Inference on stochastic time-varying coefficient models

L. Giraitis*
Queen Mary, University of London

G. Kapetanios
Queen Mary, University of London

T. Yates
Bank of England

October 22, 2012

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, rigourously, alternative kernel based, nonparametric estimation approaches for such models and derives their basic properties. The use of such estimation methods for stochastic time-varying coefficient models, or any persistent stochastic process for that matter, is novel and has not been suggested previously in the literature. 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 PPP hypothesis.

Key words: Time-varying coefficient models, random coefficient models, nonparametric estimation, kernel estimation, autoregressive processes.

JEL Classification: C10, C14.

*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 very constructive comments.

1 Introduction

This paper proposes kernel-based nonparametric methods for inference on the paths of the unobserved drifting coefficient processes in random time varying coefficient (RC) models, such as (2.1). 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), Cogley and Sargent (2005), Cogley, Sargent and Primiceri (2010), Benati (2010), Pesaran, Pettenuzzo, and Timmermann (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 formalise 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 generalising 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 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. In addition, 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 2.2.1 presents the model and some of its basic properties that are of use for later developments. Section 2.2.2 contains 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 4.3 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 chance is likely to be responsible for most major forecast failures of time invariant series 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, Durbin, and Evans (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 (2004) 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, Litterman, and Sims (1984) who estimate an RC model on macroeconomic time series and emphasise 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 (2005) 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 90's and early 2000's, dubbed the 'Great Moderation'. In this work, and work influenced by it, the authors assume a random walk process for the coefficients, but bound the coefficients of the VAR model so that at each point in time the VAR is non-explosive. In the univariate case 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 (2005) 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, Sargent, and Primiceri (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 characterise evolutions in the term structure and the correspondence of changes therein with the monetary regime. Benigno, Ricci, and Surico (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 coefficents 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.

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 has meant that the Kalman filter approach may not be appropriate 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), Robinson (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 discuss the recent evolution of important macroeconomic variables. 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 analysed 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 (2005) 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), 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 process 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 drifting coefficient process is a driftless random walk.

We consider the AR(1) models

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

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

with a drifting random coefficient ρ_t, random intercept α_t and initialization y_0, where $\{u_t\}$ is a stationary ergodic martingale difference sequence (m.d.s.) with respect to the natural filtration $\mathcal{F}_t = \sigma\{u_j, \rho_j, j \le t\}$.

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 stationary process, which has different theoretical properties compared to AR processes with random coefficients. Moreover, the model (2.1) 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

$$\rho_t = \rho \frac{a_t}{\max_{0 \le k \le t} |a_k|}, \quad t \ge 1, \tag{2.2}$$

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. Although the most novel case is $Ea_t = 0$, we shall assume that ρ_t combines deterministic and random components.

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

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

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

Part (ii) shows 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 (2005). In addition, if v_1 has $2 + \delta$ finite moments, then the process $a_{[\tau n]}$, $0 \le \tau \le 1$ converges weakly in Shorokhod 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.

[1]The results of this paper remain valid also for negative $\rho \in (-1, 0)$ in (2.2).

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

$$n^{-\gamma} a_{[\tau n]} \Rightarrow_{D[0,1]} W_\tau + g(\tau), \quad 0 \le \tau \le 1, \tag{2.3}$$

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.*

In addition,

$$\begin{aligned} & n^{-\gamma}(a_{[\tau n]} - Ea_{[\tau n]}) \Rightarrow_{D[0,1]} W_\tau, \quad n^{-\gamma} Ea_{[\tau n]} \to g(\tau), \quad 0 \le \tau \le 1, \\ & |Ea_t - Ea_{t+k}| \le Ck^\gamma, \qquad 1 \le k < t. \end{aligned} \tag{2.4}$$

Remark 2.1. Assumption 2.2 is satisfied by a_t's with $Ea_t = 0$, whose first differences $v_j = a_j - a_{t-j}$ form a zero mean linear process

$$v_j = \sum_{k=0}^{\infty} \nu_k \zeta_{j-k}, \quad j \ge 0, \qquad \sum_{k=0}^{\infty} \nu_k^2 < \infty, \tag{2.5}$$

with stationary ergodic m.d. innovations ζ_k, $E\zeta_1^2 < \infty$, and has regularly increasing variance:

$$\mathrm{Var}(a_n) = \mathrm{Var}(\sum_{j=1}^{n} v_j) \sim Cn^{2\gamma}, \quad \text{for some } \gamma \in (0,1). \tag{2.6}$$

Properties (2.5) and (2.6) imply the weak convergence (2.3) with a fractional Brownian motion limit W_τ, as long as $E|\zeta_1|^p < \infty$ for some $p > \max(1/\gamma, 2)$, see Abadir, Distaso, Giraitis and Koul (2010, Theorem 3.1). ARMA(p,q) processes v_j satisfy (2.6) with $\gamma = 1/2$, while ARFIMA(p,d,q), $|d| < 1/2$ processes with $\gamma = (1/2) + d$. Seasonal long memory GARMA processes v_j, which spectral density has a singularity at frequency $\omega \ne 0$, satisfy (2.5) and (2.6) with $\gamma = 1/2$, see Section 7.2.2 of Giraitis, Koul, and Surgailis (2012).

Under Assumption 2.2, the coefficient process ρ_t, as n increases, behaves as a rescaled limit process W_τ of (2.4):

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

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 $\to_D$ denotes convergence in distribution, $=_D$ 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, is*

$$a_t = z_t + t^\gamma g_t = \{v_1 + \cdots + v_t\} + t^\gamma \times \frac{h_1 + \cdots + h_t}{t}, \tag{2.7}$$

where v_j's are stationary zero mean r.v.'s and h_j are non-random numbers such that $\max_j |h_j| < \infty$. *Then, for* $1 \le k \le t$, $|Ea_t - Ea_{t+k}| = |t^\gamma g_k - (t+k)^\gamma g_{t+k}| \le Ck^\gamma$. *Hence, a_t satisfies Assumption 2.2, if $z_t = a_t - Ea_t$ has a weak limit $n^{-\gamma} z_{[\tau n]} \Rightarrow_{D[0,1]} W_\tau$, see, e.g., Remark 2.1, and the mean $Ea_t = t^\gamma g_t \equiv t^\gamma g(t/n)$ satisfies $n^{-\gamma} Ea_t = n^{-\gamma} g([\tau n]/n) \to \tau^\gamma g(\tau)$.*

Such a_t with trending mean has asymptotically non-diminishing random and deterministic components. When v_j's are weakly dependent, one has $a_t = z_t + t^{1/2} g_t$. It also allows to generate non-random coefficients ρ_t used in modelling locally stationary processes.

Parametric random coefficients. Example (2.7) suggests a simple parametrization for an AR(1) random coefficient:

$$\rho_t = \rho \frac{c \sum_{k=1}^t u_k + t^{1/2}}{\max_{1 \le j \le t} |c \sum_{k=1}^j u_k + j^{1/2}|}, \tag{2.8}$$

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

Remark 2.2. To restrict ρ_t in the interval $[-\rho, \rho]$, we use the normalization $\rho_t = \rho\, a_t / \max_{0 \le k \le t} |a_k|$. Normalization $\rho_t = \rho\, a_t / \max_{0 \le k \le n} |a_k|$ also could 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 how to restrict ρ_t is open. Usually in the macroeconometric literature restriction is based on computational convenience without discussing the properties of the resulting model, and what is the best way how to restrict the process ρ_t from an economic point. In general, such restriction can be tackled in a variety of ways none of which detracts from the main findings of the paper.

Remark 2.3. The paper narrows its interest to AR(1) time-varying random framework and identifying conditions that allow rigorous inference on it. It is shown, that kernel estimation and inference extends to coefficients composed of time varying random and deterministic parts. Such a finding neither is intuitively obvious nor has a trivial formal justification. Establishing the AR(1) framework opens the possibility for 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

$$\rho_{t,i} = \rho \frac{a_{t,i}}{\max_{0\le k\le t} \sum_{i=1}^{p} |a_{k,i}|}, \quad t \ge 1, \tag{2.9}$$

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

$$A_t = \begin{pmatrix} \rho_{t,1} & \rho_{t,2} & \ldots & \rho_{t,p} \\ 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 1 & 0 \end{pmatrix}, \tag{2.10}$$

or its spectral norm $||A_t||_{sp}$, are bounded above by one, for all t. Such requirement on companion matrix in (2.10) plays a similar role to standardization of a_t in (2.2) and is reminiscent of the standard stationarity condition for fixed coefficient $AR(p)$ models.
(2) Time Varying $VAR(1)$ model is given by

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

where $\boldsymbol{y}_t$ is an $m-$dimensional vector and $\boldsymbol{\Psi}_{t-1} = [\psi_{t-1,ij},\ i, j = 1, \cdots, m]$. The necessary bounding can be implemented by defining, similarly to (2.9),

$$\psi_{t-1,ij} = \rho \frac{a_{t-1,ij}}{\max_{1\le i\le t} \sum_{j=1}^{m} |a_{t-1,ij}|}, \quad t \ge 1,$$

where $0 < \rho < 1$, and $a_{t,ij} = a_{t-1,ij} + u_{t,ij}$, where $u_{t,ij}$ are zero mean m.d. sequences with finite variance. This ensures that the maximum eigenvalue of $\boldsymbol{\Psi}_{t-1}$ is bounded above by one in absolute value. A third extension concerns modeling 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 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 properties of its covariance function. To write y_t as a moving average of the noise u_j, define the (random) weights

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

Note that

$$|c_{t,j}| \leq |\rho|^j, \quad 1 \leq j \leq t. \tag{2.11}$$

Next theorem describes basic properties of AR(1) process y_t, $t = 1, \cdots, n$, (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

$$y_t = \sum_{j=0}^{t-1} c_{t,j} u_{t-j} + c_{t,t} y_0, \quad t \geq 1. \tag{2.12}$$

(ii) The second and fourth moments satisfy

$$Ey_t^2 \leq 2(1-\rho)^{-2}(\sigma_u^2 + Ey_0^2), \quad Ey_t^4 \leq 4(1-\rho)^{-4}(Eu_1^4 + Ey_0^4). \tag{2.13}$$

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

$$z_t(\rho_t) := \sum_{k=0}^{t-1} \rho_t^j u_{t-k}, \quad t \geq 1, \tag{2.14}$$

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

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

$$y_t = z_t(\rho_t) + o_p(1), \quad Ey_t \to 0. \tag{2.15}$$

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

$$E[y_t y_{t+k}] = \sigma_u^2 E[\rho_t^{|k|}(1-\rho_t^2)^{-1}] + o(1), \quad t \to \infty, \quad \forall k, \tag{2.16}$$
$$|\mathrm{Cov}(y_t, y_s)| \leq \rho^{|t-s|}(1-\rho^2)^{-1}(\sigma_u^2 + Ey_0^2), \quad \forall t, s \geq 1. \tag{2.17}$$

2.2.2 Estimation and Inference

In this section we construct a feasible estimation procedure of the drifting coefficient $\rho_1, \cdots, \rho_n$, based on observables $y_1, \cdots, y_n$ following the AR(1) model (2.1) with no intercept. We consider an estimate of ρ_t, that can be written as a weighted sample autocorrelation at lag 1. We shall show that under Assumptions 2.1 and 2.2, it is consistent and asymptotically normally distributed. Computation of standard errors modified to accommodate a martingale difference noise u_t is straightforward, and the method allows the construction of pointwise confidence intervals for the drifting coefficient under minimal restrictions on ρ_t.

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

$$H \to \infty, \quad H = o(n). \tag{2.18}$$

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 $y_{t-H}, \cdots, y_{t+H}$. We shall consider a general class of estimators

$$\hat{\rho}_{n,t} := \frac{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_k y_{k-1}}{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_{k-1}^2}, \tag{2.19}$$

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

$$K(x) = O(e^{-cx^2}),\ \exists c > 0,\ |(d/dx)K(x)| = O(|x|^{-2}), \quad x \to \infty. \tag{2.20}$$

Examples of K include

$$\begin{aligned} &K(x) = (1/2)I(|x| \leq 1), \quad \text{flat kernel,} \\ &K(x) = (3/4)(1-x^2)I(|x| \leq 1), \quad \text{Epanechnikov kernel,} \\ &K(x) = (1/\sqrt{2\pi})e^{-x^2/2}, \quad \text{Gaussian kernel.} \end{aligned}$$

The flat and Epanechnikov kernels have a finite support, whereas Gaussian kernel has an infinite support. The above kernels satisfy (2.20).

Now we discuss the asymptotic properties of the estimator $\hat{\rho}_{n,t}$ of (2.19). Denote for $1 \leq t, k \leq n$,

$$b_{tk} := K(\frac{t-k}{H}), \quad \hat{\sigma}^2_{Y,t} = \sum_{k=1}^{n} b_{tk} y_{k-1}^2, \quad \hat{\sigma}^2_{Yu,t} = \sum_{k=1}^{n} b_{tk}^2 y_{k-1}^2 u_k^2.$$

To establish asymptotic normality we will assume that the conditional variance of the noise $E[u_j^2|F_{j-1}]$, $j \geq 1$ is bounded away from 0. In the next theorem and below, we set $\bar{H} = H$, if kernel K has finite support, and $\bar{H} = H \log^{1/2} H$, if kernel K has infinite support.

Theorem 2.3. *Let $y_1, \cdots, y_n$ be defined as in (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 (2.18) and (2.20), respectively.*
(i) Then,

$$\hat{\rho}_{n,t} - \rho_t = O_p\big((\bar{H}/n)^{\gamma} + H^{-1/2}\big). \tag{2.21}$$

(ii) In addition, if $E[u_1^2|u_0, u_{-1}, \cdots] \geq c > 0$, for some $c > 0$, and $(\bar{H}/n)^\gamma = o(H^{-1/2})$, then

$$\frac{\widehat{\sigma}^2_{Y,t}}{\widehat{\sigma}_{Yu,t}}\Big(\hat{\rho}_{n,t} - \rho_t\Big) \to_D N(0,1). \tag{2.22}$$

In particular, for $\gamma \geq 1/2$, (2.22) holds, if $H = o(n^{1/2}/\log^{1/4} n)$.

Notice that studentization under martingale difference noise u_j, is different from that used in i.i.d. case, while the requirement for conditional variance $E[u_1^2|u_0, u_{-1}, \cdots]$ to be bounded away from 0 is satisfied by i.i.d. and ARCH noises u_j.

The next corollary establishes the asymptotic of random normalization $\widehat{\sigma}^2_{Y,t}/\widehat{\sigma}_{Yu,t}$ and the $\sqrt{H}$ rate of convergence in (2.22). We will use the following notation: $v^2_{1,t} = \sigma^2_u(1-\rho_t^2)^{-1}$ and $v^2_{2,t} = \beta_K \sum_{i,k=0}^\infty \rho_t^{i+k} E[u_{-i}u_{-k}u_1^2]$ where $\beta_K = \int K^2(x)dx$.

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

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

where $c_1 \leq v^2_{1,t}/v_{2,t} \leq c_2$, for some constants $c_1, c_2 > 0$.

The above corollary follows from Lemma 5.2. The following corollary shows that studentization in (2.22) can be based on the residuals $u_j = y_j - \hat{\rho}_t y_{j-1}$, $j = 1, \cdots, n$, computed using $\hat{\rho}_t$. Let $\widehat{\sigma}^2_{Y\widehat{u},t} := \sum_{j=1}^n b^2_{tj} y^2_{j-1} \widehat{u}^2_j$.

