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

March 29, 2011

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 sample path of time series of 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 examines, in a rigorous manner, alternative kernel based estimation approaches for such models in a nonparametric framework 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 studentisation. 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 time series of the unobserved coefficient processes of random, or time varying, coefficient (RC) models, such as (3.1). As the next section on the related literature makes clear, 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 documenting changes in the degree of exchange rate pass-through. At this stage it is sufficient to cite a selection of papers that make use of such models. These are Cogley and Sargent (2001), Cogley and Sargent (2005), Cogley, Sargent, and Primiceri (2010), Benati (2010), Benati and Surico (2008), Mumtaz and Surico (2009), Pesaran, Pettenuzzo, and Timmermann (2006), Stock and Watson (1998), Koop and Potter (2008) and Koop and Potter (2007). It is clear that RC models provide a de facto benchmark technology for analysing structural change. The breadth of this previous work means that the results of this paper can 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 estimating RC models. This is especially the case for providing inference on the unobserved random coefficient processes of 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 such 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 those that are 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 devices can be used to bound the persistent processes serving as a model for time-varying coefficients. The question of how to bound the process is important. We use a simple approach to achieve that, which is illustrative and allows us to focus on the novelty of the kernel approach to estimating these models. Many other approaches can be used that result in different data generating mechanisms but which, we note, are estimable with the kernel approach. It is important to note further on this matter, that the question of how to restrict the coefficient process is not clearly addressed in the macroeconometric literature. Invariably, the restriction is imposed in a way that is computationally convenient without discussing the properties of the resulting model. As a result, it is unclear what is the best way to restrict the process from an economic point of view or what are the properties of the alternative models 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. We also consider the possibility that coefficient processes 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 coefficient process, our results are relevant much more widely. They apply to general regression models, multivariate VAR-type models and can be straightforwardly extended to models that allow for time-varying stochastic volatility such as those used widely in applied macroeconometrics.

The theoretical analysis in this paper is coupled with an extensive Monte Carlo study that addresses a number of issues arising out of our theoretical investigations. In particular, we find that our proposed estimator has the desirable properties identified in our theoretical analysis. For example, the theoretically optimal choice of bandwidth is also one of the best in small samples. Finally, we illustrate the usefulness of the kernel estimator in two applications that have received attention in previous work. The first uses our kernel estimator to document changes in inflation persistence over time. The second documents changes in the persistence of deviations from purchasing power parity (PPP) has fallen or not.

The rest of the paper is structured as follows: Section 2 discusses the existing literature and provides a framework for our contribution. Section 3 presents the model and some of its basic properties that are of use for later developments. Section 4 contains main theoretical results on the asymptotic properties of the new estimator. Section 5 provides an extensive Monte Carlo study while Section 6 discusses the application of the new inference methods to an empirical application on CPI inflation and real exchange rate data. Finally, Section 7 concludes. The proofs of all results are relegated to an Appendix.

2 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 such 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. Cogley and Sargent (2005) find evidence of change in the coefficients of the inflation process despite the inclusion of time-varying volatilities. Subsequent work by these authors in 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 an 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 find low variances for the shock to the equation governing the law of motion of the coefficients; 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 Dahlhaus (1997) and others who 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. It is important to note that 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, it is this kernel approach that we consider in the context of carrying out inference on RC models.

3 The model and its basic properties

3.1 The model

In this section we introduce a class of autoregressive models driven by a random drifting autoregressive parameter that evolves as a non-stationary process, standardised to take values in the interval $(-1, 1)$.

Such an autoregressive model is aimed to replicate patterns of evolution of autoregressive coefficients that are relevant for the modelling of the evolution of macroeconomic variables such as inflation. Such 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 forecasting and estimation.

The limit theory for stationary autoregressive models with non-random coefficients is well developed and understood. The asymptotic theory for AR models with time-invariant coefficients 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 a random coefficient, 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) model

$$y_t = \rho_{n,t-1} y_{t-1} + u_t, \quad t = 1, 2, \cdots, n, \tag{3.1}$$

with a drifting coefficient $\rho_{n,t}$ and initialization y_0, where $\{u_t\}$ is an i.i.d. sequence with zero mean and variance σ_u^2. Formally, $y_t = y_{tn}$ and $\rho_{n,t}$, $t = 0, \cdots, n$ are triangular arrays, where $\rho_{n,1}, \cdots, \rho_{n,n}$ represents a history between time moments 1 and n, which is the object of interest of estimation. For simplicity of notations we skip the index n for y_t.

The definition of $\rho_{n,t}$, is based on the following structural assumption. Given a (non-stationary) process $\{a_t\}$ and a parameter $\rho \in (-1, 1)$, the time-varying parameter $\rho_{n,t}$ is defined as a standartized version of $\{a_t\}$:

$$\rho_{n,t} = \rho \frac{a_t}{\max_{0 \le k \le n} |a_k|}, \quad t = 1, 2, \cdots, n, \tag{3.2}$$

where the stochastic process $\{a_t\}$ determines the random drift and ρ restricts $\rho_{n,t}$ away from the boundary points -1 and 1. Both $\{a_t\}$ and ρ are unknown. Observe that $\rho_{n,k} \in [-\rho, \rho] \subset (-1, 1)$, for all $k = 1, \cdots, n$.

To assure asymptotic stabilization of $\{y_t\}$ and enable statistical inference of the coefficient process $\rho_{n,t}$, we need additional assumptions on a_t and initialization y_0.

Assumption 3.1. *The random variables $(a_0, \cdots, a_n)$ are independent of the errors $(u_1, \cdots, u_n)$; $Ea_0^2 < \infty$ and $Ey_0^2 < \infty$.*

We assume that a_t evolves as

$$a_t = a_{t-1} + v_t, \quad t = 1, \cdots, n, \tag{3.3}$$

where $\{v_t\}$ is a stationary process with the zero mean. Denote by

$$S_n(\tau) := \sum_{j=1}^{[n\tau]} v_j, \quad 0 \le \tau \le 1$$

the partial sum process of v_j.

