules. Journal of Monetary Economics 51, 473–502.
Benigno, P., Ricci, L.A., Surico, P., 2010. Unemployment and productivity in the long run: the role of macroeconomic volatility. NBER Working Paper no. 16374.
Brown, R.L., J, Durbin, Evans, J.M., 1974. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Association, Series A 138, 149–163.
Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.
Chari, V., Kehoe, P., McGrattan, E., 2002. Can sticky price models generate volatile and persistent real exchange rates? Review of Economic Studies 69, 533–563.
Chen, S., Engel, C., 2005. Does 'aggregation bias' explain the PPP puzzle? Pacific Economic Review 10, 49–72.
Chortareas, G.E., Kapetanios, G., 2009. Getting PPP right: identifying mean-reverting real exchange rates in panels. Journal of Banking and Finance 33, 390–404.
Chortareas, G.E., Kapetanios, G., Shin, Y., 2002. Nonlinear mean reversion in real exchange rates. Economics Letters 77, 411–417.
Chow, A., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.
Christiano, L., Eichenbaum, M., Evans, C., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113, 1–45.
Cogley, T., Sargent, T.J., 2001. Evolving post World War 2 US inflation dynamics. NBER Macroeconomics Annual 16, 331–337.
Cogley, T., Sargent, T.J., 2005a. The conquest of US inflation: learning and robustness to model uncertainty. Review of Economic Dynamics 8, 528–563.
Cogley, T., Sargent, T.J., 2005b. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics 8, 262–302.
Cogley, T., Sargent, T., Primiceri, G., 2010. Inflation gap persistence in the US. American Economic Association: Macroeconomics 2, 43–69.
Cooley, T.F., Prescott, E.C., 1973. An adaptive regression model. International Economic Review 14, 364–371.
Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Annals of Statistics 25, 1–37.
Dahlhaus, R., Giraitis, L., 1998. On the optimal segment length for parameter estimates for locally stationary time series. Journal of Time Series Analysis 19, 629–656.
Dalla, V., Giraitis, L., Koul, H., 2012. Studentising weighted sums of linear processes. Journal of Time Series Analysis (forthcoming).
Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Reviews 3, 1–100.
Frankel, J.A., Rose, A.K., 1996. A panel project on purchasing power parity: mean reversion within and between countries. Journal of International Economics 40, 209–224.
Giraitis, L., Koul, H., Surgailis, D., 2012. Large Sample Inference for Long memory Processes. Imperial College Press, London.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press.
Imbs, J., Mumtaz, H., Ravn, M., Rey, H., 2005. PPP strikes back: aggregation and the real exchange rate. Quarterly Journal of Economics 120, 1–43.
Kapetanios, G., Tzavalis, E., 2010. Modeling structural breaks in economic relationships using large shocks. Journal of Economic Dynamics and Control 34, 417–436.
Kapetanios, G., Yates, T., 2008. An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change. Mimeo, Bank of England.
Koop, G., Potter, S., 2008. Time varying VARs with inequality restrictions. University of Strathclyde/FRB New York, mimeo.
Lai, T.L., Wei, C.Z., 2010. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Annals of Statistics 10, 154–166.
Lucas, R.E., 1976. Econometric policy evaluation: a critique. In: Brunner, K., Meltzer, K. (Eds.), The Phillips Curve and Labour Markets. North-Holland, Amsterdam.
Muller, U.K., Petalas, P.M., 2010. Efficient estimation of the parameter path in unstable time series models. Review of Economic Studies 77, 1508–1539.
Muller, U.K., Watson, M.W., 2008. Testing models of low-frequency variability. Econometrica 76, 979–1016.
Mumtaz, H., Surico, P., 2009. Time varying yield curve dynamics and monetary policy. Journal of Applied Econometrics 24, 895–913.
Nelson, E., 2005. The great inflation of the seventies: what really happened? The B.E. Journal of Macroeconomics 0 (1), 3.
Papell, D.H., 1997. Searching for stationarity: purchasing power parity under the current float. Journal of International Economics 43, 313–332.
Papell, D.H., Prodan, R., 2006. Additional evidence of long run purchasing power parity with restricted structural change. Journal of Money, Credit and Banking 38, 1329–1350.
Papell, D.H., Theodoridis, H., 1998. Increasing evidence of purchasing power parity over the current float. Journal of International Money and Finance 17, 41–50.
Papell, D.H., Theodoridis, H., 2001. The choice of numeraire currency in panel tests of purchasing power parity. Journal of Money, Credit and Banking 33, 790–803.
Pesaran, M.H., Pettenuzzo, D., Timmermann, A., 2006. Forecasting time series subject to multiple structural breaks. Review of Economic Studies 73 (4), 1057–1084.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. Journal of Econometrics 136, 115–130.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.
Ploberger, W., Kramer, W., 1992. The CUSUM test with OLS residuals. Econometrica 60, 271–285.
Priestley, M., 1965. Evolutionary spectra and nonstationary processes. Journal of Royal Statistical Society, Series B 27, 204–237.
Robinson, P.M., 1989. Nonparametric estimation of time varying parameters. In: Hackl, P. (Ed.), Economic Structural Change: Analysis and Forecasting. Springer, Berlin, pp. 253–264.
Robinson, P.M., 1991. Time varying nonlinear regression. In: Hackl, P. (Ed.), Economic Structural Change: Analysis and Forecasting. Springer, Berlin, pp. 179–190.
Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34, 647–668.
Sargent, T.J., 2001. The Conquest of American Inflation. Princeton University Press.
Sargent, T.J., Williams, N., Zha, T., 2006. Shocks and government beliefs: the rise and fall of American inflation. American Economic Review 96, 1193–1224.
Segalotto, J.-F., Arnone, M., Laurens, B., 2006. Measures of Central Bank autonomy: Empirical evidence for OECD, developing, and emerging market economies. IMF Working Papers 06/228, International Monetary Fund.
Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilbrium model for the Euro Area. JEEA 1, 1123–1175.
Stock, J.H., Watson, M.W., 1998. Median unbiased estimation of coefficient variance in a time-varying parameter model. Journal of the American Statistical Association 93 (441), 349–358.
Stout, W.F., 1974. Almost Sure Convergence. Academic Press.