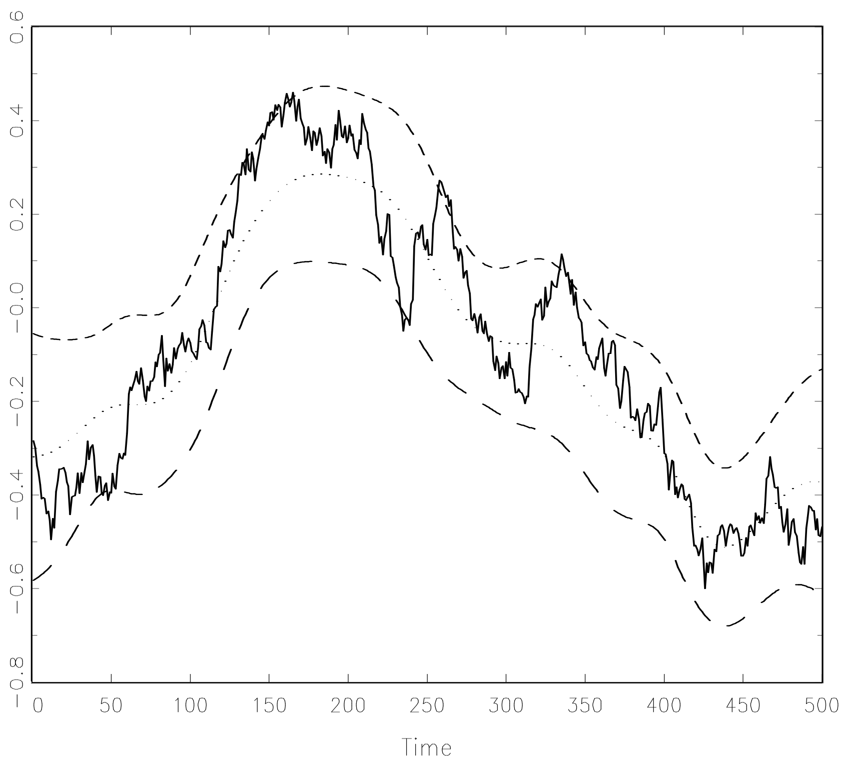
Corollary 2.2. *Under assumptions of Theorem 2.3(ii), in addition to (2.22), it holds*

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

Remark 2.4. The consistency of the estimator $\hat{\rho}_{n,t}$ requires persistence of the process ρ_t, generated by non-stationarity (stochastic or deterministic trending behavior) of the process a_t. The main restriction on a_j is to satisfy the functional central limit theorem $n^{-\gamma} a_{[\tau n]} \to_{D[0,1]} W_\tau + g(\tau)$ with some normalization $n^{-\gamma}$, $0 < \gamma < 1$. Larger values of γ correspond to a stronger persistence in a_t. By assumption, the first difference process $v_j = \{a_j - Ea_j\} - \{a_{j-1} - Ea_{j-1}\}$ is a stationary process and may have short, long or negative memory in cases $\gamma = 1/2$, $1/2 < \gamma < 1$ and $0 < \gamma < 1/2$, respectively. Application of the normal approximation (2.22) does not require knowledge of γ. It is practical to choose $H = o(n^{1/2})$, leading to a asymptotic normality (2.22) for $1/2 \leq \gamma < 1$. When γ is close to 0, the pattern of trending behavior of a_j and the quality of approximation (2.22) deteriorate. In case of a stationary process a_j, the above estimation of ρ_j is not consistent.

To give an idea of the nature of the pointwise confidence intervals, implied by Theorem 2.3, we include Figure 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 $H = \sqrt{n}$,

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



and 90% confidence intervals. It shows that the process is well tracked and the point-wise confidence band contains the true process most of the time (for 92.8% of t's).

We complete this section describing properties of the estimate $\bar{y}_t := B_{1t}^{-1} \sum_{j=1}^{n} b_{tj} y_j$, $B_{1t} := \sum_{k=1}^{n} b_{tk}$, which in case of non-random coefficient is the estimate of the mean. The following proposition shows that if y_t follows AR(1) model with no intercept, then $\bar{y}_t \to_p 0$ and satisfies the standard normal approximation. The martingale approximation of (2.23) to the weighted sum $\bar{y}_t$ resembles the well-known approximation result for stationary linear processes, based on Beveridge-Nelson decomposition, see Phillips and Solo (1992).

Let $B_{2t}^2 = \sum_{k=1}^{n} b_{tj}^2$, 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,*

$$\bar{y}_t = (1 - \rho_t)^{-1} \bar{u}_t + O_p\big((\bar{H}/n)^{\gamma}\big) + o_p(H^{-1/2}), \tag{2.23}$$

$$\frac{B_{1t}}{B_{2t}}\frac{1-\rho_t}{\sigma_u}\bar{y}_t \to_D N(0,1), \quad \textit{if } (\bar{H}/n)^\gamma = o(H^{-1/2}). \tag{2.24}$$

2.3 AR(1) process with a persistent random attractor

In this section we extend the AR(1) model by adding a persistent term which, in standard AR models, plays the role of the mean. Recall that, by (2.15), $Ey_t \to 0$, as $t \to \infty$.

Given observations $y_1, \cdots, y_n$, 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 (less persistent) dynamic component $y_t - \mu_t$, which evolves as AR(1) process (2.1):

$$y_t - \mu_t = \rho_{t-1}(y_{t-1} - \mu_{t-1}) + u_t. \tag{2.25}$$

Clearly, this model can be written as an AR(1) process $y_t = \alpha_t + \rho_{t-1}y_{t-1} + u_t$ with the intercept $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$.

To make such decomposition meaningful, a random process μ_t has to be sufficiently persistent, so that both μ_t and ρ_t can be extracted from the data. Although such μ_t can be estimated by the weighted average $\bar{y}_t$, in general, μ_t cannot be interpreted as the mean Ey_t.

Permissible classes of processes μ_t allowing estimation of μ_t and AR(1) dynamics is described by the following assumption.

Assumption 2.3. *μ_t is $\mathcal{F}_t = \sigma(u_j, a_j, j \le t)$ measurable, $\max_j E\mu_j^2 < \infty$, and for some $\beta \in (0, 1]$, satisfies either (i) or (ii).*

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

(ii) *One can write $\mu_t - \mu_j = m(t,j) + m'(t,j)$, $t > 1$, where $Em^2(t,j) \le C(|t-j|/t)^{2\beta}$, for $|t-j| \le t/2$, and $\max_{j:|t-j|\le k} |m'(t,j)| = O_p((k/t)^\beta)$, as $t \to \infty$, for $\log t << k \le t/2$.*

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

$$\mu_t = t^{-\beta}(\nu_1 + \cdots + \nu_t) + t^{-1}(m_1 + \cdots + m_t), \tag{2.26}$$

where ν_j's are stationary zero mean r.v.'s such that $E(\nu_1 + \cdots + \nu_k)^2 \le Ck^{2\beta}$, and m_j's are non-random numbers such that $\max_j |m_j| < \infty$. It covers the constant case $\mu_t = t^{-1}(\mu + \cdots + \mu) = \mu$, the non-random case $\mu_t = g(t/n)$, and the persistent random case where $\mu_t = t^{-1/2}\sum_{j=1}^t \nu_j$, where ν_j is an i.i.d. noise or weakly dependent stationary process. Persistent processes μ_t, satisfying Assumption 2.3(ii), arise analyzing an AR(1) model with intercept in Section 2.4.

The process y_t of (2.25) is *bounded* in the sense that $\max_j Ey_j^2 < \infty$, which in turn, implies that

$$\max_j P(|y_j| \ge K) \le K^{-2} \max_j Ey_j^2 \to 0, \quad K \to \infty.$$

To verify the first claim, notice, that $Ey_j^2 = E\big(\mu_j + \{y_j - \mu_j\}\big)^2 \le 2E\mu_j^2 + 2E(y_j - \mu_j)^2$. According to (2.25), $y_t - \mu_t$ follows AR(1) model with no intercept, and therefore by (2.13), $\max_j E(y_t - \mu_t)^2 < \infty$. Together with Assumption 2.3, this implies $\max_j Ey_j^2 < \infty$.

To evaluate μ_t, ρ_t and α_t, we use estimates $\bar{y}_t$,

$$\hat{\rho}_{n,t} := \frac{\sum_{k=1}^n b_{tk}(y_k - \bar{y}_t)(y_{k-1} - \bar{y}_t)}{\sum_{k=1}^n b_{tk}(y_{k-1} - \bar{y}_t)^2}, \quad \widehat{\alpha}_t = \bar{y}_t - \widehat{\rho}_{n,t}\bar{y}_t. \tag{2.27}$$

We also use notation $y'_k = y_k - \mu_k$, $k \ge 1$, $\widehat{\sigma}^2_{Y',t} = \sum_{k=1}^n b_{tk} y'^2_{k-1}$ and $\widehat{\sigma}^2_{Y'u,t} = \sum_{k=1}^n b^2_{tk} y'^2_{k-1} u_k^2$.

The next theorem provides conditions for inference about μ_t ρ_t and α_t. It assumes, that persistency parameter β of Assumption 2.3 is not smaller than the parameter γ of Assumption 2.2.

Theorem 2.4. *Let y_j be defined as in (2.25), and $t = [n\tau]$, where $0 < \tau < 1$. Assume that H and K satisfy (2.18) and (2.20), Assumptions 2.1 and 2.2 hold with the parameter γ, and μ_t satisfies Assumption 2.3 with the parameter $\beta \ge \gamma$.*

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

$$\bar{y}_t - \mu_t = O_p(\kappa_n), \quad \widehat{\rho}_{n,t} - \rho_t = O_p(\kappa_n), \quad \widehat{\alpha}_{n,t} - \alpha_t = O_p(\kappa_n). \tag{2.28}$$

(ii) *Moreover, if $(\bar{H}/n)^\gamma = o(H^{-1/2})$, and $E[u_1^2|u_0, u_{-1}, \cdots] \ge c > 0$, for some $c > 0$, then*

$$\frac{B_{1t}}{B_{2t}} \frac{(1-\rho_t)}{\sigma_u}\big(\bar{y}_t - \mu_t\big) \to_D N(0,1), \quad \frac{\widehat{\sigma}^2_{Y',t}}{\widehat{\sigma}_{Y'u,t}}\Big(\widehat{\rho}_{n,t} - \rho_t\Big) \to_D N(0,1), \tag{2.29}$$
$$\frac{B_{1t}}{B_{3t}}\Big(\widehat{\alpha}_{n,t} - \alpha_t\Big) \to_D N(0,1), \quad B_{3t}^2 := \sum_{j=1}^n b_{nj}^2\Big(1 - \mu_t \frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}} y'_{j-1}\Big)^2 u_j^2,$$

where $B_{1t}/B_{3t} = O_p(H^{1/2})$ and $B_{3t}/B_{1t} = O_p(H^{-1/2})$.

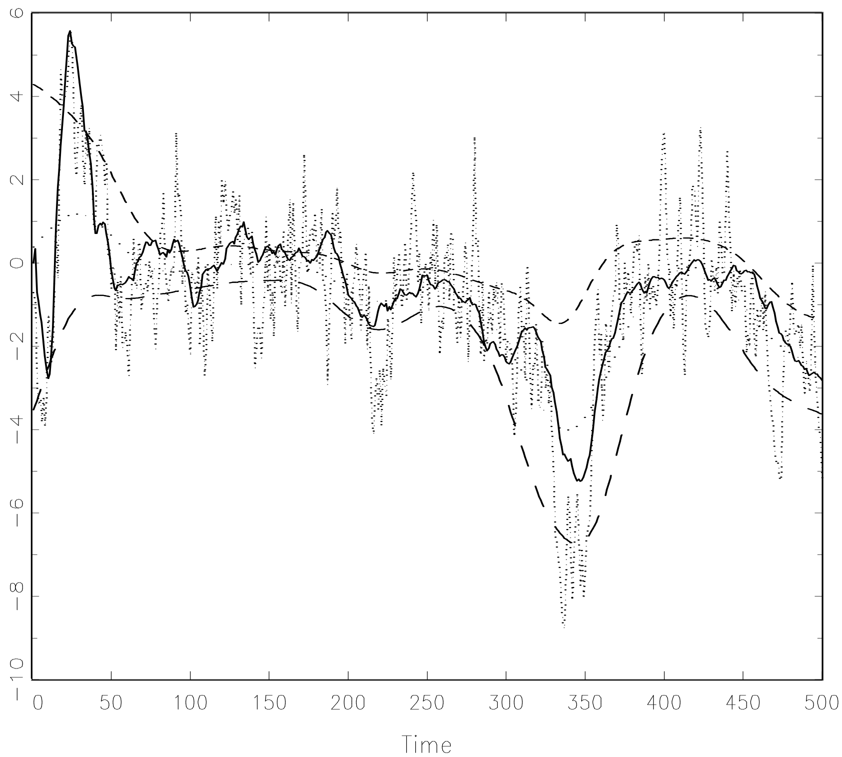
To make asymptotic normality operational, studentization in (2.29) has to be based on the residuals $\hat{u}_k = y_k - \hat{\rho}_t y_{k-1} - \widehat{\alpha}_t$, $k = 1, \cdots, n$. Set $\widehat{\sigma}^2_{\hat{Y}\hat{u},t} := \sum_{k=1}^n b^2_{tk}(y_{k-1} - \bar{y}_t)^2 \hat{u}_k^2$ and $\widehat{\sigma}^2_{\hat{Y},t} := \sum_{k=1}^n b_{tk}(y_{k-1} - \bar{y}_t)^2$.

Corollary 2.3. *Let assumptions of Theorem 2.4(ii) hold and $\max_j E\mu_j^4 < \infty$. Then*

$$\Big|\frac{1 - \widehat{\rho}_{n,t}}{1 + \widehat{\rho}_{n,t}}\Big|^{1/2} \frac{B_{1t}^{3/2}}{B_{2t}\,\widehat{\sigma}_{\hat{Y},t}}(\bar{y}_t - \mu_t) \to_D N(0,1), \quad \frac{\widehat{\sigma}^2_{\hat{Y},t}}{\widehat{\sigma}_{\hat{Y}\hat{u},t}}\Big(\hat{\rho}_{n,t} - \rho_t\Big) \to_D N(0,1), \tag{2.30}$$
$$\frac{B_{1t}}{\widehat{B}_{3t}}\Big(\widehat{\alpha}_{n,t} - \alpha_t\Big) \to_D N(0,1), \quad \widehat{B}_{3t}^2 := \sum_{j=1}^n b_{nj}^2\Big(1 - \bar{y}_t \frac{B_{1t}}{\widehat{\sigma}^2_{\hat{Y},t}}(y_{j-1} - \bar{y}_t)\Big)^2 \hat{u}_j^2.$$

The results of this section show, that a bounded non-stationary process y_t combining a persistent attractor μ_t, and a dynamic AR(1) component $y_t - \mu_t$, allows extraction of μ_t and

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



ρ_t together with their confidence bands. It presents a useful tool for inference, analysis and modeling of dynamics of bounded non-stationary processes.

To justify the use of the attractor terminology, we include Figure 2 showing a realization of y_t and μ_t, based on random walk models for ρ_t and μ_t, for a sample size of 500, the estimate $\bar{y}_t$ based on a normal kernel and a bandwidth $H = \sqrt{n}$, and 90% confidence intervals for $\bar{y}_t$. It shows that the process 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

In this section we discuss an AR model $y_t = \alpha_t + \rho_{t-1} y_{t-1} + u_t$, $t \geq 1$ with a persistent and possibly random intercept α_t. Using similar algebraic relations as in (2.12), y_t can be written as the sum

$$y_t = \rho_{t-1} y_{t-1} + \{\alpha_t + u_t\} = \{\sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i}\} + \{\sum_{i=0}^{t-1} c_{t,i}u_{t-i} + c_{t,t}y_0\} =: \mu_t + y'_t, \qquad (2.31)$$

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

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 processes α_t, allowing inference about AR(1) model with intercept.

Assumption 2.4. *α_t is $\mathcal{F}_t = \sigma(u_j, a_j,\, j \leq t)$ measurable, $\max_j E\alpha_j^2 < \infty$, and for some $\beta \in (0, 1]$, $E(\alpha_t - \alpha_{t+k})^2 \leq C(k/t)^{2\beta}$, $t \geq 1$, $1 \leq k \leq t$.*

For a standard example of the process α_t satisfying Assumption 2.4, see (2.26). Next we show that if α_t satisfies Assumption 2.4, then μ_t satisfies Assumption 2.3.

Proposition 2.2. *Suppose that a_t satisfies Assumptions 2.1 and 2.2 with parameter γ, and α_t satisfies Assumption 2.4 with parameter $\beta \geq \gamma$. Then μ_t in (2.31) satisfies Assumption 2.3(ii) with $\beta = \gamma$.*

Moreover, as $t \to \infty$,

$$\mu_t = (1 - \rho_t)^{-1}\alpha_t + o_p(1). \qquad (2.32)$$

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

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

using restriction $\rho_t = \rho a_t / \max_{0 \leq j \leq t} |a_j|$, bounding 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 5.1 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 (2.3) with $\gamma = 1/2$. It is evident that, for $\phi = 0$ and "optimal" bandwidth $H = n^{0.5}$, the average MSE falls substantially with 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 5.2 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 (2.3) with parameter $\gamma = d - 1/2$, taking values $\gamma = 0.01, 0.25, 0.75, 0.99$, and 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 of and inference for ρ_t, as suggested by the theory.

The first part of the simulation analysis is focused on the AR(1) model y_t with i.i.d. errors u_t and no intercept. Our theory allows for a model

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

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_j on estimation of ρ_t, we set $\alpha_t = 0$ in y_t, and 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 normalised to have unit variance. Table 5.3 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 5.3 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 in studentization, see e.g. (2.22), but in general, it does not affected estimation, as suggested by 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 contribute new evidence to 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.