The popular empirical choice of v_t as i.i.d. sequence of random variables (with zero mean and variance σ_v^2) corresponds to a driftless random walk a_t (see, e.g. Cogley and Sargent (2005)). In i.i.d. case, the additional moment assumption $E|v_1|^{2+\delta} < \infty$ for some $\delta > 0$, assures weak convergence

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

in Shorokhod space $D[0,1]$ to a standard Brownian motion B_τ.

In our work, which covers the i.i.d. case, the variables v_t's are allowed to be dependent. The only assumption imposed on v_t, is a weak convergence of a renormalized partial sums process to a possibly non-Gaussian limit process. We denote it by W_τ to indicate that it may be different from the standard Brownian motion B_τ, and may be even non-Gaussian.

Assumption 3.2. *There exists $\gamma \in (0,1)$ such that*

$$n^{-\gamma} S_n(\tau) \Rightarrow_{D[0,1]} \sigma_v^2 W_\tau, \quad 0 \le \tau \le 1 \tag{3.4}$$

converges weakly in Shorokhod space $D[0,1]$ to some limit process $(W_\tau,\, 0 \le \tau \le 1)$ with zero mean and variance $\mathrm{Var}(W_\tau) = 1$ *and continuous paths in* $[0,1]$, *for some* $\sigma_v^2 > 0$.

Remark 3.1. Assumption 3.2 (weak convergence) is satisfied by a wide class of linear models

$$v_j = \sum_{k=0}^{\infty} a_k \zeta_{j-k}, \quad j \ge 0, \tag{3.5}$$

where $\{\zeta_k\}$ is a sequence of i.i.d. variables with zero mean and variance 1, $\sum_{k=0}^{\infty} a_k^2 < \infty$, such that

$$\mathrm{Var}(\sum_{j=1}^{n} v_j) \sim C n^{2\gamma}, \quad \gamma \in (0,1). \tag{3.6}$$

If a linear model (3.5) satisfies (3.6), then weak convergence of Assumption 3.2 holds true, if $\gamma > 1/2$, or $0 < \gamma \le 1/2$ and $E|\zeta_1|^p < \infty$ for some $p > 1/\gamma$, see, e.g., Giraitis, Koul, and Surgailis (2010), Proposition 4.3.6. Conditions (3.5)-(3.6) are satisfied by short and long memory and seasonal time series models. ARMA(p,q) models satisfy them with $\gamma = 1/2$. ARFIMA(p,d,q), $|d| < 1/2$ models, which are used to model short memory ($d = 0$), long memory ($0 < d < 1/2$) and negative memory ($-1/2 < d < 0$) times series, satisfy (3.5)-(3.6) with $\gamma = (1/2) + d$. The definition of a_j also allows stationary processes $\{v_j\}$, exhibiting seasonal long memory behaviour. Such processes can be generated by GARMA(p,d,q) models. Covariance functions of GARMA models resemble slowly decaying damped sine waves, whereas a spectral density has a singularity/zero point at frequency $\omega \ne 0$ and is continuous at zero frequency. GARMA models satisfy (3.5)-(3.6) with $\gamma = 1/2$, see section 7.2.2. of Giraitis, Koul, and Surgailis (2010).

Under Assumption 3.2, the coefficient process $\{\rho_{n,t},\, t = 1, \cdots, n\}$, as n increases, converges in distribution to the limit

$$\{\rho_{n,[n\tau]},\ 0 \le \tau \le 1\} \to_D \{\rho \widetilde{W}_\tau,\ 0 \le \tau \le 1\}, \tag{3.7}$$
$$\widetilde{W}_\tau := \frac{W_\tau}{\sup_{0 \le s \le 1} |W_s|},$$

where W_τ, $\tau \in [0,1]$ is the same as in (3.4). In particularly, W_τ can be a Brownian motion or fractional Brownian motion. (3.7) shows that the parameter $\rho_{n,[n\tau]}$ evolves around mean 0, and can take any value in the interval $[-|\rho|, |\rho|]$. The variance of the limit coefficient changes with t/or u.

Remark 3.2. To restrict $\rho_{n,t}$ in the interval $[-\rho, \rho]$, we use the normalization $\rho_{n,t} = \rho\, a_t / \max_{0\le k\le n} |a_k|$. Our methods and theory may be extended also to alternative standardizations such as $\rho_{n,t} = \rho\, a_t / \max_{0\le k\le t} |a_k|$, $t = 1, \cdots, n$. Another implicit standardisation that is popular in the applied macroeconometric literature is $\rho_{n,t} = a_t$,

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

Such alternative standardisations may allow the relaxation of the assumption of independence between the processes $\{a_t\}$ and $\{u_j\}$. In general the question of how to restrict $\rho_{n,t}$ can be tackled in a variety of ways none of which detracts from the main findings of the paper. It is important to note that the question of how to restrict $\rho_{n,t}$ is not clearly addressed in the macroeconometric literature. Usually the restriction is imposed in a way that is computationally convenient without discussing the properties of the resulting model. As a result it is unclear what is the best way to restrict the process from an economic point of view.

Remark 3.3. It is important to stress that the paper restricts the model of interest to be an AR(1), so as to set up an AR time-varying random framework and identify conditions that allow rigorous inference on it. Our main finding is 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 framework and inference for AR(1) models opens the possibility for general inference theory for AR and VAR models that may possess time-varying variances, and to more general error processes, such as martingale differences. 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.