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 *i-th* 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 (*i-th* currency units per one unit of the *j-th* currency), $p_{j,t}$ the price level (CPI) in the *j-th* country, and $p_{i,t}$ the price level of the *i-th* country. That is, a rise in $q_{i,t}$ implies a real appreciation of the *j-th* country's currency against the *i-th* country's currency.

We consider a model whereby 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. We estimate the model using the kernel estimators presented in Section 2 but having obtained similar results for both kernels, choose to report results only for the normal kernel due to space considerations. We use a bandwidth H equal to $n^{1/2}$ as suggested by theory. Results are reported pictorially in Figures 3-4. Figure 3 relates to CPI inflation and Figure 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 standard errors.

4.2 Empirical Results

The empirical results presented in Figures 3-4 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, Sargent, and Primiceri (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: 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, Eichenbaum, and Evans (2005) and Smets and Wouters (2003) and the more likely it is that this persistence reflects changes in the monetary regime, something more plausibly viewed as subject to time-variation, as policymakers, or their conclusions about what constitutes good policies, change. 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.

Figure 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 alow 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 of significant time-variation in inflation persistence 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, Arnone, and Laurens (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), Papell and Theodoridis (2001), Chortareas, Kapetanios, and Shin (2002) and Chortareas and Kapetanios (2009). One reason that persistent deviations from PPP can open up is because of nominal rigidities. But Chari, Kehoe, and McGrattan (2002) note that this persistence in the data - they report an autoregressive coefficient of around 0.8 for 8 U.S 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, Mumtaz, Ravn, and Rey (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 Figure 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 these 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, that real exchange rate persistence might have been a function solely of nominal rigidities was already strongly at odds with plausible sticky price models; our findings that this persistence is time-varying serve to emphasise this. Consistent with our study of inflation persistence, one is naturally led therefore to conjecture that this persistence in the real exchange rate must have some source that is more plausibly time varying - in the dynamics of monetary or fiscal policy. For example, recalling the paper by Benigno (2002) already cited above, one might imagine that inertia in monetary policy rules could have changed.

4.3 Concluding Remarks

This paper has proposed a new class of time-varying coefficient models, allowing the decomposition of a nonstationary 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 interesting and 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 summarising 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 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. This case is strengthened by our Monte Carlo evidence which shows that these estimators work well in small samples.

5 Appendix

5.1 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 (2.20):

$$B_{1t} := \sum_{k=1}^{n} b_{tk} \sim H \int K(x)dx = H, \quad B_{2t}^2 := \sum_{k=1}^{n} b_{tk}^2 \sim H \int K^2(x)dx =: H\beta_K. \tag{5.1}$$

Proof of Theorem 2.1. (i) Equation of (2.12) follows using recursions

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

The above recursion also implies

$$y_t = \sum_{j=0}^{k} c_{t,j}u_{t-j} + c_{t,k+1}y_{t-k-1}, \quad (1 \le k \le t-1). \tag{5.2}$$

(ii) To prove (2.13), use $|c_{t,j}| \le |\rho|^j$, $E|u_j u_k| \le Eu_1^2$, and (2.12), to obtain

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

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

$$Ey_t^4 \le E\Big(\sum_{j=0}^{t-1}\rho^j|u_{t-j}| + \rho^t y_0\Big)^4 \le 4((\sum_{j=0}^{t-1}\rho^j)^4 Eu_1^4 + \rho^{4t}Ey_0^4) \le 4(1-\rho)^{-4}(Eu_1^4 + Ey_0^4). \ \Box$$

Proof of Theorem 2.2. The first equality in (2.15) holds, because (5.17) of Lemma 5.1 implies $E(y_t - z_t(\rho_t))^2 = o(1)$, as $t \to \infty$. To show the second claim of (2.15), write $y_t = z_t(\rho_{t-\log t}) + \{y_t - z_t(\rho_{t-\log t})\}$. Note that $|Ez_t(\rho_{t-\log t})| = |E\sum_{k=0}^{t-1}\rho_{t-\log t}^j u_{t-k}|$

$\le \big|E\big[\sum_{k=0}^{\log t-1}\rho_{t-\log t}^k E[u_{t-k}|F_{t-k-1}]\big]\big| + \sum_{k=\log t}^{t-1}\rho^k E|u_{t-k}| = 0 + O(\rho^{\log t}) \to 0$. Since by (5.17), $E(y_t - z_t(\rho_{t-\log t}))^2 = o(1)$, this proves $Ey_t \to 0$, as $t \to \infty$.

Proof of (2.16). Let $k \ge 0$. It suffices to show that, as $t \to \infty$,

$$E\big[y_{t+k}y_t - z_t(\rho_t)z_{t+k}(\rho_t)\big] \to 0, \quad E\big[z_t(\rho_t)z_{t+k}(\rho_t)\big] - \sigma_u^2 E[\rho_t^{|k|}(1-\rho_t^2)^{-1}\big] \to 0.$$

By (5.17), $E(y_t - z_t(\rho_t))^2 = o(1)$ and $E(y_{t+k} - z_{t+k}(\rho_t))^2 = o(1)$. By (2.13), $Ey_t^2 < \infty$ and $Ez_{t+k}^2 < \infty$. This, using Cauchy inequality, implies $E|y_{t+k}y_t - z_t(\rho_t)z_{t+k}(\rho_t)| \le E|(y_{t+k} - z_{t+k}(\rho_t))y_t| + E|z_{t+k}(\rho_t)(y_t - z_t(\rho_t))| \le C\{E(y_{t+k} - z_{t+k}(\rho_t))^2\}^{1/2} + C\{E(y_t - z_t(\rho_t))^2\}^{1/2} = o(1)$, which verifies the first claim. To show the second claim, use independence between ρ_j's and u_j's, to obtain $E[z_t(\rho_t)z_{t+k}(\rho_t)] = \sigma_u^2 E[\sum_{i=0}^{t-1}\rho_t^{k+2i}] \to \sigma_u^2 E[\rho_t^k \sum_{i=0}^{\infty}\rho_t^{2i}] = \sigma_u^2 E[\rho_t^k(1-\rho_t^2)^{-1}]$, which proves (2.16) for $k \ge 0$. For $k < 0$, (2.16) follows using the same argument.

Proof of (2.17). By (5.2), and mutual independence of ρ_t's, u_t's and y_0,

$$\begin{aligned} Ey_t^2 &= E\big(\textstyle\sum_{j=0}^{t-1} c_{t,j}u_{t-j} + c_{t,t}y_0\big)^2 = E\big(\textstyle\sum_{j=0}^{t-1} c_{t,j}u_{t-j}\big)^2 + E\big(c_{t,t}y_0\big)^2 \\ &\le \textstyle\sum_{j=0}^{t-1} E[c_{t,j}^2]\sigma_u^2 + c_{t,t}^2 Ey_0^2 \le (\sigma_u^2 + Ey_0^2)\sum_{j=0}^{\infty}\rho^{2j} \le (\sigma_u^2 + Ey_0^2)(1-\rho^2)^{-1}, \end{aligned} \tag{5.3}$$

because $E[u_j u_k] = 0$. Next, by (5.2) and property $E[u_s y_t] = E[y_t E[u_s|F_s]] = 0$, $s > t$,

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

which together with (5.3) implies (2.17). □

Proof of Theorem 2.3. (i) Using $y_j = \rho_{j-1}y_{j-1} + u_j$, and $\widehat{\sigma}_{Y,t}^2 = \sum_{j=1}^n b_{tj}y_{j-1}^2$, write

$$\widehat{\rho}_{n,t} - \rho_t = \widehat{\sigma}_{Y,t}^{-2}\sum_{j=1}^n b_{tj}y_{j-1}u_j + \widehat{\sigma}_{Y,t}^{-2}\sum_{j=1}^n b_{tj}(\rho_{j-1} - \rho_t)y_{j-1}^2 =: \widehat{\sigma}_{Y,t}^{-2}S_{Yu,t} + \widehat{\sigma}_{Y,t}^{-2}r_{n,t}. \tag{5.4}$$

By (5.21), $\widehat{\sigma}_{Y,t}^{-2} = O_p(H^{-1})$. Next, by definition of u_t, (2.13) and (5.1), $ES_{Yu,t}^2 = \sum_{j=1}^n b_{tj}^2 E[y_{j-1}^2 u_j^2] \le C\sum_{j=1}^n b_{tj}^2 \le CH$. Thus, $S_{Yu,t} = O_p(H^{1/2})$, and $\widehat{\sigma}_{Y,t}^{-2}S_{Yu,t} = O_p(H^{-1/2})$.

Next we verify that

$$r_{n,t} = O_p\big((\bar{H}/n)^\gamma H + 1\big), \tag{5.5}$$

which implies $\widehat{\sigma}_{Y,t}^{-2}r_{n,t} = O_p\big((\bar{H}/n)^\gamma + H^{-1}\big)$, completing the proof of (2.21).

To show (5.5), let $h := b\bar{H}$ be such that (5.13) holds. Then, by (5.17) of Lemma 5.1, $R_{t,h} \equiv \max_{|t-j|\le h}|\rho_j - \rho_t| = O_p((h/t)^\gamma) = O_p((\bar{H}/n)^\gamma)$, since $t \sim \tau n$. Hence, $|r_{n,t}| \le \sum_{j=1}^n b_{tj}|\rho_{j-1} - \rho_t|y_{j-1}^2 \le R_{t,h}\sum_{|j-t|\le h} b_{tj}y_{j-1}^2 + \sum_{|j-t|>h} b_{tj}2y_{j-1}^2$. By (2.13) and (5.1), $E\sum_{|j-t|\le h} b_{tj}y_{j-1}^2 \le C\sum_{|j-t|\le h} b_{tj} \le CH$, and by (5.13), $E\sum_{|j-t|>h} b_{tj}\rho^j y_{j-1}^2 \le C\sum_{|j-t|>h} b_{tj} = o(1)$. Hence, $r_{n,t} = O_p\big((\bar{H}/n)^\gamma H + 1\big)$, completing the proof.

(ii) Under restriction $(\bar{H}/n)^\gamma = o(H^{-1/2})$, by (5.5) $r_{n,t} = o_p(H^{1/2})$, while by Lemma 5.2, $\widehat{\sigma}_{Yu,t}^{-1} = O_p(H^{-1/2})$. Hence, $\widehat{\sigma}_{Yu,t}^{-1} r_{n,t} = o_p(1)$, and by (5.4),

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

Since $\widehat{\sigma}_{Yu,t}^{-1} S_{Yu,t}$ satisfies (5.30) of Lemma 5.4, this completes the proof of (2.22). □

Proof of Corollary 2.2. In view of (2.22), it suffices to show that $\widehat{\sigma}_{Y\widehat{u},t}^2/\widehat{\sigma}_{Yu,t}^2 \to_p 1$. By Lemma 5.2, $\widehat{\sigma}_{Yu,t}^{-2} = O_p(H^{-1})$. Whence, it remains to prove that

$$\widehat{\sigma}_{Y\widehat{u},t}^2 - \widehat{\sigma}_{Yu,t}^2 = o_p(H), \tag{5.6}$$

which implies $|\widehat{\sigma}_{Y\widehat{u},t}^2/\widehat{\sigma}_{Yu,t}^2 - 1| = \widehat{\sigma}_{Yu,t}^{-2}|\widehat{\sigma}_{Y\widehat{u},t}^2 - \widehat{\sigma}_{Yu,t}^2| = o_p(1)$, completing the proof.

Since $\widehat{u}_j^2 - u_j^2 = (\widehat{u}_j - u_j)^2 + 2(\widehat{u}_j - u_j)u_j$, setting $q_n^2 := \sum_{j=1}^n b_{tj}^2 y_{j-1}^2 (\widehat{u}_j - u_j)^2$, and using Cauchy inequality, we obtain

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

By Lemma 5.2, $\widehat{\sigma}_{Yu,t} = O_p(H^{1/2})$. Hence, to prove (5.6), it suffices to verify that

$$q_n^2 = o_p(H). \tag{5.7}$$

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

$$q_n^2 \le 2\textstyle\sum_{j=1}^n b_{tj}^2 y_{j-1}^4 (\rho_{j-1} - \rho_t)^2 + 2(\rho_t - \hat{\rho}_{n,t})^2 \sum_{j=1}^n b_{tj}^2 y_{j-1}^4 =: 2(q_{n,1} + q_{n,2}). \tag{5.8}$$

To bound $q_{n,1}$, note that $|\rho_j| \le \rho$, $\max_j b_{tj} < \infty$ and, by (2.13), $\max_j Ey_j^4 < \infty$. Hence, by the same argument as in the proof of (5.5), it follows $q_{n,1} \le C \sum_{j=1}^n b_{tj} |\rho_t - \rho_t| y_{j-1}^4 = O_p\big((\bar{H}/n)^\gamma\big)H + 1\big) = o_p(H)$.

To bound $q_{n,2}$, note that by (2.21), $\rho_t - \hat{\rho}_{n,t} = o_p(1)$, while $\sum_{j=1}^n b_{tj}^2 Ey_{j-1}^4 \le C \sum_{j=1}^n b_{tj} = O(H)$, by (5.1). This implies $q_{n,2} = o_p(H)$, which proves (5.7). □

Proof of Proposition 2.1. The claim (2.23) in shown in (5.39) of Lemma 5.5. In addition, if $(\bar{H}/n)^\gamma = o(H^{-1/2})$, then (5.39) implies $(1 - \rho_t)\bar{y}_t = \bar{u}_t + o_p\big(H^{1/2}\big)$, which together with (5.40) implies (2.24). □

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

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

By (5.1), $B_{1t} \sim H$, which together with (5.43) of Lemma 5.6 implies $r_{1t} = O_p\big((\bar{H}/n)^\beta + H^{-1}\big)$. By Lemma 5.5, $\bar{y}'_t = O_p\big((\bar{H}/n)^\gamma + H^{-1/2}\big)$, which proves (2.28), noting that $\beta \ge \gamma$. In addition, if $(\bar H/n)^\gamma = o(H^{-1/2})$, then $r_{1t} = o_p(H^{-1/2})$, and $\bar y_t - \mu_t = \bar y'_t + o_p(H^{-1/2})$. Since $\bar y'_t$ satisfies (2.24), this proves (2.29).

Next we prove claims about $\widehat\rho_{n,t}$. Denote $S_{\widehat Y\widehat Y,t} := \sum_{j=1}^n b_{tj}(y_j - \bar y_t)(y_{j-1} - \bar y_t)$ and $S_{Y'Y',t} := \sum_{j=1}^n b_{tj} y'_j y'_{j-1}$, $\hat\sigma^2_{\hat Y,t} = \sum_{j=1}^n b_{tj}(y_{j-1} - \bar y_t)^2$ and $\hat\sigma^2_{Y',t} = \sum_{j=1}^n b_{tj} y'^2_{j-1}$. By definition (2.27), $\widehat\rho_{n,t} = S_{\hat Y\hat Y,t}/\widehat\sigma^2_{\hat Y,t}$, while the quantity $\widetilde\rho_{n,t} := S_{Y'Y',t}/\widehat\sigma^2_{Y',t}$ is the estimator of ρ_t in the model with no intercept.

Hence, by (5.44) of Lemma 5.6, with $\xi_n := O_p\big((\bar H/n)^\gamma H + 1\big)$,

$$
\begin{aligned}
|\widehat\rho_{n,t} - \widetilde\rho_{n,t}| = \Big|\frac{S_{\widehat Y\widehat Y,t}}{\widehat\sigma^2_{\hat Y,t}} - \frac{S_{Y'Y',t}}{\widehat\sigma^2_{Y',t}}\Big| &= \Big|\frac{S_{Y'Y',t} + \xi_n}{\widehat\sigma^2_{Y',t} + \xi_n} - \frac{S_{Y'Y',t}}{\widehat\sigma^2_{Y',t}}\Big| \le \frac{|\xi_n|\{|S_{Y'Y',t}/\widehat\sigma^2_{Y',t}| + 1\}}{\widehat\sigma^2_{Y',t} + \xi_n} \\
&= O_p(|\xi_n| H^{-1}) = O_p\big((\bar H/n)^\gamma + H^{-1}\big), \qquad (5.9)
\end{aligned}
$$

since $|\widetilde\rho_{n,t}| = |S_{Y'Y',t}/\widehat\sigma^2_{Y',t}| = O_p(1)$ by (2.21), and $\widehat\sigma^2_{Y',t} \ge cH$ for some $c > 0$ by (5.21).

Hence, $\hat\rho_{n,t} = \widetilde\rho_{n,t} + O_p\big((\bar H/n)^\gamma + H^{-1}\big)$. Since $\widetilde\rho_{n,t}$ satisfies (2.21), this implies (2.28), while for $(\bar H/n)^\gamma = o(H^{-1/2})$, $\widetilde\rho_{n,t}$ satisfies (2.22) which implies (2.29).

Finally, we prove claims about $\widehat\alpha_{n,t}$ which is an estimate of $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$. Recall that $y'_t = y_t - \mu_t$ denotes AR(1) process with no intercept and $\widetilde\rho_{n,t} = \widehat\sigma^{-2}_{Y',t}\sum_{j=1}^n b_{nj} y'_j y'_{j-1}$ is the estimator of the parameter ρ_t in such model. Let $\delta_n := O_p((\bar H/n)^\gamma) + o_p(H^{-1/2})$. Then

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

by (2.23) of Proposition 2.1 and (5.43), noting that $\beta \ge \gamma$. Moreover, by (5.9), $\widehat\rho_{n,t} - \widetilde\rho_{n,t} = \delta_n$, while by (5.40), $\bar u_t = O_p(H^{-1/2})$, and by (2.29) $\widehat\rho_{n,t} - \rho_t = O_p(H^{-1/2}) + \delta_n$. Hence,

$$
\begin{aligned}
\widehat\alpha_{n,t} = \bar y_t(1 - \widehat\rho_{n,t}) &= \Big(\mu_t + \bar u_t(1-\rho_t)^{-1} + \delta_n\Big)(\{1-\rho_t\} + \{\rho_t - \widehat\rho_{n,t}\}) \qquad (5.10) \\
&= \mu_t(1-\rho_t) + \bar u_t + \mu_t(\rho_t - \widehat\rho_{n,t}) + \delta_n = \mu_t(1-\rho_t) + \bar u_t + \mu_t(\rho_t - \widetilde\rho_{n,t}) + \delta_n.
\end{aligned}
$$

Assumption 2.3 and (5.17) straightforwardly imply that $\alpha - \mu_t(1-\rho_t) = \delta_n$. 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, above bounds imply $Z_{n,t} = O_p(H^{-1/2}) + \delta_n$, which proves (2.28). 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 (5.31), and $B_{1t}/B_{3t} = O_p(H^{1/2})$ by (5.22), this implies (2.29) for $\widehat\alpha_{n,t}$ and completes the proof. □

Proof of Corollary 2.3. To prove the first claim, notice that by Lemma 5.6 and Lemma 5.2, $\widehat\sigma^2_{\hat Y,t} = \widehat\sigma^2_{Y',t} + o_p(H) = Hv^2_{1,t} + o_p(H)$, where $v^2_{1,t} = \sigma^2_u(1-\rho^2_t)^{-1}$, and $\widehat\rho_{n,t} - \rho_t = o_p(1)$ by (2.28). Since $B_{1t} \sim H$ by (5.1), this implies

$$
\Big|\frac{1-\hat\rho_{n,t}}{1+\hat\rho_{n,t}}\Big|^{1/2}\frac{B_{1t}^{3/2}}{B_{2t}\widehat\sigma_{\hat Y,t}} = \Big|\frac{1-\rho_t}{1+\rho_t}\Big|^{1/2}\frac{B_{1t}(1-\rho_t^2)^{1/2}}{B_{2t}\sigma_u}(1+o_p(1)) = \frac{B_{1t}}{B_{2t}}\frac{(1-\rho_t)}{\sigma_u}(1+o_p(1)),
$$

which, in view of (2.29), proves the the first claim in (2.30).

To prove the second and third claims in (2.30), in view of (2.29), it suffices to show that

$$
(\widehat\sigma^2_{\hat Y,t}/\widehat\sigma_{\hat Y\hat u,t})/(\widehat\sigma^2_{Y',t}/\widehat\sigma_{Y'u,t}) \to_p 1, \quad \hat B^2_{3t}/B^2_{3t} \to_p 1. \qquad (5.11)
$$

To prove the first result, note that by Lemma 5.6, $\widehat{\sigma}^2_{\hat{Y},t} = \widehat{\sigma}^2_{Y',t} + o_p(H)$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u},t} = \widehat{\sigma}^2_{Y'u,t} + o_p(H)$, while by (5.21) $\widehat{\sigma}^{-2}_{Y,t} = O_p(H^{-1})$. 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 (5.11).

To prove the second result, note that $B_{3t}^{-2} = O_p(H^{-1})$ by Lemma 5.2. We shall show that $|\widehat{B}^2_{3t} - B^2_{3t}| = o_P(H)$, which implies the required bound: $|\widehat{B}^2_{3t}/B^2_{3t} - 1| = B_{3t}^{-2}|\widehat{B}^2_{3t} - B^2_{3t}| = o_p(1)$. Let $\widehat{\sigma}^2_{Y'uu,t} := \sum_{j=1}^n b_{tj} y'_{j-1} u_j^2$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u},t} := \sum_{j=1}^n b_{tj} \hat{y}_{j-1} \hat{u}_j^2$. Then

$$B^2_{3t} - \widehat{B}^2_{3t} = -2\mu_t \frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}} \widehat{\sigma}^2_{Y'uu,t} + 2\bar{y}_t \frac{B_{1t}}{\widehat{\sigma}^2_{\hat{Y},t}} \widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u}} + \mu_t^2 \frac{B^2_{1t}}{\widehat{\sigma}^4_{Y',t}} \widehat{\sigma}^2_{Y'u} - \bar{y}_t^2 \frac{B^2_{1t}}{\widehat{\sigma}^4_{\hat{Y},t}} \widehat{\sigma}^2_{\hat{Y}\hat{u}}. \tag{5.12}$$

Recall the above bounds for $\widehat{\sigma}^2_{Y',t}$, $\widehat{\sigma}^2_{\hat{Y},t}$ and $\widehat{\sigma}^2_{\hat{Y}\hat{u},t}$. By (2.28) $\bar{y}_t = \mu_t + o_p(1)$, by Lemma 5.6 $\widehat{\sigma}^2_{\hat{Y}\hat{u}\hat{u}} = \widehat{\sigma}^2_{Y'uu} + o_p(H) = O_p(H)$ and $B_{1t} \sim H$. Hence,

$$|B^2_{3t} - \widehat{B}^2_{3t}| \le o_p(H)\big|\mu_t \frac{B_{1t}}{\widehat{\sigma}^2_{Y',t}} \widehat{\sigma}^2_{Y'uu,t}\big| + o_p(1)\bar{y}_t^2 \frac{B^2_{1t}}{\widehat{\sigma}^4_{\hat{Y},t}} \widehat{\sigma}^2_{\hat{Y}\hat{u}} = o_p(H),$$

completing the proof of (5.11). □

Proof of Proposition 2.2. To verify Assumption 2.3(ii) for μ_t, write

$$\mu_{t+j} - \mu_t = \textstyle\sum_{i=0}^{t+j-1} c_{t+j,i}\alpha_{t+j-i} - \sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i} = m(t,j) + m'(t,j),$$

where $m'(t,j) := \sum_{i=0}^{t-1}(c_{t+j,i} - c_{t,i})\alpha_{t-i}$, and $m(t,j) = \mu_{t+j} - \mu_t - m'(t,j)$. It remains to show that $m(t,j)$ and $m'(t,j)$ satisfy conditions of Assumption 2.3(ii) with parameter $\beta = \gamma$.

To bound $m(t,j)$, recall that $|c_{t,i}| \le \rho^i$, while by Assumption 2.4, $\max_i E|\alpha_i| < \infty$ and $E|\alpha_{t+j-i} - a_{t-i}| \le C(j/(t-i))^\beta \le C(j/t)^\beta$ for $i \le t/2$. Therefore,

$$E|m(t,j)| \le C\big(\textstyle\sum_{i=t/2}^{t+j-1} \rho^i + \sum_{i=1}^{t/2} \rho^i E|\alpha_{t+j-i} - a_{t-i}|\big) \le C(\rho^{[t/2]} + |j/t|^\beta) \le C|j/t|^\beta,$$

since $\rho < 1$, which verifies the required bound for $m(t,j)$.

Next, for $\log t << k < t/2$, write $m'(t,j) = \sum_{i=0}^{k}[\cdots] + \sum_{i=k+1}^{t-1}[\cdots] =: q_{1tj} + q_{2tj}$. To bound q_{1tj}, note that by (5.14), for $i = 1, \cdots, k$, $|c_{t+j,i} - c_t^i| \le 3i\rho^{i-1}R_{t,k}$, where $R_{t,k} = \max_{j:|j-t|\le h} |\rho_t - \rho_j| = O_p((k/t)^\gamma)$, by (5.17). Hence, with $r_n := \sum_{i=0}^k \rho^{i-1} i|\alpha_{t-i}|$, $\max_{|j-t|\le k} |q_{1tj}| = O_p((k/t)^\gamma) r_n = O_p((k/t)^\gamma)$, because $Er_n \le C\sum_{i=0}^\infty \rho^{i-1} i < \infty$ implies $r_n = O_p(1)$.

On the other hand, $E\max_{|j-t|\le k} |q_{2tj}| \le \sum_{i=k+1}^{t-1} \rho^i \le C\rho^k = O(t^{-1}) = O((k/t)^\gamma)$, because $k/\log t \to \infty$, which completes verification of Assumption 2.3(ii) with $\beta = \gamma$.

Proof of (2.32). Let $m = \log t$. Write $|\mu_t - (1-\rho_t)^{-1}\alpha_t| = |\sum_{i=0}^{t-1} c_{t,i}\alpha_{t-i} - \sum_{i=0}^\infty \rho_t^i \alpha_t|$ $\le \sum_{i=0}^m |(c_{t,i} - \rho_t^i)\alpha_{t-i}| + \sum_{i=0}^m |\rho_t^i(\alpha_{t-i} - \alpha_t)| + \{|\sum_{i=m+1}^{t-1} c_{t,i}\alpha_{t-i}| + |\sum_{i=m+1}^\infty \rho_t^i \alpha_t|\} =: r_{1t} + r_{2t} + r_{3t}$. It remains to show that $r_{lt} = o_p(1)$, for $l = 1, 2, 3$.

Using the same argument as above, $\max_{1\le i\le m} |c_{t,i} - \rho_t^i| \le i\rho^{i-1} R_{t,m} = O_p((m/t)^\gamma)$. Thus, $|r_{1t}| = O_p((m/t)^\gamma) \sum_{i=0}^m i\rho^{i-1}|\alpha_{t-i}| = O_p((m/t)^\gamma) = o_p(1)$.

Next, by Assumption 2.4, $E|\alpha_{t-i} - \alpha_t| \le C(i/(t-i))^\gamma \le C(i/t)^\gamma$, for $1 \le i \le m$. Thus, $E|r_{2t}| \le Ct^{-\gamma} \sum_{i=0}^m \rho^i i^\gamma \to 0$, and so $r_{2t} = o_p(1)$.

Finally, $E|r_{3t}| \le \sum_{i=m+1}^{t-1} \rho^i E|\alpha_{t-i}| + \sum_{i=m+1}^\infty \rho^i E|\alpha_i| \le C\sum_{i=m+1}^\infty \rho^i \le C\rho^m \to 0$, and therefore, $r_{3t} = o_p(1)$, which completes the proof. □

5.2 Auxiliary results

In this section we provide auxiliary results used to prove the main theorems.

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

$$\sum_{1 \le j \le n,\ |t-j| \ge b\bar{H}} b_{tj} = o(1). \tag{5.13}$$

Indeed, if kernel K has finite support, then $K(x) = 0$, $|x| \ge p$ for some finite $p > 0$. If K has infinite support, use b such that $b^2 c \ge 1$, where c is the same as in $K(x) = O(e^{-cx^2})$ in (2.20). Then, $\sum_{|t-j|\ge b\bar{H}} b_{tj} \le C \int_{b\bar{H}}^{\infty} e^{-c(x/H)^2} dx \le b\bar{H}^{-1} \int_{b\bar{H}}^{\infty} e^{-c(x/H)^2} x dx \le CH\bar{H}^{-1} = o(1)$.

We shall use notation $R_{t,h} := \max_{j:|j-t|\le h} |\rho_t - \rho_j|$.

Lemma 5.1. *(i) For $t \ge 1$, and $1 \le j, s \le t-1$,*

$$|c_{t,j} - \rho_t^j| \le \rho^{j-1} j \max_{k=1,\cdots,j} |\rho_t - \rho_{t-k}|, \quad |\rho_t^j - \rho_{t-s}^j| \le \rho^{j-1} j |\rho_t - \rho_{t-s}|. \tag{5.14}$$

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

$$|\rho_{t+j} - \rho_t| \le 2\rho |a_{t_0}|^{-1} \max_{k=1,\cdots,j} |a_{t+k} - a_t|. \tag{5.15}$$

(iii) Let $t, j \ge 1$ and $h, m \ge 1$ be such that $h + m < t$. Then there exist random variables $Y_{j,m}$, such that for $|j - t| \le h$,

$$|y_j - z_j(\rho_t)| \le (R_{t,h+m} + \rho^m) Y_{j,m}, \quad EY_{j,m}^4 \le C, \tag{5.16}$$

where C does not depend on t, h, m and j.

In addition, if $t \to \infty$, $h \to \infty$ and $h = o(t)$, and Assumptions 2.1 and 2.2 hold, then

$$R_{t,h} = O_p\big((h/t)^\gamma\big), \qquad \max_{j:|t-j|\le \log t} E(y_j - z_j(\rho_t))^2 = o(1). \tag{5.17}$$

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

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

which proves the first claim of (5.14), while the second claim follows by the same argument.

(ii) To prove (5.15), 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

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

We show that

$$|m_{t+j} - m_t| \le \max_{k=1,\cdots,j} |a_{t+k} - a_t|, \tag{5.18}$$

which completes the proof of (5.15). Let $m_{t+j} = |a_{j_*}|$. If $j_* \le t$, then $m_{t+j} - m_t = 0$ and (5.18) 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 (5.18) holds.

(iii) *Proof of (5.16).* We start with the first claim. By (2.12) and (2.14),

$$|y_j - z_j(\rho_t)| = \big| \textstyle\sum_{k=1}^{j-1} (c_{j,k} - \rho_t^k) u_{j-k} + c_{j,j} y_0 \big|.$$

We show that for $|j - t| \le h$ and $k = 1, \cdots, m$,

$$|c_{j,k} - \rho_t^k| \le 3 R_{t,h+m} \rho^{k-1} k. \tag{5.19}$$

Indeed, by (5.14), $|c_{j,k} - \rho_t^k| \le |c_{j,k} - \rho_j^k| + |\rho_j^k - \rho_t^k| \le \rho^{k-1} k \{\max_{i=1,\cdots,k} |\rho_j - \rho_{j-i}| + |\rho_j - \rho_t|\} \le 3\rho^{k-1} k R_{t,h+m}$. The last inequality follows by definition of $R_{t,h}$, noting that $|\rho_j - \rho_{j-k}| \le |\rho_t - \rho_j| + |\rho_t - \rho_{j-k}|$, and recalling that $|t - j| \le h$ and $|t - (j - k)| \le |t - j| + k \le h + m$. Notice that the latter restrictions imply $m \le t - h \le j$.

Use (5.19), and $|c_{j,k}| \le \rho^k$, $|\rho_t|^k \le \rho^k$, to obtain

$$\begin{aligned} |\textstyle\sum_{k=1}^{m-1} (c_{j,k} - \rho_t^k) u_{j-k}| &\le R_{t,h+m} y_{j,m,1}, \quad y_{j,m,1} := 3 \textstyle\sum_{k=1}^{m} k \rho^{k-1} |u_{j-k}|, \\ |\textstyle\sum_{k=m}^{j-1} (c_{j,k} - \rho_t^k) u_{j-k} + c_{j,j} y_0| &\le \rho^m y_{j,m,2}, \quad y_{j,m,2} := \textstyle\sum_{k=m}^{j-1} \rho^{k-m} |u_{j-k}| + |y_0|, \end{aligned}$$

which yields $|y_j - z_j(\rho_t)| \le (R_{t,h+m} + \rho^m) Y_{j,m}$, $Y_{j,m} := \max(3y_{j,m,1}, y_{j,m,2})$. This proves the first claim of (5.16). Verification of $\max_j EY_{j,m}^4 < \infty$ reduces to showing that $\max_j Ey_{j,m,l}^4 < \infty$, $l = 1, 2$, which is straightforward, using $0 < \rho < 1$, $Eu_k^4 = Eu_1^4 < \infty$ and $Ey_0^4 < \infty$. For example, $Ey_{j,m,1}^4 \le (\sum_{k=1}^{\infty} k\rho^{k-1})^4 Eu_1^4 < \infty$.