$$y_t = \sum_{i=1}^{p} \rho_{n,t-1,i} y_{t-i} + u_t, \quad t = 1, 2, \cdots, n,$$

can be defined using the bounding condition

$$\rho_{n,t,i} = \rho \frac{a_{t,i}}{\max_{0\le k\le n} \sum_{i=1}^{p} |a_{k,i}|}, \quad t = 1, 2, \cdots, n,$$

where $0 < \rho < 1$, and each $a_{t,i}$ are independent versions of the a_t process used above. These bounding restrictions provide a sufficient condition for the maximum eigenvalue in absolute value of the matrix

$$A_{n,t} = \begin{pmatrix} \rho_{n,t,1} & \rho_{n,t,2} & \cdots & \rho_{n,t,p} \\ 1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 1 & 0 \end{pmatrix}$$

to be bounded above by one, for all t.
(2) Time Varying $VAR(1)$ model

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_{n,t-1}\boldsymbol{y}_{t-i} + \boldsymbol{u}_t, \quad t = 1, 2, \cdots, n,$$

where $\boldsymbol{y}_t$ is an $m-$dimensional vector, involves a bounding condition obtained by setting

$$\boldsymbol{\Psi}_{n,t-1} = \boldsymbol{Q}_t \boldsymbol{\Lambda}_{n,t-1} \boldsymbol{Q}_t'$$

where $\boldsymbol{Q}_t$ and $\boldsymbol{\Lambda}_{n,t-1}$ are defined as follows. Let

$$\tilde{\boldsymbol{\Psi}}_{t-1} = [\psi_{t-1,ij}], \qquad \psi_{t,ij} = \psi_{t,ij} + v_{\psi t,ij}, \quad t = 1, \cdots, n; \;\; i, j = 1, \cdots, m$$

where $v_{\psi t,ij}$ is a zero mean i.i.d. sequence with finite variance. Further, let

$$\tilde{\boldsymbol{\Psi}}_{t-1} = \boldsymbol{Q}_{t-1} \tilde{\boldsymbol{\Lambda}}_{t-1} \boldsymbol{Q}_{t-1}'$$

be the Schur decomposition of $\tilde{\boldsymbol{\Psi}}_{t-1}$. Then, $\boldsymbol{\Lambda}_{n,t-1}$ is obtained from $\tilde{\boldsymbol{\Lambda}}_{t-1}$ by replacing the i-th diagonal element of $\tilde{\boldsymbol{\Lambda}}_{t-1}$, denoted by $\tilde{\lambda}_{t-1,i}$, by $\lambda_{n,t-1,i}$ where

$$\lambda_{n,t,i} = \lambda \frac{\tilde{\lambda}_{t,i}}{\max_{1 \le i \le m,\ 0 \le k \le n} |\tilde{\lambda}_{k,i}|}$$

and $0 < \lambda < 1$. This ensures that the maximum eigenvalue of $\boldsymbol{\Psi}_{n,t-1}$ is bounded above by one in absolute value.

A third extension can be obtained by allowing time variation in the variance of the error term of the autoregressive model. Then, kernel estimators can be used to estimate time varying processes modelling that variance. The latter two extensions are currently the topic of further research by the authors. It is clear that there is great scope for adapting our framework to suit the needs of empirical researchers in applied macroeconometrics.

Remark 3.4. Our formal analysis assumes that the sequence u_t is i.i.d. This assumption can be straightforwardly relaxed to allow, for example, that u_t be a martingale difference sequence such that $\max_s \sum_t |Cov(u_t^2, u_s^2)| < \infty$ and $E(u_t^4)$ is uniformly bounded over t. We prefer to stick to the i.i.d. assumption for simplicity and clarity of exposition.

3.2 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

$$c_{t,0} := 1, \quad c_{t,j} := \prod_{k=1}^{j} \rho_{n,t-k}, \quad 1 \le j \le t \le n.$$

Note that

$$|c_{t,j}| \le |\rho|^j, \quad 1 \le j \le t \le n.$$

Next theorem describes basic properties of $y_t,\ t = 1, \cdots, n$.

Theorem 3.1. *Under Assumption 3.1, the random process $\{y_t,\ t=1,\cdots,n\}$ of (3.1) has the following properties.*
(i) y_t can be written as

$$y_t = \sum_{j=0}^{t-1} c_{t,j}u_{t-j} + c_{t,t}y_0 \tag{3.8}$$
$$= \sum_{j=0}^{k} c_{t,j}u_{t-j} + c_{t,k+1}y_{t-k-1}, \quad (1 \le k \le t-1).$$

(ii) The variance and covariance functions satisfy:

$$|\mathrm{Var}(y_t)| \le \frac{\sigma_u^2 + Ey_0^2}{1-\rho^2}, \quad Ey_t^2 \le \frac{\sigma_u^2 + Ey_0^2}{1-\rho^2}, \tag{3.9}$$

$$|\mathrm{Cov}(y_{t+k}, y_t)| \le |\rho|^k \mathrm{Var}(y_t), \quad t \ge 1,\ k \ge 0, \tag{3.10}$$
$$\le \frac{|\rho|^k}{1-\rho^2}(\sigma_u^2 + Ey_0^2).$$

The next theorem derives the asymptotic autocovariance $\mathrm{Cov}(y_{t+k}, y_t)$, as $t \to \infty$. In addition to Assumption 3.1, we need also Assumption 3.2.

Theorem 3.2. *Suppose, in addition, that in Theorem 3.1 Assumption 3.2 is satisfied.*
If $t = [n\tau]$, $\tau \in (0,1)$, then as $n \to \infty$,

$$\mathrm{Cov}(y_{t+k}, y_t) \to E\Big\{\frac{(\rho\widetilde{W}_\tau)^k}{1-(\rho\widetilde{W}_\tau)^2}\Big\}\sigma_u^2, \quad \forall k \ge 0, \tag{3.11}$$

$$y_t = \sum_{j=0}^{\infty}(\rho\widetilde{W}_\tau)^j u_{t-j} + o_P(1). \tag{3.12}$$

4 Estimation and Inference

In this section we construct a feasible estimation procedure of the drifting coefficient $\rho_{n,1},\cdots,\rho_{n,n}$ based on observables $y_1,\cdots,y_n$. We consider an estimate of $\rho_{n,t}$, that can be written as a weighted sample autocorrelation at lag 1. We shall show that under Assumptions 3.1 and 3.2, it is consistent and asymptotically normally distributed. Since computation of standard errors is straightforward, the method allows to construct the confidence band for the drifting coefficient under minimal restrictions on a_t, as long as $\{a_t\}$ is independent of the errors $\{u_t\}$. Finally, we propose and analyse an extension of the model to allow for a deterministic as well as a stochastic component in the unobserved coefficient process.