Proof of (5.17). We start with the first claim. Let $t_0 := t - h$. Then $|\rho_t - \rho_j| \le |\rho_t - \rho_{t_0}| + |\rho_j - \rho_{t_0}|$, and therefore, by (5.15),

$$R_{t,h} \le 2 \max_{k=1,\cdots,2h} |\rho_{t_0+k} - \rho_{t_0}| \le 4\rho |a_{t_0}|^{-1} \max_{k=1,\cdots,2h} |a_{t_0+k} - a_{t_0}|. \tag{5.20}$$

Under Assumption 2.2, $t_0^{-\gamma} a_{t_0} \to_D W_1 + g(1)$, which implies $|a_{t_0}|^{-1} = O_p(t_0^{-\gamma}) = O_p(t^{-\gamma})$, because W_1 has continuous distribution, and $t_0 = t - h \to \infty$.

Let $a'_j = a_j - Ea_j$ and $i_n := \max_{k=1,\cdots,2h} |a'_{t_0+k} - a'_{t_0}|$. By Assumption 2.1 and (2.4), $\max_{k=1,\cdots,2h} |a_{t_0+k} - a_{t_0}| \le i_n + \max_{k=1,\cdots,2h} |Ea_{t_0+k} - Ea_{t_0}|\} \le i_n + Ch^{\gamma}$. Stationarity of v_i's implies that $i_n =_D \max_{k=1,\cdots,2h} |a'_k - a'_0|$. Thus, by the weak convergence (2.4), as $h \to \infty$,

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

Hence, the r.h.s. of (5.20) is of order $O_p((h/t)^{\gamma})$ which proves the first claim of (5.17).

To show the second claim, use (5.16) with $h = m = \log t$, for $|j-t| \le h$, to obtain $E(y_j - z_j(\rho_t))^2 \le E[(R_{t,h+m} + \rho^m)^2 Y_{j,m}^2] \le (ER_{t,h+m}^4)^{1/2}(EY_{j,m}^4)^{1/2} + C\rho^{2m} \le C(ER_{t,2h})^{1/2} + C\rho^{2h}$, because $R_{t,2h} \le 2\rho$. Since $\rho^{2h} \to 0$, it remains to show that $ER_{t,2h} \to 0$.

Let $t_0 = t - 2h$ and $A = \{|a_{t_0}| \le t_0^{\gamma}/\log t_0\}$. Then $P(A) \to 0$, $t_0 \to \infty$. Indeed, by Assumption 2.2, $t_0^{-\gamma} a_{t_0} \to_D W_1 + g(1)$, and W_1 has continuous probability distribution. Therefore, $\forall \delta > 0$, $P(A) \le P(t_0^{-\gamma}|a_{t_0}| \le \delta) \to P(|W_1 + g(1)| \le \delta) \to 0$, as $\delta \to 0$. Hence, $E[1(A)R_{t,2h}] \le 2P(A) \to 0$, as $t \to \infty$, $\delta \to 0$.

On the complimentary event A^c, $|a_{t_0}|^{-1} \le t_0^{-\gamma}\log t_0$, and $R_{t,2h} \le 2\max_{k=1,\cdots,4h}|\rho_{t_0+k} - \rho_{t_0}|$. Thus, by (5.15), $R_{t,2h} \le 4|a_{t_0}|^{-1}\max_{k=1,\cdots,4h}|a_{t_0+k} - a_{t_0}|$. Therefore,

$$E[I(A^c)R_{t,2h}] \le t_0^{-\gamma}\log t_0 \textstyle\sum_{k=1}^{4h} E|a_{t_0+k} - a_{t_0}| \le C t_0^{-\gamma}\log^3 t_0 \to 0,$$

because under Assumption 2.1(ii) and (2.4), $E|a_{t_0+k} - a_{t_0}| = E|\sum_{i=t_0+1}^{t_0+k} v_j| + |Ea_{t_0+k} - Ea_{t_0}| \le kE|v_1| + Ck \le Ch$, and $h = \log t$. The above estimates imply $ER_{t,2h} \to 0$, as $t \to \infty$, completing the proof of the second claim in (5.17) and the lemma. $\Box$

Let $v_{1,t}^2$ and $v_{2,t}^2$ be as in Corollary 2.1, B_{3t} be as in (2.29) of Theorem 2.4. Denote $v_{3,t}^2 := \beta_K E[(1-\delta U_0(a))^2 u_1^2]|_{a=\rho_t,\,\delta=\delta_t}$, where $\delta_t := \mu_t(1-\rho_t^2)/\sigma_u^2$, and $U_0(a) = \sum_{k=0}^{\infty} a^k u_{-k}$.

Lemma 5.2. *Suppose that ρ_t, u_t, K and H satisfy assumptions of Theorem 2.3(i), μ_t satisfies Assumption 2.3 and $\bar{H} = o(n)$. Then*

$$\widehat{\sigma}_{Y,t}^2/(Hv_{1,t}^2) \to_p 1, \quad \widehat{\sigma}_{Yu,t}^2/(Hv_{2,t}^2) \to_p 1, \tag{5.21}$$

$$B_{3,t}^2/(Hv_{3,t}^2) \to_p 1, \tag{5.22}$$

where $c_1 \le v_{1,t}^2 \le c_2$, $v_{2,t}^2 \le c_2$ and $v_{3,t}^2 = O_p(1)$, for some constants $c_1, c_2 > 0$. Moreover, if $V_1 := E[u_1^2|u_0, u_{-1}, \cdots] \ge c > 0$, for some $c > 0$, then also $c_1 \le v_{2,t}^2$ and $c_1 \le v_{3,t}^2$.

Proof. First we verify claims about $v_{l,t}^2$, $l = 1, 2, 3$. The claim $c_1 \le v_{1,t}^2 \le c_2$ holds, because $|\rho_t| \le \rho$ implies $\sigma_u^2 \le v_{1,t}^2 \le \sigma_u^2(1-\rho^2)^{-1}$. Next, equality $v_{2,t}^2 = \beta_K E[U_0^2(a)u_1^2]|_{a=\rho_t}$ and assumption $V_1 \ge c > 0$ imply that $E[U_0^2(a)u_1^2] = E[U_0^2(a)V_1] \ge cE[U_0^2(a)] \ge c\sigma_u^2 > 0$, while for $|a| \le \rho$, $E[U_0^2(a)u_1^2] \le E[U_0^4(a)] + Eu_1^4 \le (1-\rho)^{-4}Eu_1^4 + Eu_1^4 < \infty$, which implies $c_1 \le v_{2,t}^2 \le c_2$. To bound $v_{3,t}^2$, note that $E[(1-\delta U_0(a))^2 u_1^2] = E[(1-\delta U_0(a))^2 V_1] \ge cE[(1-\delta U_0(a))^2] = c + \delta^2 EU_0^2(a) \ge c$, while $E[(1-\delta U_0(a))^2 u_1^2] \le 4 + 4\delta^4 E[U_0^4(a)] + Eu_1^4 \le 4 + 4\delta^4(1-\rho)^{-4}Eu_1^4 + Eu_1^4 < \infty$. Moreover, $\delta_t = O_p(1)$, because $E|\delta_t| \le E|\mu_t|2\sigma_u^{-2} < \infty$, which implies $c \le v_{3,t}^2 = O_p(1)$.

Proof of (5.21). We verify the second claim, (the first claim can be verified similarly). Let $h = b\bar{H}$ be as in (5.13), and let $m = H$, $t_0 = t - h - m$. We shall approximate y_j's by

$$z_{t_0,j} := \textstyle\sum_{k=1}^{m} \rho_{t_0}^k u_{j-k}, \quad |j-t| \le h, \tag{5.23}$$

and subsequently, $\widehat{\sigma}_{Yu,t}^2$ by $s_{t_0}^2 := \sum_{j:|t-j|\le h} b_{tj}^2 z_{t_0,j-1}^2 u_j^2$. We shall show that

$$\widehat{\sigma}_{Yu,t}^2 - s_{t_0}^2 = O_p\big((\bar{H}/n)H + 1\big) = o_p(H), \tag{5.24}$$

$$s^2_{t_0}/(Hv^2_{2,t_0}) \to_p 1, \quad v^2_{2,t_0}/v^2_{2,t} \to_p 1, \tag{5.25}$$

which implies (5.21): $(Hv^2_{2,t})^{-1}\widehat{\sigma}^2_{Yu,t} = (Hv^2_{2,t})^{-1}(\widehat{\sigma}^2_{Yu,t} - s^2_{t_0}) + (Hv^2_{2,t})^{-1}s^2_{t_0} \to_p 1$. The specific structure of $z_{t_0,j}$ and the choice of h, m allow to show the convergence (5.25), while keeping approximation error (5.24) negligible. Recall definition (5.1) of β_K.

Proof of (5.24). The result (5.16) of Lemma 5.1 remains valid for $z_{t_0,j}$, as easily seen from the proof. Namely,

$$|y_j - z_{t_0,j}| \le (R_{t,h+m} + \rho^m)Y_{j,m}, \quad \max_j EY^4_{j,m} \le C. \tag{5.26}$$

Bound $|\widehat{\sigma}^2_{Yu,t} - s^2_{t_0}| \le \sum_{j:|t-j|\le h} b^2_{tj}|y^2_{j-1} - z^2_{t,j-1}|u^2_j \le |\sum_{|j-t|>h}[\cdots]| + |\sum_{|j-t|\le h}[\cdots]| =: r_{n,1} + r_{n,2}$. By (2.13), $\max_j E[y^2_{j-1}u^2_j] \le \max_j(Ey^4_{j-1} + Eu^4_j) < \infty$. Then, by (5.13), $Er_{n,1} \le C\sum_{|j-t|\ge h} b_{tj} = o(1)$, which implies $r_{n,1} = o_p(1)$.

Next, by (5.26), for $|j-t| \le h$, $|y^2_{j-1} - z^2_{t_0,j-1}|u^2_j \le (R_{t,h+m} + \rho^m)\tau_{m,j}$, where $\tau_{m,j} = Y_{j-1,m}(|y_{j-1}| + |z_{t_0,j-1}|)u^2_j$. Hence, $r_{n,2} \le C(R_{t,h+m} + \rho^m)q_n$ where $q_n = \sum_{|j-t|\le h} b_{n,j}\tau_{m,j}$. By (5.17), $R_{t,h+m} = O_p\big(((h+m)/t)^\gamma\big) = O_p((\bar{H}/n)^\gamma)$, since $t \sim \tau n$ and $\bar{H} = o(n)$. In addition, $\max_j E\tau_{m,j} \le C\max_j E(Y^4_{j,m} + y^4_j + z^4_{t_0,j} + u^4_j) < \infty$, by (5.26), (2.13), and noting that $\max_j Ez^4_{t_0,j} < \infty$ which follows similarly as in the proof of (2.13). So, $Eq_n \le C\sum_{|j-t|\le h} b_{n,j} \le CH$, which implies $q_n = O_p(H)$ and $r_{n,2} = O_p\big((\bar{H}/n)^\gamma H + \rho^H H\big)$, proving (5.24), because $\rho^H H = O(1)$.

Proof of (5.25). Let $\bar{s}^2_{t_0} := \sum_{i,s=0}^m \rho^{i+s}_{t_0}\sum_{j:|t-j|\le h} b^2_{tj}E[u_{j-1-i}u_{j-1-s}u^2_j]$. To prove the first result of (5.25), it suffices to show

$$\bar{s}^2_{t_0} = Hv^2_{2,t_0}(1 + o_p(1)), \quad E|s^2_{t_0} - \bar{s}^2_{t_0}| = o(H). \tag{5.27}$$

To verify the first claim, note that by stationarity $E[u_{j-1-i}u_{j-1-s}u^2_j] = E[u_{-i}u_{-s}u^2_1]$. Recall that $m \to \infty$ and $|\rho_j| \le \rho < 1$. Therefore, $|\sum_{s=m+1}^\infty \rho^{2s}_{t_0}| \le \sum_{s=m+1}^\infty \rho^{2s} \to 0$, and $\sum_{j:|t-j|\le h} b^2_{tj} \sim \sum_{j=1}^n b^2_{tj} \sim H\beta_K$, by (5.13), which implies

$$\bar{s}^2_{t_0} = \sum_{i,s=0}^\infty \rho^{i+s}_{t_0}E[u_{-i}u_{-s}u^2_1]H\beta_K(1 + o_p(1)) = Hv^2_{2,t_0}(1 + o_p(1)).$$

To verify the second claim in (5.27), notice that

$$H^{-1}|s^2_{t_0} - \bar{s}^2_{t_0}| = \big|\sum_{i,s=0}^m \rho^{i+s}_{t_0}\nu_{n,is}\big| \le \sum_{i,s=0}^m \rho^{i+s}|\nu_{n,is}|,$$

where $\nu_{n,is} := H^{-1}\sum_{j:|t-j|\le h} b^2_{tj}\big(u_{j-1-i}u_{j-1-s}u^2_j - E[u_{j-1-i}u_{j-1-s}u^2_j]\big)$. We show that

$$\max_{i,s}|\nu_{n,is}| \le C, \quad \nu_{n,is} \to 0, \quad \forall s, t, \tag{5.28}$$

which by of dominated convergence theorem implies $H^{-1}|s^2_{t_0} - \bar{s}^2_{t_0}| \to 0$.

Note that $|\nu_{n,is}| \le H^{-1}\sum_{j:|t-j|\le h} b^2_{tj}2Eu^4_1 \le C$, uniformly in i and s, which proves the first claim. To show the second claim in (5.28), write $\nu_{n,is} = T_n - ET_n$, where $T_n = \sum_{j\in\mathbb{Z}} c_{nj}x_j$, $c_{nj} := H^{-1}b^2_{tj}I(|j-t| \le h)$ and $x_j := u_{j-1-i}u_{j-1-s}u^2_j$. Since u_j is stationary ergodic, then x_j is stationary ergodic. Notice, that $\sum_{j\in\mathbb{Z}}|c_{nj}| = H^{-1}\sum_{j\in\mathbb{Z}} b_{tj}^2 \le C$ by (5.1), while $\sum_{j\in\mathbb{Z}}|c_{nj}-c_{n,j-1}| = H^{-1}(\sum_{j:|t-j|\le h}|b_{nj}^2-b_{n,j-1}^2|+|b_{n,t-h}^2|+|b_{n,t+h}^2|) = o(1)$, under (2.20). Thus, by Lemma 5.3, $E|T_n - ET_n| = o(1)$, which completes the proof of (5.28) and (5.25).

Proof of the second claim of (5.25). Note that $E|u_{-i}u_{-s}u_1^2| \le Eu_1^4 < \infty$. Thus, by (5.14), $\left|v_{2,t}^2 - v_{2,t_0}^2\right| \le C\sum_{i,s=0}^{\infty}|\rho_{t_0}^{i+s}-\rho_t^{i+s}| \le C|\rho_{t_0}-\rho_t|\sum_{i,s=1}^{\infty}(i+s)\rho^{i+s-1} \le C|\rho_{t_0}-\rho_t|$. By (5.17), $|\rho_{t_0}-\rho_t| \le R_{t,h+m} = O_p(\{(h+m)/t\}^{\gamma}) = o_p(1)$, since $h+m = b\bar{H}+H = O(\bar{H})$, and $t\sim\tau n$, which completes the proof of (5.25).

Proof of (5.22). Let $G_{t,1} := \sum_{j=1}^n b_{tj}^2u_j^2$, $G_{t,2} := \sum_{j=1}^n b_{tj}^2 y'_{j-1}u_j^2$ and $G_{t,3} := \widehat{\sigma}^2_{Y'u,t}$ and $c_t^* := \mu_t(B_{1t}/\widehat{\sigma}^2_{Y',t})$. Write $B_{3t}^2 := G_{t,1} - 2c_t^*G_{t,2} + c_t^{*\,2}G_{t,3}$. Hence to show (5.22) it suffices to prove that $B_{3t}^2 = Hv_{3,t}^2(1+o_p(1))$. Recall that y'_j is an AR(1) process with no intercept. Thus, by (5.21), $G_{t,3}/(Hv_{2,t}^2) \to_p 1$, and $c_t^*/\big(\mu_t(1-\rho_t^2)\sigma_u^{-2}\big) \to_p 1$. Similarly, as in the proof of (5.21), it can be shown that $G_{t,1}/(H\sigma_u^2\beta_K) \to_p 1$, and $G_{t,2}/(H\beta_K D) \to_p 1$, where $D = E[U_0(a)u_1^2]|_{a=\rho_t} = \sum_{k=0}^{\infty}\rho_t^k E[u_{-k}u_1^2]$. Hence $B_{3t}^2 = H\beta_K\big(Eu_1^2 - 2c_t^*E[U_0(a)u_1^2] + c_t^{*\,2}E[U_0^2(a)u_1^2]\big)_{a=\rho_t}(1+o_p(1)) = Hv_{3,t}^2(1+o_p(1))$, which completes the proof of (5.22).

The following fact will be used in the proof of Lemma 5.4:

$$\widehat{\sigma}_{Y,t}^{-2} - \widehat{\sigma}_{Y,t_0}^{-2} = o_p\big(H^{-1}\big). \tag{5.29}$$

Indeed, a similar to (5.24) approximation implies, that $|\widehat{\sigma}^2_{Y,t}-\widehat{\sigma}^2_{Y,t_0}| = O_p\big((\bar{H}/n)H+1\big)$, while by (5.21), $\widehat{\sigma}_{Y,t}^{-2} = O_p(H^{-1})$ and $\widehat{\sigma}_{Y,t_0}^{-2} = O_p(H^{-1})$. Hence, $|\widehat{\sigma}_{Y,t}^{-2}-\widehat{\sigma}_{Y,t_0}^{-2}| = \sigma_{Y,t}^{-2}\sigma_{Y,t_0}^{-2}|\widehat{\sigma}^2_{Y,t}-\widehat{\sigma}^2_{Y,t_0}| = o_p\big(H^{-1}\big)$. $\Box$

We state for convenience the following result, which is shown in (Dalla, Giraitis, and Koul, 2012, Lemma 4.7).

Lemma 5.3. *Let $T_n = \sum_{j\in\mathbb{Z}} c_{nj}Y_j$, where $\{Y_j\}$ is a stationary ergodic sequence, $E|Y_1| < \infty$, and c_{nj} are real numbers such that for some $\nu_n < \infty$, $\sum_{j\in\mathbb{Z}}|c_{nj}| = O(\nu_n)$ and $\sum_{j\in\mathbb{Z}}|c_{nj} - c_{n,j-1}| = o(\nu_n)$. 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 (5.4), and of $Z_{n,t} = \bar{u}_t + \mu_t(\rho_t - \widetilde{\rho}_{n,t})$, used in (5.10). Recall definition (2.29) of B_{3t}.

Lemma 5.4. *Suppose that ρ_t, u_t, K and H satisfy assumptions of Theorem 2.3(ii) and μ_t satisfies Assumption 2.3. Then*

$$\widehat{\sigma}_{Yu,t}^{-1}S_{Yu,t} \to_D N(0,1), \tag{5.30}$$

$$B_{3t}^{-1}B_{1t}Z_{n,t} \to_D N(0,1). \tag{5.31}$$

Proof. To prove (5.30), set $m = H$, $h = b\bar{H}$ and $t_0 = t-h-m$. We shall approximate variables y_j by $z_{t_0,j}$ of (5.23), and, consequently, $S_{Yu,t}$ by $S^{(apr)}_{Yu,t} = \sum_{|j-t|\le h} b_{tj}z_{t_0,j-1}u_j$.

We show that

$$|S_{Yu,t} - S^{(apr)}_{Yu,t}| = o_p(H^{1/2}), \tag{5.32}$$

$$d_n^{-1} S_{Yu,t}^{(apr)} \quad \to_D \quad \mathcal{N}(0,1), \quad d_n^2 := H v_{2,t_0}^2. \tag{5.33}$$

By (5.21) and (5.25), $d_n^2/\widehat{\sigma}_{Yu,t}^2 \to_p 1$, and $\widehat{\sigma}_{Yu,t}^{-1} = O_p(H^{-1/2})$, which implies (5.30):

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

Proof (5.32). Similarly as in the proof of (5.24) it follows that $S_{Yu,t} - S_{Yu,t}^{(apr)} = O_p\big((\bar{H}/n)^\gamma H + 1\big) = o_p(H^{1/2})$, because $(\bar{H}/n)^\gamma = o_p(H^{-1/2})$, which proves (5.32).

Proof (5.33). Notice that d_n is $\mathcal{F}_{t_0}$ measurable and therefore, $\theta_{nj} := d_n^{-1} b_{tj} z_{t_0,j-1} u_j$ are martingale differences with respect to filtration $\mathcal{F}_j$: i.e. $E[\theta_{nj}|\mathcal{F}_j] = d_n^{-1} b_{tj} z_{t_0,j-1} E[u_j|\mathcal{F}_j] = 0$, $|j-t| \le h$.

By the central limit theorem for martingale differences, see e.g. (White, 1980, Corollary 3.1), to show asymptotic normality (5.33), it suffices to prove that

$$\sum_{|j-t|\le h} E[\theta_{nj}^2|\mathcal{F}_{j-1}] \to_p 1, \tag{5.34}$$

$$\sum_{|j-t|\le h} E\big[\theta_{nj}^2 I(|\theta_{nj}| \ge \varepsilon)\big] \to 0, \quad \forall \varepsilon > 0. \tag{5.35}$$

Note that $\sum_{|j-t|\le h} E[\theta_{nj}^2|\mathcal{F}_{j-1}] = d_n^{-2} \sum_{|j-t|\le h} b_{tj}^2 z_{t_0,j-1}^2 V_j =: d_n^{-2} q_n$, where $V_j := E[u_j^2|\mathcal{F}_{j-1}]$. Repeating the steps of the proof of (5.25), and using equality $E[u_{-i}u_{-s}V_1] = E[E[u_{-i}u_{-s}u_1^2|F_1]] = E[u_{-i}u_{-s}u_1^2]$, it follows that

$$q_n = \{\sum_{i,s=0}^{\infty} \rho_{t_0}^{i+s} E[u_{-i}u_{-s}V_1]\} H\beta_K(1 + o_p(1)) = v_{2,t_0}^2 H(1 + o_p(1)) = d_n^2(1 + o_p(1)),$$

which proves (5.34).

Next, we verify (5.35). Let $L = \log H$, and notice that $d_n^2 \ge cH$, for some $c > 0$, by Lemma 5.2. Then

$$\theta_{nj} I\big(|\theta_{nj}| \ge \varepsilon, \ |u_j| \le L\big) \le \theta_{nj}^2 \varepsilon^{-2} I(|u_j| \le L) \le d_n^{-4} b_{tj}^4 z_{t_0,j-1}^4 L^4 \varepsilon^{-2} \le C L^4 H^{-2} b_{tj} z_{t_0,j-1}^4,$$
$$|\theta_{nj}| I(|u_j| > L) \le d_n^{-2} b_{tj}^2 z_{t_0,j-1}^2 u_j^2 I(|u_j| > L) \le C H^{-1} b_{tj} z_{t_0,j-1}^2 u_j^2 I(|u_j| > L).$$

Denote by s_n the l.h.s. of (5.35) and recall, that $\max_j E z_{t_0,j}^4 < \infty$. Then,

$$\begin{aligned} E s_n &\le \textstyle\sum_{|j-t|\le h} b_{tj} \{L^4 H^{-2} E z_{t_0,j-1}^4 + H^{-1}(E z_{t_0,j-1}^4)^{1/2} (E u_j^4 I(|u_j| > L))^{1/2}\} \\ &\le o(H^{-1}) \textstyle\sum_{j=1}^n b_{tj} + O(H^{-1}) \sum_{j=1}^n b_{tj} (E u_1^4 I(|u_1| > L))^{1/2} = o(1), \end{aligned}$$

because of $L = \log H \to \infty$, stationarity of u_j and (5.1). This proves (5.35).

Proof of (5.31). Recall that $y_j' = y_j - \mu_j$ is as in (2.25). Let $c_k^* := \mu_k(B_{1t}/\widehat{\sigma}_{Y',k}^2)$, $k = t, t_0$. By definition, $B_{1t} Z_{n,t} = \sum_{j=1}^n b_{tj} u_j - c_t^* S_{Y'u,t} = \sum_{j=1}^n b_{tj}(1 - c_t^* y_{j-1}') u_j$. We shall approximate this sum by $Q^{(apr)}_{n,t} = \sum_{|j-t|\le h} b_{tj}(1 - c^*_{t_0} z_{t_0,j-1})u_j$, where t_0 and $z_{t_0,j}$ are as in the proof of (5.30). We show that

$$|B_{1t}Z_{n,t} - Q^{(apr)}_{n,t}| \quad = \quad o_p(H^{1/2}), \tag{5.36}$$

$$\tilde d_n^{-1} Q^{(apr)}_{n,t} \quad \to_D \quad \mathcal{N}(0,1), \quad \tilde d_n^2 := H v^2_{3,t_0}. \tag{5.37}$$

By (5.22), $B_{13}^{-1} = O_p(H^{-1/2})$, which implies (5.31):

$$\frac{B_{1t}Z_{n,t}}{B_{3,t}} = \frac{B_{1t}Z_{n,t} - Q^{(apr)}_{n,t}}{B_{3,t}} + \frac{Q^{(apr)}_{n,t}}{d_n}\frac{\tilde d_n}{B_{3,t}} = o_p(1) + \frac{Q^{(apr)}_{n,t}}{\tilde d_n}(1 + o_p(1)) \to_D \mathcal{N}(0,1),$$

because $\tilde d_n / B_{3,t} \to_p 1$. The latter claim follows from (5.22), noting that $v^2_{3,t_0}/v^2_{3,t} \to_p 1$, which can be shown similarly in the proof of the second claim of (5.25), noting that Assumption 2.3 implies $\mu_t - \mu_{t_0} = o_p(1)$ and that $v^2_{3,t} \ge c_1 > 0$ by Lemma 5.2.

Proof of (5.36). First, notice that with $i_n := \sum_{|j-t|>h} b_{tj}u_j$,

$$|B_{1t}Z_{n,t} - Q^{(apr)}_{n,t}| \le |i_n| + |c^*_t S_{Y'u,t} - c^*_{t_0} S^{(apr)}_{Y'u,t}| \le |i_n| + |(c^*_t - c^*_{t_0})S_{Y'u,t}| + |c^*_{t_0}(S_{Y'u,t} - S^{(apr)}_{Y'u,t})|.$$

Notice that $i_n = o_p(1)$ because $E|i_n| = o(1)$ by (5.13), and $S_{Yu,t} = O_p(H^{1/2})$ because $ES^2_{Yu,t} \le \sum_{j=1}^n b_{tj} E y^2_{j-1} u^2_j \le C \sum_{j=1}^n b_{tj} \le CH$. Since by (5.32) $S_{Yu,t} - S^{(apr)}_{Y'u,t} = o_p(H^{1/2})$, this together with (5.38) implies $|B_{1t}Z_{n,t} - Q^{(apr)}_{n,t}| = o_p(1) + o_p(1)O_P(H^{1/2}) + O_p(1)o_p(H^{1/2}) = o_p(H^{1/2})$, implying (5.36). It remains to show

$$|c^*_t - c^*_{t_0}| = o_p(1), \quad c^*_{t_0} = O_p(1). \tag{5.38}$$

Bound,

$$|c^*_t - c^*_{t_0}| \le B_{1t}\{|(\mu_t - \mu_{t_0})\widehat\sigma^{-2}_{Y',t}| + |\mu_{t_0}(\widehat\sigma^{-2}_{Y',t} - \widehat\sigma^{-2}_{Y',t_0})|\},$$

where $B_{1t} = O(H)$, and by Assumption 2.3, $\mu_{t_0} = O_p(1)$, $\mu_t - \mu_{t_0} = O_p((|t-t_0|/t)^\beta) = O_p((\bar H/n)^\beta) = o_p(1)$. Next, $\widehat\sigma^{-2}_{Y',t} = O_p(H^{-1})$ by Lemma 5.2, and $|\widehat\sigma^{-2}_{Y',t} - \widehat\sigma^{-2}_{Y',t_0}| = o_p(H^{-1})$ by (5.29). Hence, $|c^*_t - c^*_{t_0}| = O_p\big(H\{o_p(1)O_p(H^{-1}) + O_p(1)o_p(H^{-1})\}\big) = o_p(1)$. Finally, $|c^*_{t_0}| = B_{1t}|\mu_{t_0}|\widehat\sigma^{-2}_{Y',t_0} = O_p(H)O_p(1)O_p(H^{-1}) = O_p(1)$ proving (5.38) and (5.36).

Proof of (5.37). It follows by the same arguments as in the proof of (5.33) and noting that $\tilde d_n \ge H^{1/2}c$, for some $c > 0$, by Lemma 5.2. □

Let $\bar y_t$ and $\bar u_t$ be as in Proposition 2.1.

Lemma 5.5. *Let y_t satisfy assumptions of Theorem 2.3(i), and $\bar H = o(n)$. Then,*

$$\bar y_t = (1 - \rho_t)^{-1}\bar u_t + O_p\big((\bar H/n)^\gamma\big) + o_p(H^{-1/2}), \tag{5.39}$$

$$(B_{2t}\sigma_u)^{-1} B_{1t}\bar u_t \to_D N(0,1), \quad \text{if } (\bar H/n)^\gamma = o(H^{-1/2}). \tag{5.40}$$

Proof. First we prove (5.39). Let $h := b\bar H$ be as in (5.13) 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 $S'_{u,t} := \sum_{j:|t-j|\le h} b_{tj}u_j$. We shall show below that

$$S_{y,t} - (1 - \rho_{t_0})^{-1} S'_{u,t} = O_p\big((\bar H/n)^\gamma H + 1\big), \tag{5.41}$$

$$\rho_t - \rho_0 = o_p(1). \tag{5.42}$$

By (5.1), $B_{1t} \sim H$, and $E|S_{u,t}| \le CB_{1t}$, $E|S'_{u,t}| \le CB_{1t}$, while by (5.13), $E(S_{u,t} - S'_{u,t})^2 \le C\sum_{|j-t|>h} b_{tj}^2 = o(1)$, which implies $S_{u,t} = O_p(H^{1/2})$, $S'_{u,t} = O_p(H^{1/2})$ and $S_{u,t} - S'_{u,t} = o_p(H^{1/2})$. Thus, (5.41) and (5.42) imply (5.39):

$$\begin{aligned} |\bar y_t - (1-\rho_t)^{-1}\bar u_t| &= B_{1t}^{-1}|S_{y,t} - (1-\rho_t)^{-1}S_{u,t}| \le CH^{-1}\{|S_{y,t} - (1-\rho_{t_0})^{-1}S'_{u,t}| \\ &\quad + |(1-\rho_{t_0})^{-1}S'_{u,t} - (1-\rho_t)^{-1}S_{u,t}|\} = O_p((\bar H/n)^\gamma) + o_p(H^{-1/2}), \end{aligned}$$

estimating $|(1-\rho_{t_0})^{-1}S'_{u,t} - (1-\rho_t)^{-1}S_{u,t}| \le |(1-\rho_{t_0})^{-1} - (1-\rho_t)^{-1}|\,|S'_{u,t}| + (1-\rho_t)^{-1}|S_{u,t} - S'_{u,t}| \le |\rho_{t_0} - \rho_t| O_p(H^{1/2}) + o_p(H^{1/2}) = o_p(H^{1/2})$.

Proof of (5.41). Set $m = H$, and let $z_{t_0,j}$ be as in (5.23). Then,

$$\begin{aligned} |S_t - (1-\rho_{t_0})^{-1}S'_t| &\le |\sum_{j=1}^n b_{tj}y_j - \sum_{|t-j|\le h} b_{tj}z_{t_0,j}| + |\sum_{|t-j|\le h} b_{tj}z_{t_0,j} - (\sum_{k=0}^m \rho_{t_0}^k)\, S'_t| \\ &\quad + |\{\sum_{k=0}^m \rho_{t_0}^k - (1-\rho_{t_0})^{-1}\}S'_t| =: t_{n,1} + t_{n,2} + t_{n,3}. \end{aligned}$$

It suffices to show that $t_{n,j}$, $j = 1, 2, 3$ satisfy the bound (5.41).