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

$$H \to \infty, \quad H = o(n). \tag{4.1}$$

To estimate $\hat{\rho}_{n,t}$, one can use the following 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 also consider a more general class of estimators

$$\hat{\rho}_{n,t} := \frac{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_k y_{k-1}}{\sum_{k=1}^{n} K(\frac{t-k}{H}) y_{k-1}^2}, \tag{4.2}$$

where $K(x) \geq 0$, $x \in \mathbb{R}$ is a continuous bounded function (kernel) such for some $\delta > 1$,

$$|K(x)| \leq C|x|^{-1-\delta}, \quad x \to \infty. \tag{4.3}$$

K does not require to be an even function. For example,

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

The flat and Epanechnikov kernels have a finite support, whereas Gaussian kernel has an infinite support.

In case when K has a finite support, asymptotic properties of $\hat{\rho}_{n,t}$ will be derived under Assumption 3.2, whereas if K has an infinite support, we shall need the following slightly stronger assumption.

Assumption 4.1. *Assumption 3.2 is satisfied with some* $\gamma \in (0, 1)$ *and*

$$\begin{aligned}
\mathrm{Var}(n^{-\gamma} S_n(1)) \leq C, \quad n \geq 1, \quad \forall n \geq 1, \\
E u_1^4 < \infty, \quad E y_0^4 < \infty.
\end{aligned} \tag{4.4}$$

Assumption (4.4) about the variance $\mathrm{Var}(n^{-\gamma} S_n(1))$ is closely related to (3.4) and in most of cases is easy verifiable under conditions that imply the weak convergence (3.4). We include it because formally (3.4) does not imply (4.4).

Now we discuss the asymptotic properties of the estimator $\hat{\rho}_{n,t}$ of (4.2).

Denote

$$b_{tk} := K(\frac{t-k}{H}), \quad 1 \leq t, k \leq n, \quad T_{H,t} := \frac{\sum_{k=1}^{n} b_{tk}}{\left(\sum_{k=1}^{n} b_{tk}^2\right)^{1/2}}, \tag{4.5}$$

$$\xi_{n,t} := \frac{\sum_{k=1}^{n} b_{tk} u_k y_{k-1}}{\sum_{k=1}^{n} b_{tk} y_{k-1}^2}, \quad t = 1, \cdots, n. \tag{4.6}$$

Theorem 4.1. *Let $y_1, \cdots y_n$ be defined as in (3.1), and $t = [n\tau]$, where $0 < \tau < 1$ is fixed. Assume that Assumptions 3.1 and 3.2 hold true with some $\gamma \in (0,1)$, and H and K satisfy (4.1) and (4.3), respectively. If K has an infinite support, assume, in addition, that Assumption 4.1 holds true.*
Then,

$$\begin{aligned}\hat{\rho}_{n,t} - \rho_{n,t} &= \xi_{n,t} + O_P((H/n)^\gamma) \qquad (4.7)\\ &= O_P(1/\sqrt{H}) + O_P((H/n)^\gamma),\end{aligned}$$

$$\frac{T_{H,t}}{(1-\rho_{n,t}^2)^{1/2}}\xi_{n,t} \to_D N(0,1). \qquad (4.8)$$

In addition, if $H = o(n^{\gamma/(0.5+\gamma)})$, then

$$\frac{T_{H,t}}{\sqrt{1-\hat{\rho}_{n,t}^2}}\Big(\hat{\rho}_{n,t} - \rho_{n,t}\Big) \to_D N(0,1). \qquad (4.9)$$

In particular, for $\gamma \geq 1/2$, (4.9) holds true, if $H = o(n^{1/2})$.

Since K is a continuous function, by (4.3) and the theorem of dominated convergence (TDC), for $t = [n\tau]$, $0 < \tau < 1$, as $n \to \infty$,

$$T_{H,t} \sim H^{1/2}\int_{\mathbb{R}} K(x)dx \Big/ \Big(\int_{\mathbb{R}} K^2(x)dx\Big)^{1/2}.$$

In particular, for a flat kernel, $T_{H,t} \sim \sqrt{H}$. The above estimator requires persistence of the process $\rho_{n,t}$, and non-stationarity (stochastic or deterministic trending behavior) of a_t, which is measured by the parameter $0 < \gamma < 1$, that defines the magnitude of the error term in the normal approximation:

$$\hat{\rho}_{n,t} - \rho_{n,t} = O_P(H^{-1/2} + (H/n)^\gamma), \qquad (4.10)$$

$$\frac{\sqrt{H}}{\sqrt{1-\hat{\rho}_{n,t}^2}}\Big(\hat{\rho}_n - \rho_n\Big) \sim N(0,1) + O_P(H^{1/2}(H/n)^\gamma). \qquad (4.11)$$

Larger values of γ correspond to a stronger persistence in a_t; a deterministic coefficient $\rho_{n,j}$ corresponds to $\gamma = 1$ in (4.11). Also, $\rho_{n,t} \equiv const$ corresponds to $\gamma = \infty$. Application of the normal approximation (4.11) does not require knowledge of γ. A process $\{v_j\}$ in $a_t = a_{t-1} + v_t$ can have short, long or negative memory. The main restriction on $\{v_j\}$ is to satisfy the functional central limit theorem with some normalization $n^{-\gamma}$, $0 < \gamma < 1$. In applications, it is practical to choose $H = o(n^{1/2})$. Such a bandwidth leads to a negligible error in (4.11) for short memory processes $\{v_j\}$ ($\gamma = 1/2$) and long memory processes ($1/2 < \gamma < 1$). When γ tends to 0, the pattern of trending behavior of a_t and the quality of approximation (4.11) deteriorate. In case of a stationary process a_t, the above estimation is not consistent.

In order to give an idea of the nature of the confidence bands implied by Theorem 4.1, we present in Figure 2 an $\rho_{n,t}$ realisation based on a random walk model for a sample size of 500, its estimate based on flat kernel and a bandwidth of $\sqrt{n}$ together with 90% confidence bands. As we can see the process is well tracked and the confidence band contains the true process most of the time (85.4% of the time to be exact).

Next, we consider the case when the process a_t, defining the AR(1) coefficient, $\rho_{n,t}$, includes a deterministic drift:

$$a_t = a_{t-1} + \mu(t/n) + v_t, \quad t = 1, \cdots, n, \tag{4.12}$$

where $\mu(x)$ is a continuous function on $[0, 1]$, such that $\sup_{0 \le x \le 1} |\mu(x)| > 0$, and v_t's are the same as in (3.3). If $\mu(x) \ne 0$, then the non-stationary process a_t is a trending unit root process:

$$\begin{aligned} a_t &= \sum_{j=1}^{t} \mu(j/n) + \sum_{j=1}^{t} v_j + a_0 \\ &= \text{"Deterministic trend" + "stochastic trend"}. \end{aligned} \tag{4.13}$$

The following theorem shows that results of Theorem 4.1 extend to the model (4.12). Comparing to (3.3), in (4.13) the deterministic trend dominates the stochastic trend which improves quality of estimation. It also indicates that asymptotic results obtained for an AR(1) model with a random coefficient remain valid for an AR(1) model with a time varying deterministic coefficient:

$$a_t = \varphi(t/n), \quad t = 1, \cdots, n, \tag{4.14}$$

where $\varphi(x)$, $x \in [0, 1]$ is a continuous function with a bounded derivative, such that $\sup_{0 \le x \le 1} |\varphi(x)| > 0$.

Theorem 4.2. *Let $y_1, \cdots y_n$ and a_t be defined as in (3.1), and (4.1) and (4.3) be valid.*
(i) Suppose that a_t is as in (4.12) and satisfies Assumptions 3.1-3.2. If K has an infinite support, assume in addition, that Assumption 4.1 is satisfied.

Then

$$\begin{aligned} \hat{\rho}_{n,t} - \rho_{n,t} &= \xi_{n,t} + O_P((H/n)) \\ &= O_P(1/\sqrt{H}) + O_P((H/n)), \end{aligned} \tag{4.15}$$

and (4.8) holds true.

If $H = o(n^{2/3})$, then $\hat{\rho}_{n,t}$ satisfies (4.9).
(ii) If a_t is defined as in (4.14), then (4.15) remains true.

5 Monte Carlo study

In this section, we report results of a Monte Carlo study on the small sample properties of the new kernel based estimator of a coefficient process. We consider the following model which accords with that analysed in the preceding section.

$$y_{n,t} = \rho_{n,t} y_{n,t-1} + u_t, \qquad 1 \leq t \leq n,$$
$$\rho_{n,t} = \rho \frac{a_t}{max_{0 \leq i \leq n} |a_t|}, \qquad |\rho| \leq 1. \tag{5.1}$$

and

$$a_t = a_{t-1} + v_t.$$

(5.1) uses the same specification to bound $\rho_{n,t}$ between $-\rho$ and ρ as that applied in the previous section. As we have noted earlier, this is only one of a multitude of ways in which boundedness can be imposed on $\rho_{n,t}$ and should not be viewed as either restrictive or unique. For example, we note that (5.1) can be replaced with

$$\rho_{n,t} = \rho \frac{a_t}{max_{0 \leq i \leq t} |a_t|}, \qquad |\rho| \leq 1, \tag{5.2}$$

without affecting in any significant way the conclusions reached in our Monte Carlo study. Detailed results supporting this statement are available upon request.

While the baseline case is one where both $\{v_t\}$ and $\{u_t\}$ are martingale difference processes, we will consider cases where this assumption is relaxed. The martingale difference assumption is much more crucial for $\{u_t\}$, whereas theoretical result allow for a wide class of dependent v_t's. Therefore, to investigate robustness of the estimation to dependence in u_t's, we only consider one form of deviation from i.i.d-ness by assuming that

$$u_t = \theta u_{t-1} + \epsilon_{1t},$$

where $\theta = 0,\ 0.2,\ 0.5$. For v_t we assume that it is either a short memory process given by

$$v_t = \phi v_{t-1} + \epsilon_{2t}$$

or a long memory process given by

$$(1 - L)^{d-1} v_t = \epsilon_{2t}.$$

We let $\phi = 0, 0.2, 0.5, 0.9$, $\theta = 0, 0.2, 0.5$ and $d = 0.51, 0.75, 1.25, 1.49$. Both ϵ_{1t} and ϵ_{2t} are assumed to be standard normal variates. We set $\rho = 0.9,\ 1$. Note that although $\rho = 1$ is not covered by the theory, which assumes that $|\rho| < 1$, it is of interest to see how the estimator works in that case. We consider two-sided kernel estimators with two different kernels: a normal kernel and a flat kernel. The bandwidth H for both kernels are set to n^{α} where $\alpha = 0.2, 0.4, 0.5, 0.6, 0.8$. Note that the value $\alpha = 0.5$ corresponds to the optimal value for the bandwidth derived in the previous section. Finally we set

$n = 50, 100, 200, 400, 800, 1000$. Results are reported in Tables 8.1-8.3 for the various experiments discussed above. The performance measure chosen is an MSE type measure given by $MSE_n = \frac{1}{n}\sum_{t=1}^{n}(\hat{\rho}_{n,t} - \rho_{n,t})^2$. Averages of MSE_n over 1000 replications are reported.

Table 8.1 reports results for $\rho = 0.9$ and allowing short memory for v_t's. In results not fully reported in the paper, due to space considerations but available on request, we consider the case where $\rho = 1$ which is not covered by theory but is in line with the models used in the empirical literature. In general, the same patterns emerge as in Table 8.1 but the estimators perform slightly worse with higher average MSEs. We comment next on some clear patterns that emerge from Table 1, including the case $\theta = 0$ covered by the theory, and cases $\theta = 0.2, 0.5$ analyzing the impact of dependence in u_t's (cases $\theta = 0.2, 0.5$ are not reported due to space considerations but are available on request). We focus on the normal kernel estimator as very similar patterns occur for the flat kernel. In the case $\theta = 0$, the consistency of $\hat{\rho}_{n,t}$ is clear as the average MSE falls substantially with sample size, from, say, 0.036 for $\phi = 0$, and an optimal bandwidth, for $n = 50$, to 0.011 for $n = 1000$. This fall is observed for all choices of bandwidths.