The bound $t_{n,1} = O_p\big((\bar H/n)^\gamma H + 1\big)$ follows using the same argument in the proof of (5.24). Next we show that $E|t_{n,2}| = O(1)$, which implies $t_{n,2} = O_p(1)$. Notice, that for large n, summation in $t_{n,2}$ is carried in $j \ge t - h > m$. Hence,

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

Defining $b'_{tj} := b_{tj} I(|j-t| \le h)$, write $\theta_k = \sum_{j=2}^n (b'_{t,j+k} - b'_{tj})u_j$. Recall $\max_j b_{nj} < \infty$. Then,

$$\begin{aligned} E\theta_k^2 &= \sigma_u^2 \textstyle\sum_{j=2}^n (b'_{t,j+k} - b'_{tj})^2 \le C\sum_{j=2}^n |b'_{t,j+k} - b'_{tj}| \le Ck\sum_{j=2}^n |b'_{t,j+1} - b'_{tj}| \\ &\le Ck\textstyle\sum_{|t-j|\le h} |b_{t,j+1} - b_{tj}| + Ck\sum_{|t-j|>h} b_{tj} \le Ck, \end{aligned}$$

because, under (2.20), $\sum_{j=1}^n |b_{t,j+1} - b_{tj}| \le C$, while $\sum_{|t-j|>h} b_{tj} = o(1)$ by (5.13). Thus, $E|t_{n,2}| \le \sum_{k=0}^m \rho^k (E\theta_k^2)^{1/2} \le C\sum_{k=0}^\infty \rho^k k^{1/2} < \infty$.

To bound $t_{n,3}$, note that $|(1-\rho_{t_0})^{-1} - \sum_{k=0}^m \rho_{t_0}^k| \le \sum_{k=m+1}^\infty \rho_{t_0}^k \le \rho^{2m}(1-\rho)^{-2} = O(H^{-1})$, when $m = H$. Hence, $t_{n,3} = O_p(H^{-1}S'_{u,t}) = O_P(1)$, which completes the proof of (5.41).

Proof of (5.42). Recall that $t_0 = t - 2h$. Using notation $R_{t,h}$ of Lemma 5.1, bound $|\rho_t - \rho_{t_0}| \le R_{t,2h} = O_p\big((h/t)^\gamma\big) = o_p(1)$, by (5.17), because $t \sim \tau n$ and $h = O(\bar H) = o(n)$.

Proof of (5.40). Write $(\sigma_u B_t)^{-1} B_{1t}\bar u_t = \sum_{j=1}^n \theta_{nj}$, where $\theta_{nj} := (\sigma_u B_{2t})^{-1} b_{tj} u_j$ is a m.d.s. It suffices to verify (5.34) and (5.35). By the same argument as proving the second claim in (5.28), $\sum_{j=1}^n E[\theta_{nj}^2|\mathcal{F}_{j-1}] = (1 + o_p(1))\sum_{j=1}^n E\theta_{nj}^2 \to_p 1$, yielding (5.34). Finally, $\sum_{j=1}^n E\big[\theta_{nj}^2 I(|\theta_{nj}| \ge \varepsilon)\big] \le \varepsilon^{-2}\sum_{j=1}^n E\theta_{nj}^4 \le CB_{2t}^{-4}\sum_{j=1}^n b_{nj}^4 \le CB_{2t}^{-2} \to 0$ by (5.1) and $\max_j b_{nj} < \infty$, which verifies (5.35). □

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 $\widehat\sigma^2_{\hat Y\hat u,t}$, $\widehat\sigma^2_{Y'uu,t}$ and $\widehat\sigma^2_{\hat Y\hat u\hat u,t}$ as in Corollary 2.3 and (5.12).

Lemma 5.6. *Suppose assumptions of Theorem 2.4(i) hold and $\bar H = o(n)$. Then, with $\xi_n := (\bar H/n)^\gamma H + 1$,*

$$\textstyle\sum_{j=1}^n b_{tj}(\mu_j - \mu_t) = O_p\big((\bar H/n)^\beta H + 1\big), \tag{5.43}$$

$$S_{\hat Y \hat Y,t} - S_{Y'Y',t} = O_p(\xi_n), \quad \widehat\sigma^2_{\hat Y,t} - \widehat\sigma^2_{Y',t} = O_p(\xi_n). \tag{5.44}$$

In addition, if $\max_j E\mu_j^4 < \infty$, then

$$\widehat\sigma^2_{\hat Y \hat u,t} - \widehat\sigma^2_{Y'u,t} = o_p(H), \quad \widehat\sigma^2_{\hat Y \hat u \hat u,t} = \widehat\sigma^2_{Y'uu,t} + o_p(H) = O_p(H). \tag{5.45}$$

Proof. First we show (5.43). Let $h = b\hat H$ be as in (5.13). Recall that $t \sim \tau n$ and $h = o(n)$. Split the sum (5.43) in two parts, $\sum_{j=1}^n = \sum_{|j-t|>h} + \sum_{|j-t|\le h}$. 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 (5.13), so the first sum is $o_p(1)$.

To bound the second sum, consider two cases. First, let μ_t satisfies Assumption 2.3(i). Then, for $|j-t| \le h$, $E|\mu_j - \mu_t| \le (E(\mu_j - \mu_t)^2)^{1/2} \le C(|t-j|/t)^\gamma \le C(h/n)^\gamma$. Therefore, $E\big|\sum_{|j-t|\le h} b_{tj}(\mu_j - \mu_t)\big| \le C(\bar H/n)^\beta \sum_{|j-t|\le h} b_{tj} \le C(\bar H/n)^\beta H$ by (5.1), which yields (5.43).

Second, let μ_t satisfies Assumption 2.3(ii). Then $|\mu_j - \mu_t| \le |m(t,j)| + |m'(t,j)|$, and $E\sum_{|j-t|\le h} b_{tj}|m(t,j)| \le C(\bar H/n)^\beta H$ by the same argument as under Assumption 2.3(i). Next, by assumption $\max_{|j-t|\le h} |m'(t,j)| = O_p\big((h/n)^\beta\big) = O_p\big((\bar H/n)^\beta\big)$, and therefore, $\sum_{|j-t|\le h} b_{tj}|m'(t,j)| = O_p\big((\bar H/n)^\beta H\big)$. Clearly, the above bounds imply (5.43).

Proof of (5.44). We start with the first claim. Since $\sum_{j=1}^n b_{tj}(y_j - \bar y_t) = 0$, then

$$\begin{aligned} S_{\hat Y \hat Y,t} - S_{Y'Y',t} &= \textstyle\sum_{j=1}^n b_{tj}\{(y_j - \bar y_t)(y_{j-1} - \mu_t) - y'_j y'_{j-1}\} \\ &= \textstyle\sum_{j=1}^n b_{tj}\{(y_j - \mu_t)(y_{j-1} - \mu_t) - y'_j y'_{j-1}\} + (\mu_t - \bar y_t)\sum_{j=1}^n b_{tj}(y_{j-1} - \mu_t) =: q_{1t} + q_{2t}. \end{aligned}$$

So, it suffices show that $q_{it} = O_p(\xi_n)$, $i = 1, 2$.

Write $q_{1t} = \sum_{j=1}^n b_{nj}\theta_{tj}$, where $\theta_{tj} := (y_j - \mu_t)(y_{j-1} - \mu_t) - y'_j y'_{j-1}$. Then $|\theta_{tj}| \le |(y'_j + \mu_j - \mu_t)(y'_{j-1} + \mu_{j-1} - \mu_t) - y'_j y'_{j-1}| \le |(\mu_j - \mu_t)y'_{j-1}| + |(\mu_{j-1} - \mu_t)y'_j| + |(\mu_j - \mu_t)(\mu_{j-1} - \mu_t)|$. Notice that $E|\theta_{tj}| \le C(\max_j E\mu_j^2 + \max_j E y_j'^2) < \infty$, by Assumption 2.3 and (2.13).

Let h be as in the proof of (5.43). Then, for j such that $|t-j| \le h$, θ_{tj} has the following properties: Under Assumption 2.3 (i), $E|\theta_{tj}| \le C(E(\mu_j - \mu_t)^2)^{1/2} \le C(|t-j|/t)^\beta \le C(h/n)^\beta = O((\bar H/n)^\beta)$, while under Assumption 2.3 (ii), bound $|\theta_{tj}| \le \max_{j:|t-j|\le h} |\mu_j - \mu_t|\xi_j$, where $\xi_j := |y'_{j-1}| + |y'_j| + |\mu_{j-1} - \mu_t|$ satisfies $\max_j E\xi_j < \infty$. Consequently, $q_{1t} = O_p(\xi_n)$ follows by the argument as in the proof of (5.43).

To bound q_{2t}, observe that $\mu_t - \bar y_t = B_{1t}^{-1}\{\sum_{j=1}^n b_{tj}(\mu_j - \mu_t) + \sum_{j=1}^n b_{tj} y'_{j-1}\} = O_p\big((\bar H/n)^\beta + H^{-1/2}\big)$, by (5.43) and (5.39) and because $B_{1t} \sim H$. Then, $|\sum_{j=1}^n b_{tj}(y_{j-1} - \mu_t)| = B_{1t}|\mu_t - \bar y_t| = O_p\big(H\{(\bar H/n)^\gamma + H^{-1/2}\}\big)$, and $q_{2t} = O_p(\xi_n)$, completing the proof of the first claim.

Proof of the second claim in (5.44) is similar to that of the first one.

Proof of (5.45). To show the first claim, use $|\hat y_{j-1}^2 \hat u_j^2 - y_{j-1}'^2 u_j^2| \le |\hat y_{j-1}^2 - y_{j-1}'^2| u_j^2 + \hat y_{j-1}^2 |\hat u_j^2 - u_j^2|$, to bound

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

It suffices to show that $r_{n,l} = o_p(H)$, $l = 1, 2$.

To bound, $r_{n,1}$, for $L \geq 1$, let $u_j^* := |u_j| I(|u_j| \geq L)$ and use $|u_j| \leq L + u_j^*$, to obtain

$$|r_{n,1}| \leq \sum_{j=1}^n b_{nj} |\hat{y}_{j-1}^2 - y_{j-1}'^2| L + \sum_{j=1}^n b_{nj} |\hat{y}_{j-1}^2 - y_{j-1}'^2| u_j^{*\,2} = r_{n,11} L + r_{n,12}.$$

Arguing similarly as in the proof of (5.44), it follows that $r_{n,11} = o_p(H)$. To bound $r_{n,12}$, notice that $\max_j E\hat{y}_j^4 < \infty$, because $y_j = \mu_j + y_j'$, and $\max_j E\mu_j^4 < \infty$ by assumption of lemma, while $\max_j E y_j'^4 < \infty$ by (2.13). This and stationarity of u_j implies $E|\hat{y}_{j-1}^2 - y_{j-1}'^2| u_j^{*\,2} \leq C \max_j \{(E\hat{y}_{j-1}^4 + E y_{j-1}'^4)^{1/2} (E u_j^{*\,4})^{1/2}\} \leq C(E u_1^{*\,4})^{1/2} =: \varepsilon_L \to 0$ as $L \to \infty$. Hence $H^{-1} E|r_{n,12}| \leq \varepsilon_L H^{-1} \sum_{j=1}^n b_{nj} \leq C\varepsilon_L$ which implies $H^{-1} r_{n,1} \to_p 0$, as $H \to \infty$, $L \to \infty$.

It remains to bound $r_{n,2}$. Since $y_j = \rho_{j-1} y_{j-1} + u_j + \alpha_j$, then $\hat{u}_j - u_j = y_j - \hat{\rho}_{n,t} y_{j-1} - \widehat{\alpha}_t = (\rho_{j-1} - \hat{\rho}_{n,t}) y_{j-1} + \alpha_j - \widehat{\alpha}_t = (\rho_{j-1} - \hat{\rho}_{n,t}) y_{j-1} + \{\alpha_t - \widehat{\alpha}_t\} + \{\alpha_j - \alpha_t\}$. Hence,

$$\begin{aligned} |r_{n,2}| &\leq 3\Big(\sum_{j=1}^n b_{nj} \hat{y}_{j-1}^2 y_{j-1}^2 (\rho_{j-1} - \hat{\rho}_{n,t})^2 + (\alpha_t - \widehat{\alpha}_t)^2 \sum_{j=1}^n b_{nj} \hat{y}_{j-1}^2 \\ &\quad + \sum_{j=1}^n b_{nj} \hat{y}_{j-1}^2 (\alpha_j - \alpha_t)^2 \Big) =: 3(r_{n,21} + r_{n,22} + r_{n,23}). \end{aligned}$$

Then $r_{n,21} = o_p(H)$ follows using the same argument as bounding $q_{n,2}$ in (5.8). Next, $r_{n,22} = (\alpha_t - \widehat{\alpha}_t)^2 O_p(H) = o_p(H)$ by (2.28) and because $E \sum_{j=1}^n b_{nj} \hat{y}_{j-1}^2 = O(H)$.

To bound $r_{n,23}$, set $\theta_{n,j} := b_{nj}^{1/2} \hat{y}_{j-1}^2 |\alpha_j - \alpha_t|$ and denote $A := \cup_{j=1}^n I\big(\theta_{n,j} > 1\big)$. Then $P(A) \leq \sum_{j=1}^n E\theta_{n,j} = \sum_{j=1}^n b_{nj}^{1/2} \hat{y}_{j-1}^2 |\alpha_j - \alpha_t| =: q_n$. Notice that $|\alpha_j - \alpha_t| = |\mu_j - \mu_{j-1}\rho_{j-1} - (\mu_t - \mu_{t-1}\rho_{t-1})| \leq |\mu_j - \mu_t| + |\mu_{j-1} - \mu_{t-1}| \rho_{j-1} + |\mu_{t-1}(\rho_{j-1} - \rho_{t-1})|$. Then by the same argument as in the proof of (5.44) and (5.8), it follows that $q_n = O_p((\bar{H}/n)^\gamma H + 1) = o_p(H)$. Similarly, $\sum_{j=1}^n b_{nj} \hat{y}_{j-1}^2 (\alpha_j - \alpha_t)^2 I(A^c) \leq \sum_{j=1}^n b_{nj}^{1/2} |\alpha_j - \alpha_t| = o_p(H)$, which implies $r_{n,23} = o_p(H)$, completing the proof of the first claim in (5.45).

The second claim follows using the same argument as in the proof of the first one. □

References

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., and P. Guggenberger (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, 1636–52.

Benati, L., and P. Surico (2008): "Evolving U.S. Monetary Policy and The Decline of Inflation Predictability," *Journal of the European Economic Association*, 6, 634–646.

Benigno, G. (2002): "Real exchange rate persistence and monetary policy rules," *Journal of Monetary Economics*, 51, 473–502.

Benigno, P., L. A. Ricci, and P. Surico (2010): "Unemployment and productivity in the long run: the role of macroeconomic volatility," *NBER working paper no 16374*.

Brown, R. L., J. Durbin, and J. M. Evans (1974): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statical Association, Series A*, 138, 149–163.

Chan, N. H., and C. Z. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, 15, 1050–1063.

Chari, V., P. Kehoe, and E. McGrattan (2002): "Can sticky price models generate volatile and persistent real exchange rates?," *Review of Economic Studies*, 69, 533–563.

Chen, S., and C. Engel (2005): "Does 'aggregation bias' explain the PPP puzzle?," *Pacific Economic Review*, 10, 49–72.

Chortareas, G. E., and G. Kapetanios (2009): "Getting PPP right: Identifying mean-reverting real exchange rates in panels," *Journal of Banking and Finance*, 33, 390–404.

Chortareas, G. E., G. Kapetanios, and Y. Shin (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., M. Eichenbaum, and C. Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113, 1–45.

Cogley, T., T. Sargent, and G. Primiceri (2010): "Inflation gap persistence in the U.S.," *American Economic Association: Macroeconomics*, 2, 43–69.

Cogley, T., and T. J. Sargent (2001): "Evolving post World War 2 U.S. inflation dynamics," *NBER Macroeconomics Annual*, 16, 331–337.

——— (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302.

Dahlhaus, R. (1997): "Fitting time series models to nonstationary processes," *Annals of Statistics*, 25, 1–37.

Dahlhaus, R., and L. Giraitis (1998): "On the optimal segment length for parameter estimates for locally stationary time series," *Journal of Time Series Analysis*, 19, 629–656.

Dalla, V., L. Giraitis, and H. Koul (2012): "Automatic studentization in nonparametric regression," *Working paper, Queen Mary, University of London*.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.

Frankel, J. A., and A. K. Rose (1996): "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," *Journal of International Economics*, 40, 209–224.

Giraitis, L., H. Koul, and D. Surgailis (2012): *Large Sample Inference for Long memory Processes*. Imperial College Press, London.

Imbs, J., H. Mumtaz, M. Ravn, and H. Rey (2005): "PPP strikes back: aggregation and the real exchange rate," *Quarterly Journal of Economics*, 120, 1–43.

Kapetanios, G., and E. Tzavalis (2004): "Modeling structural breaks in economic relationships using large shocks," *Journal of Economic Dynamics and Control*, 34, 417–436.