The choice of bandwidth has a substantial effect on the performance of the estimator. Rather neatly, it is clear that the theoretically optimal choice, $H = n^{0.5}$, of the bandwidth is also very good in finite samples. For the case $\rho = 0.9$, this is clear for the larger sample sizes (≥ 800), and for all sample sizes for $\rho = 1$. So for, say, $\phi = 0$, $\rho = 0.9$ and $n = 1000$ the optimal bandwidth has an MSE equal to 0.011, compared to 0.012 for the second best bandwidth choice and 0.049 for the worst such choice. This superiority of the theoretically optimal bandwidth, is further accentuated for larger samples.

The presence of short memory in v_t does not seem to affect the estimator adversely. If anything the performance of the estimator improves as v_t becomes more persistent which corresponds to a stronger persistence for $\rho_{n,t}$. For example, when $\phi = 0$, $n = 50$ and $\alpha = 0.5$, the MSE is 0.036 while for $\phi = 0.9$, $n = 50$ and $\alpha = 0.5$, the MSE is 0.032. This is, in fact, reasonable if one notes that it is the high persistence of $\rho_{n,t}$ that allows kernel estimators to be consistent in this setting. On the contrary dependence in u_t is problematic as expected. Low levels of persistence can be tolerated as is the case when the AR coefficient θ of u_t is 0.2. When this coefficient θ rises to 0.5 problems of inconsistency are much more evident. As a result, and to save space, we only consider nonzero values for θ in Table 8.1. The estimator based on the flat kernel performs only slightly worse but otherwise the patterns are similar to those observed for the normal kernel.

Table 8.2 reports results for the case where v_t is a strongly persistent process. There, we clearly see once again the familiar pattern whereby more persistent processes for $\rho_{n,t}$ allow for better estimation when kernel estimators are used. So when a_t is $I(0.51)$ or $I(0.75)$, the performance of the estimator is somewhat worse compared to the case where a_t is $I(1)$ which is itself somewhat worse than the performance of the estimator when a_t is $I(1.25)$ or $I(1.49)$. For example, when $d = 0.51$, $n = 50$ and $\alpha = 0.5$, the MSE is 0.13 while for $d = 1.49$, $n = 50$ and $\alpha = 0.5$, the MSE is 0.068. This accords with the theory for $|\rho| < 1$ developed in the previous section. Otherwise, the same patterns emerge as in Table 8.1.

Finally, to compare the performance of the estimator in case of a random and non-random coefficient we consider the case where in fact there is no time variation in $\rho_{n,t}$, and $\rho_{n,t} = 0.9$. Results for this case are presented in Table 8.3. The estimators in this case also work very well and are consistent as standard theory would immediately suggest. Clearly, here the best bandwidth is the highest one. Otherwise, similar patterns to those apparent in Tables 8.1-8.2, also emerge.

6 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.

6.1 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: 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 a 'constant' term which is allowed to vary over time as well. We have considered fitting the simpler model where the constant term is time-invariant but found that, in the majority of cases, allowing for time variation in this coefficient makes material difference in the results suggesting that indeed the 'constant' term needs to be allowed to vary over time. 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 2-3. Figure 2 relates to CPI inflation and Figure 3 to real exchange rates. They report the estimated time-varying AR coefficient and the standard time-invariant $AR(1)$ coefficient together with their standard errors.

6.2 Empirical Results

The empirical results presented in Figures 2-3 can help provide answers to two important empirical topics: Inflation persistence and the validity of the PPP hypothesis. We will examine each issue in turn.

6.2.1 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. Benati (2010) adopts exactly this tactic, and infers 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 2 record our results. Overall, it is quite clear that persistence has varied considerably and, once confidence bands are taken into account, statistically significantly, over time. It is also clear that assuming a fixed autoregressive coefficient is problematic since for most cases the time-varying coefficients and fixed coefficients are significantly different for most of the sample. Further, we note that both the autoregressive coefficient and the 'constant' term in the autoregression vary over time since, allowing for time variation in the 'constant' term, changes the profile of the time-varying autoregressive coefficient.

Our findings confirm the claims of previous work that inflation was both high and persistent in the 70's and the beginning of the 80's. From then on, we see that persistence fell in most countries, through to the early 2000s. Although persistence seems to have risen through the last decade, it remains a weakly autocorrelated process compared to the degree of autocorrelation observed prior to that. (An exception here is the US, where inflation persistence has not risen latterly).

We estimate using alternative kernels. The flat kernel produces estimated processes that look more 'stochastic' than those produced by the normal kernel. This is a mechanical consequence of the extra weight that the flat kernel gives to observations entering and leaving the sample for the coefficient estimated for a particular time period. However, the overall estimated pattern for inflation persistence is the same, suggesting a clear degree of robustness to the choice of the kernel used.

Our findings of significant time-variation in inflation persistence shed light on the economic structure underlying inflation. In particular, if it is not plausible to argue that indexation or other frictions that give rise to persistence vary over time, then such rigidities are less likely to be the cause of inflation persistence. The likelihood is that such persistence has changed for other reasons, for example, because of some change in monetary policy. Such an inference would correspond with the spread of monetary regimes involving independent central banks, inflation or similarly transparent targets; and also with the increasing 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.

6.2.2 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 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 difference between monetary policy rules in operation in two countries. 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.

The estimator proposed in this paper can uncover the potential evolution in the persistence of deviations from PPP. Turning to the results, there is clear evidence of time-variation in both the 'constant' term and the autoregressive coefficient, when a time-varying autoregressive model is fitted to the data. The time-varying autoregressive coefficients and the fixed autoregressive coefficients are clearly quite different most of the time. Real exchange data are considerably more persistent than CPI inflation data. It is very interesting to note that, across both countries and kernels, the values estimated by the time-varying autoregressive coefficient model are, in general, lower, and mostly significantly so, than those estimated by the fixed autoregressive coefficient model. This is likely to be a result, in part, of the fact that, as Perron (1990), noted, failing to account for time-variation in the mean leads one incorrectly to estimate higher persistence. There is a tendency for the time-varying coefficient to exhibit a cyclical behaviour with a number of clearly identified cycles. Nevertheless, there is also an overall tendency for the coefficient to drop over time. This tendency is more clearly visible when the normal kernel is used although it is also apparent for the flat kernel (results are not reported but available upon request).