Kapetanios, G., and T. Yates (2008): "An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change," *Mimeo, Bank of England.*

Koop, G., and S. Potter (2008): "Time varying VARs with inequality restrictions," *University of Strathclyde/FRB New York mimeo.*

Lai, T. L., and C. Z. Wei (2010): "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Annals of Statistics*, 10, 154–166.

Muller, U. K., and P. M. Petalas (2010): "Efficient Estimation of the Parameter Path in Unstable Time Series Models," *Review of Economic Studies*, 77, 1508–1539.

Muller, U. K., and M. W. Watson (2008): "Testing Models of Low-frequency Variability," *Econometrica*, 76, 979–1016.

Mumtaz, H., and P. Surico (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., and R. Prodan (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., and H. Theodoridis (1998): "Increasing Evidence of Purchasing Power Parity over the Current Float," *Journal of International Money and Finance*, 17, 41–50.

——— (2001): "The Choice of Numeraire Currency in Panel Tests of Pourchasing Power Parity," *Journal of Money, Credit and Banking*, 33, 790–803.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (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., and T. Magdalinos (2007): "Limit theory for Moderate deviations from a unit root," *Journal of Econometrics*, 136, 115–130.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for linear processes," *Annals of Statistics*, 20, 971–1001.

Ploberger, W., and W. Kramer (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 *Statistics, Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Springer Berlin.

——— (1991): "Time Varying Nonlinear Regression," in *Statistics, Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Springer Berlin.

Rogoff, K. (1996): "The purchasing power parity puzzle," *Journal of Economic Literature*, 34, 647–668.

Segalotto, J.-F., M. Arnone, and B. Laurens (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., and R. Wouters (2003): "An estimated dynamic stochastic general equilbrium model for the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

Stock, J. H., and M. W. Watson (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93(441), pp. 349–358.

White, H. (1980): *Asymptotic Theory for Econometricians.* Academic Press.

Table 5.1: MSE and 90% Coverage Probability results for $\widehat{\rho}_{n,t}$ for the normal kernel.

		MSE						Coverage Probability, $t=[n/2]$					
ϕ	$H\backslash$n	50	100	200	400	800	1000	50	100	200	400	800	1000
0	$n^{0.2}$	0.098	0.077	0.069	0.060	0.051	0.048	0.785	0.790	0.793	0.822	0.837	0.825
	$n^{0.4}$	0.048	0.035	0.026	0.020	0.016	0.014	0.839	0.836	0.836	0.839	0.836	0.852
	$n^{0.5}$	0.036	0.024	0.016	0.013	0.012	0.011	0.839	0.849	0.882	0.833	0.793	0.768
	$n^{0.6}$	0.026	0.018	0.014	0.012	0.012	0.013	0.863	0.831	0.791	0.736	0.672	0.609
	$n^{0.8}$	0.019	0.017	0.017	0.023	0.030	0.033	0.855	0.763	0.649	0.416	0.309	0.257
0.2	$n^{0.2}$	0.095	0.081	0.068	0.060	0.051	0.049	0.772	0.776	0.801	0.830	0.829	0.830
	$n^{0.4}$	0.050	0.034	0.026	0.020	0.016	0.015	0.827	0.855	0.858	0.818	0.853	0.840
	$n^{0.5}$	0.035	0.024	0.017	0.014	0.011	0.011	0.848	0.841	0.856	0.818	0.781	0.784
	$n^{0.6}$	0.025	0.019	0.014	0.012	0.013	0.012	0.864	0.815	0.781	0.728	0.646	0.639
	$n^{0.8}$	0.021	0.015	0.017	0.023	0.032	0.035	0.828	0.789	0.630	0.452	0.304	0.271
0.5	$n^{0.2}$	0.099	0.083	0.068	0.059	0.051	0.049	0.762	0.792	0.788	0.827	0.810	0.820
	$n^{0.4}$	0.053	0.034	0.026	0.020	0.015	0.014	0.792	0.838	0.860	0.852	0.861	0.849
	$n^{0.5}$	0.035	0.023	0.017	0.013	0.011	0.011	0.845	0.846	0.838	0.809	0.783	0.788
	$n^{0.6}$	0.025	0.018	0.014	0.012	0.012	0.012	0.874	0.816	0.812	0.729	0.641	0.628
	$n^{0.8}$	0.020	0.015	0.018	0.024	0.031	0.035	0.829	0.783	0.616	0.448	0.295	0.235
0.9	$n^{0.2}$	0.094	0.078	0.067	0.057	0.050	0.047	0.779	0.772	0.790	0.823	0.828	0.811
	$n^{0.4}$	0.048	0.032	0.024	0.018	0.013	0.013	0.839	0.853	0.872	0.844	0.858	0.846
	$n^{0.5}$	0.033	0.022	0.015	0.012	0.009	0.009	0.844	0.864	0.871	0.848	0.811	0.807
	$n^{0.6}$	0.024	0.017	0.012	0.010	0.011	0.011	0.869	0.843	0.846	0.767	0.669	0.647
	$n^{0.8}$	0.017	0.015	0.017	0.024	0.033	0.037	0.880	0.790	0.651	0.455	0.275	0.244

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

Table 5.2: MSE and 90% Coverage Probability results for $\widehat{\rho}_{n,t}$ for the normal kernel.

		MSE						Coverage Probability, $t=[n/2]$					
d	$H\backslash$n	50	100	200	400	800	1000	50	100	200	400	800	1000
	$n^{0.2}$	0.113	0.102	0.094	0.084	0.078	0.075	0.749	0.752	0.776	0.776	0.783	0.760
	$n^{0.4}$	0.071	0.063	0.055	0.051	0.045	0.044	0.752	0.744	0.722	0.701	0.665	0.655
0.51	$n^{0.5}$	0.058	0.054	0.048	0.044	0.041	0.040	0.752	0.722	0.678	0.628	0.568	0.544
	$n^{0.6}$	0.052	0.048	0.045	0.043	0.041	0.040	0.696	0.640	0.578	0.481	0.455	0.397
	$n^{0.8}$	0.046	0.044	0.044	0.046	0.047	0.047	0.679	0.541	0.445	0.333	0.253	0.220
	$n^{0.2}$	0.099	0.089	0.078	0.069	0.061	0.059	0.798	0.787	0.773	0.794	0.808	0.810
	$n^{0.4}$	0.056	0.044	0.035	0.030	0.026	0.025	0.792	0.825	0.805	0.809	0.785	0.785
0.75	$n^{0.5}$	0.041	0.034	0.029	0.026	0.023	0.022	0.819	0.807	0.758	0.708	0.680	0.680
	$n^{0.6}$	0.036	0.028	0.026	0.024	0.024	0.025	0.778	0.746	0.691	0.610	0.524	0.490
	$n^{0.8}$	0.028	0.026	0.029	0.034	0.040	0.041	0.786	0.644	0.498	0.364	0.280	0.213
	$n^{0.2}$	0.095	0.076	0.066	0.054	0.045	0.043	0.768	0.771	0.805	0.810	0.827	0.837
	$n^{0.4}$	0.046	0.032	0.024	0.017	0.012	0.011	0.827	0.843	0.853	0.840	0.879	0.895
1.25	$n^{0.5}$	0.034	0.021	0.014	0.009	0.007	0.006	0.834	0.851	0.862	0.874	0.865	0.863
	$n^{0.6}$	0.024	0.015	0.010	0.007	0.006	0.006	0.852	0.838	0.863	0.826	0.784	0.732
	$n^{0.8}$	0.019	0.013	0.012	0.016	0.020	0.026	0.826	0.816	0.736	0.566	0.405	0.411
	$n^{0.2}$	0.093	0.078	0.064	0.052	0.043	0.041	0.772	0.792	0.801	0.832	0.826	0.844
	$n^{0.4}$	0.045	0.032	0.022	0.015	0.010	0.009	0.840	0.850	0.865	0.874	0.852	0.891
1.49	$n^{0.5}$	0.032	0.020	0.013	0.008	0.005	0.005	0.856	0.855	0.876	0.868	0.877	0.866
	$n^{0.6}$	0.024	0.014	0.008	0.005	0.004	0.004	0.869	0.872	0.861	0.871	0.813	0.823
	$n^{0.8}$	0.017	0.010	0.009	0.012	0.017	0.020	0.870	0.854	0.790	0.663	0.557	0.513

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

Table 5.3: MSE and 90% Coverage Probability (CP) results for $\widehat{\rho}_{n,t}$, $\widehat{\alpha}_{n,t}$ and $\bar{y}_t$ for the normal kernel.

MSE $\hat{\rho}_t$									
	$u_t \sim$ i.i.d.			$u_t \sim$ GARCH			$u_t \sim$ stoch. vol.		
H\n	400	800	1000	400	800	1000	400	800	1000
$n^{0.2}$	0.073	0.063	0.060	0.080	0.072	0.070	0.085	0.076	0.071
$n^{0.4}$	0.023	0.018	0.017	0.036	0.030	0.028	0.063	0.051	0.050
$n^{0.5}$	0.020	0.018	0.019	0.033	0.029	0.028	0.058	0.050	0.049
$n^{0.6}$	0.026	0.027	0.028	0.040	0.039	0.039	0.060	0.059	0.056
$n^{0.8}$	0.062	0.077	0.083	0.066	0.084	0.092	0.080	0.092	0.099
MSE $\hat{\alpha}_t$									
$n^{0.2}$	0.710	0.674	0.561	0.778	0.636	0.674	0.564	0.501	0.506
$n^{0.4}$	0.151	0.113	0.110	0.162	0.144	0.130	0.227	0.193	0.198
$n^{0.5}$	0.153	0.133	0.137	0.167	0.172	0.156	0.229	0.184	0.204
$n^{0.6}$	0.192	0.192	0.195	0.210	0.199	0.210	0.250	0.248	0.244
$n^{0.8}$	0.368	0.358	0.376	0.376	0.393	0.355	0.375	0.355	0.405
MSE $\bar{y}_t$									
$n^{0.2}$	0.439	0.417	0.397	0.437	0.443	0.449	0.435	0.424	0.408
$n^{0.4}$	0.374	0.321	0.322	0.327	0.320	0.311	0.334	0.294	0.324
$n^{0.5}$	0.517	0.470	0.473	0.465	0.481	0.486	0.547	0.423	0.454
$n^{0.6}$	0.799	0.777	0.821	0.719	0.694	0.798	0.722	0.857	0.785
$n^{0.8}$	1.812	1.954	1.990	2.087	2.095	1.708	1.840	1.764	2.097
CP's $\hat{\rho}_t$, $t = [n/2]$									
	$u_t \sim$ i.i.d.			$u_t \sim$ GARCH			$u_t \sim$ stoch. vol.		
H\n	400	800	1000	400	800	1000	400	800	1000
$n^{0.2}$	0.785	0.780	0.788	0.775	0.769	0.780	0.771	0.758	0.753
$n^{0.4}$	0.795	0.805	0.799	0.769	0.810	0.791	0.661	0.697	0.659
$n^{0.5}$	0.742	0.680	0.671	0.711	0.701	0.694	0.585	0.620	0.599
$n^{0.6}$	0.523	0.472	0.453	0.585	0.558	0.528	0.514	0.493	0.487
$n^{0.8}$	0.213	0.156	0.139	0.314	0.247	0.229	0.350	0.288	0.274
CP's $\hat{\alpha}_t$, $t = [n/2]$									
$n^{0.2}$	0.792	0.805	0.797	0.790	0.762	0.775	0.776	0.774	0.778
$n^{0.4}$	0.806	0.804	0.822	0.775	0.802	0.788	0.651	0.693	0.700
$n^{0.5}$	0.710	0.673	0.648	0.680	0.670	0.643	0.541	0.587	0.586
$n^{0.6}$	0.456	0.433	0.409	0.518	0.465	0.431	0.460	0.417	0.416
$n^{0.8}$	0.182	0.145	0.144	0.188	0.161	0.136	0.198	0.178	0.191
CP's $\bar{y}_t$, $t = [n/2]$									
$n^{0.2}$	0.744	0.771	0.795	0.789	0.807	0.796	0.791	0.793	0.818
$n^{0.4}$	0.786	0.796	0.792	0.769	0.795	0.798	0.752	0.773	0.771
$n^{0.5}$	0.685	0.717	0.656	0.689	0.648	0.660	0.706	0.662	0.642
$n^{0.6}$	0.532	0.531	0.492	0.548	0.515	0.477	0.565	0.496	0.501
$n^{0.8}$	0.328	0.267	0.248	0.360	0.243	0.248	0.331	0.262	0.255

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

Figure 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.

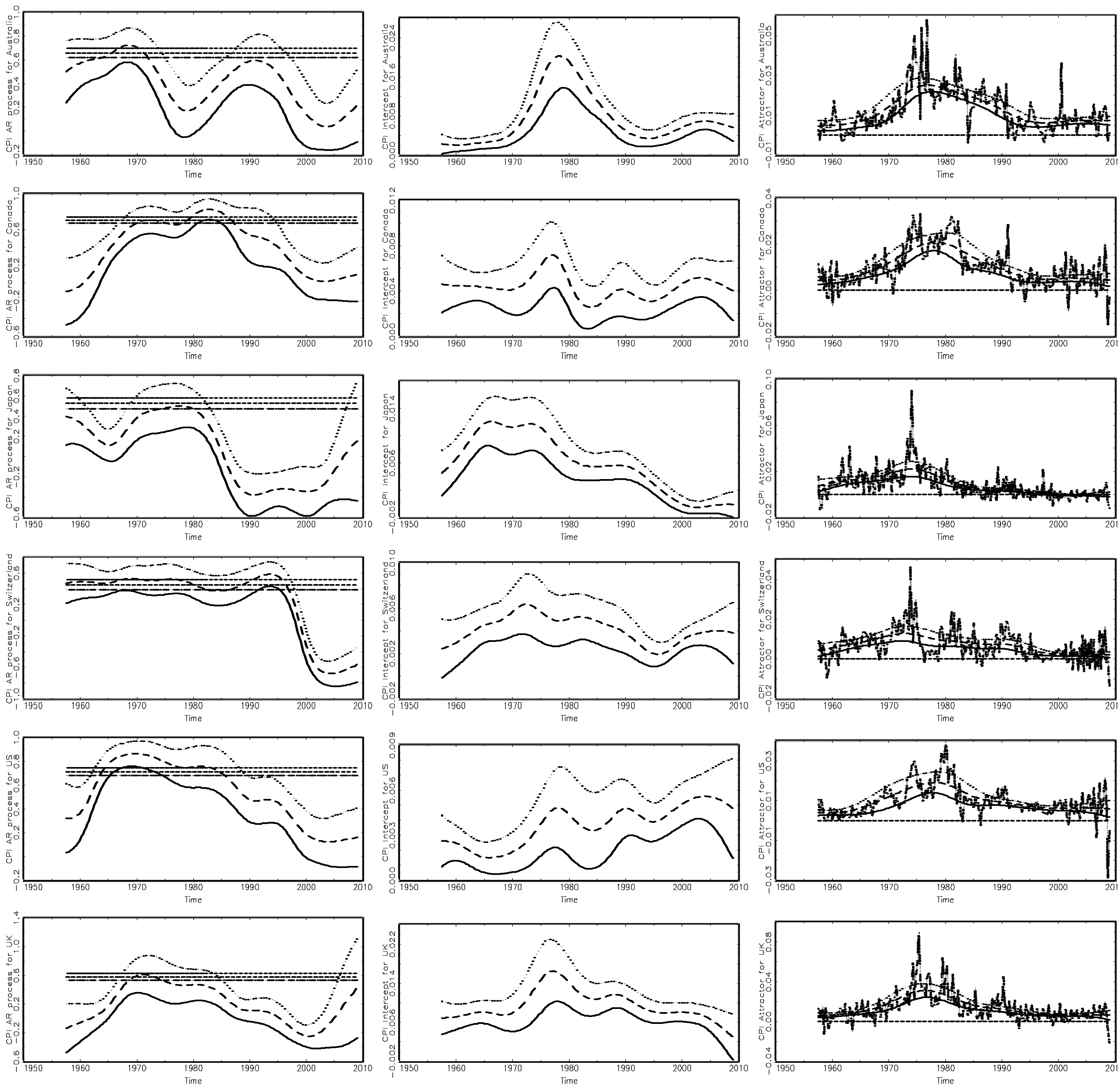


Figure 4: Time-Varying AR coefficient, intercept, attractor and 90% confidence bands in $AR(1)$ model, fitted to real exchange rates using a normal kernel for 6 countries: Australia, Canada, Japan, Switzerland, US and UK. The AR coefficient panels also report the estimate of an AR parameter in a fixed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