Overall, our results support the conclusion that the persistence of deviations from PPP has fallen. The average autoregressive coefficient peaks at above 0.95 in the early seventies, and falls to below 0.9 by the end of the sample. In economic terms this is a large fall. To put this number in perspective: note that while, say, an autoregressive coefficient of 0.98 implies a half life of about 35 quarters for a shock to the real exchange rate, one of 0.88 implies a half life of about 5 quarters.

A number of inferences can be drawn from these findings. One inference is that impediments to goods and factor price arbitrage that acts as an attractor for prices in different countries have lessened over time. Such impediments might include nominal and real rigidities or barriers to trade. The reduction of the impact of these impediments is consistent with the increase in the ratio of world trade to GDP over our sample period. This could have allowed a greater increase in competition, which would have increased the costs of firms trading across national boundaries keeping nominal prices fixed. The above clearly suggests that barriers to trade have been reduced. A second inference is that monetary policies have evolved in a way that reduced PPP deviation persistence. Finally, a third implication relates to a possible decline in the prevalence or size of shocks inducing Balassa-Samuelson effects on the real exchange rate.

7 Concluding Remarks

This paper has proposed a new and admittedly novel approach to the estimation of time-varying coefficient models. It has advocated the use of kernel inference for estimating unobserved stochastic coefficient processes in stochastic time-varying coefficient models. To our knowledge, it is the first time that kernel estimation has been proposed for inference a stochastic entity related to macroeconomic variables. The proposed estimation approach has desirable properties such as consistency and studentised 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 theoretical results provide the justification for according to kernel estimation an important role in inference of time variation. Our findings coupled with the well known applicability of kernel estimation for locally stationary processes, suggests 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, if stochastic, 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. As we noted earlier in the paper, these models have been shown to display pathologies that our approach avoids, as documented in Stock and Watson (1998) and Koop and Potter (2008).

The theoretical properties of the kernel estimator are obviously only relevant from a classical perspective. Many of the implementations of RC models adopt a Bayesian approach. To the extent that those papers are truly Bayesian the theoretical advantages of our estimator are not relevant. That said, many of the Bayesian implementations of the RC approach have only the thinnest of Bayesian veneers, using determinedly uninformative priors wherever possible.

One further extremely attractive aspect of the new estimator - which ought to appeal to Bayesians and frequentists alike - relates to its relative computational tractability. Estimation of RC models using standard methods, including Bayesian estimation, is extremely computationally demanding. Relatively small multivariate models can potentially take numerous hours to estimate, even on powerful PCs. Further, the use of these estimators requires considerable programming experience. The need for significant computing power and programming expertise, has in fact inhibited the investigation of the small sample properties of such estimators via Monte Carlo studies. On the contrary, the computational demands, associated with the use of the new estimator, are extremely 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, and by the considerable computational advantage conferred by the kernel estimator.

Before concluding, it is of interest to suggest topics of future research in this area. Firstly, it is important to generalise the theoretical framework to a general regression model. While this appears reasonably straightforward from a theoretical perspective, it is of interest to note of the possibility to relax the assumption that the errors of the model are independent or more generally martingale difference, which is desirable for autoregressive models, when regressors are exogenous. Secondly, allowing for time-variation in the variance of the error term is important and of clear relevance to applied macroeconometricians who work on issues like the relative importance of time-varying shock variances versus coefficient time variation for policy analysis. Given our work and the work of Kapetanios (2007) it is clear that consistent kernel estimation of stochastically time-varying shock variance processes is feasible. Thirdly, it is of interest to develop estimation of unobserved stochastic time varying processes standing for parameters of nonlinear models. This will allow for the introduction of time variation and its estimation in more complex models such as DSGE models in macroeconomics. Fourthly, our work allows for a wide class of unobserved processes to be estimated via the kernel approach in a semiparametric set-up. It is then of interest to investigate the possibility of using the estimated process to determine its parametric structure. While consistent estimation of parameters of an underlying model should be possible, the issue of how to carry out inference on such estimated parameters remains an open question.

8 Appendix. Proof of Theorems 3.1-4.2.

Proof of Theorem 3.1. (i) Equations of (3.8) follow using recursions

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

(ii) To prove (3.9), use $\mathrm{Var}(X) \le EX^2$, $|c_{t,j}| \le |\rho|^j$ and (3.8), to obtain

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

because by Assumption 3.1, variables u_s, $s = 1, \cdots, t$ are independent of $y_0, \cdots, y_t$. The bound (3.9) for Ey_t^2 follows using the same argument as above.

To prove (3.10), use (3.8) and the fact that the random variables u_s, $s > t$ are independent of y_t and $c_{t+k,j}$ for any $k, j, t \ge 0$, to obtain

$$\begin{aligned}
\mathrm{Cov}(y_{t+k}, y_t) &= \mathrm{Cov}\Big(\sum_{j=0}^{k-1} c_{t+k,j}u_{t+k-j} + c_{t+k,k}y_t,\, y_t\Big) \\
&= \mathrm{Cov}\Big(c_{t+k,k}y_t,\, y_t\Big) \le \mathrm{Var}^{1/2}(c_{t+k,k}y_t)\mathrm{Var}^{1/2}(y_t) \qquad (8.1) \\
&\le |\rho|^k \mathrm{Var}(y_t),
\end{aligned}$$

which together with (3.9) implies (3.10). □

Proof of Theorem 3.2. To show (3.11), use (8.1) and (3.8) to obtain

$$\begin{aligned}
\mathrm{Cov}(y_{t+k}, y_t) &= \mathrm{Cov}\Big(c_{t+k,k}y_t,\, y_t\Big) \\
&= E\big[c_{t+k,k}\sum_{j=0}^{t-1} c_{t,j}^2\big]\sigma_u^2 + \mathrm{Cov}\Big(c_{t+k,k}c_{t,t}y_0,\, c_{t,t}y_0\Big).
\end{aligned}$$

By $|\mathrm{Cov}(X,Y)| \le (EX^2EY^2)^{1/2}$ it follows that

$$\Big|\mathrm{Cov}\Big(c_{t+k,k}c_{t,t}y_0,\, b_{t,t}y_0\Big)\Big| \le |\rho|^{2t+k}Ey_0^2 \to 0,$$

because $|\rho| < 1$, $Ey_0^2 < \infty$ and $t = [n\tau] \to \infty$. Hence, to show (3.11), it suffices to prove that

$$E\left[c_{t+k,k}\sum_{j=0}^{t-1} c_{t,j}^2\right] \to E\Big\{\frac{(\rho\widetilde{W}_\tau)^k}{1-(\rho\widetilde{W}_\tau)^2}\Big\}. \tag{8.2}$$

We split the proof into two steps:

$$\sup_{n\geq 1} E[c_{t+k,k}\sum_{j=M}^{t-1} c_{t,j}^2] \to 0, \quad M \to \infty, \tag{8.3}$$

$$E[c_{t+k,k}\sum_{j=0}^{M} c_{t,j}^2] \to E[(\rho\widetilde{W}_\tau)^k \sum_{j=0}^{M}(\rho\widetilde{W}_\tau)^{2j}], \quad n \to \infty, \quad \forall M \geq 1, \tag{8.4}$$

$$\to E\Big\{\frac{(\rho\widetilde{W}_\tau)^k}{1-(\rho\widetilde{W}_\tau)^2}\Big\}, \quad M \to \infty. \tag{8.5}$$

Relations (8.4) and (8.5) imply (8.2)

To prove (8.3), use $|c_{t,j}| \leq |\rho|^j$, which yields

$$|E[c_{t+k,k}\sum_{j=M}^{t-1} c_{t,j}^2]| \leq \sum_{j=M}^{\infty} |\rho|^{2j+k} \to 0, \quad M \to \infty.$$

To prove (8.4), recall that $c_{t,k} = \rho_{n,t-1}\cdots\rho_{n,t-k}$, where $|\rho_{n,t-j}| \leq |\rho| < 1$, $j \geq 1$. Observe that the sum $c_{t+k,k}\sum_{j=0}^{M} c_{t,j}^2$ is a linear combination of products of bounded variables, $\rho_{n,t-M}, \cdots, \rho_{n,t+k}$, and by Assumption 3.2, for any $M > 1$,

$$(\rho_{n,t-M}, \cdots, \rho_{n,t+k}) \to_D (\rho\widetilde{W}_u, \cdots, \rho\widetilde{W}_u), \tag{8.6}$$

which by standard argument yields (8.4).

Finally, (8.5) follows from (8.4), noting that $|\widetilde{W}_\tau| \leq 1$, and therefore

$$E[(\rho\widetilde{W}_\tau)^k \sum_{j=M+1}^{\infty}(\rho\widetilde{W}_\tau)^{2j}] \leq |\rho|^k \sum_{j=M+1}^{\infty}\rho^{2j} \to 0, \quad M \to \infty. \tag{8.7}$$

To show (3.12), write (3.8) as

$$y_t = \sum_{j=0}^{M} c_{t,j}u_{t-j} + R_{t,M}, \quad R_{t,M} := \sum_{j=M+1}^{t-1} c_{t,j}u_{t-j} + c_{t,t}y_0.$$

By (8.6),

$$(c_{t,1}, \cdots, c_{t,M}) \to_D ((\rho\widetilde{W}_\tau)^1, \cdots, (\rho\widetilde{W}_\tau)^M), \quad n \to \infty.$$

Bearing in mind, that by Assumption 3.1, $\{c_{t,j}\}$ and $\{u_j\}$ are independent random variables, this and (8.7) yield

$$\sum_{j=0}^{M} c_{t,j}u_{t-j} \to_D \sum_{j=0}^{M}(\rho\widetilde{W}_\tau)^j u_{t-j}, \quad \forall M \geq 1,$$

$$\to_D \sum_{j=0}^{\infty} (\widetilde{W}_\tau)^j u_{t-j}, \quad M \to \infty.$$

On the other hand,

$$E|R_{t,M}| \le \sum_{j=M+1}^{t-1} |\rho|^j E|u_{t-j}| + |\rho^t| E|y_0| \to 0, \quad M \to \infty, \quad t \to \infty,$$

which implies $R_{t,M} = o_P(1), \quad M \to \infty$ and completes proof of (3.12) and the theorem. □

In the sequel, we shall use the following notations:

$$B_{H,t}^2 := \sum_{k=1}^{n} b_{tk}^2, \quad \beta_{H,t} := \sum_{k=1}^{n} b_{tk}, \tag{8.8}$$
$$B_{H,t}^2 \sim H \int_{\mathbb{R}} K^2(x) dx, \quad \beta_{H,t} \sim H \int_{\mathbb{R}} K(x) dx,$$

where approximations follows from continuity of K, (4.3), applying theorem of dominated convergence.

Proof of Theorem 4.1. (i) Write

$$\begin{aligned}
\sum_{k=1}^{n} b_{tk} y_k y_{k-1} &= \sum_{k=1}^{n} b_{tk} \rho_{n,k-1} y_{k-1}^2 + \sum_{k=1}^{n} b_{tk} u_t y_{t-1} \\
&= \rho_{n,t} \sum_{k=1}^{n} b_{tk} y_{k-1}^2 + \sum_{k=1}^{n} b_{tk} u_k y_{k-1} + \sum_{k=1}^{n} (\rho_{n,k-1} - \rho_{n,t}) b_{tk} y_{k-1}^2 \\
&=: \rho_{n,t} V_n + S_{n,1} + S_{n,2}.
\end{aligned}$$

Setting $r_{nt} := S_{n,2}/V_n$, write

$$\hat{\rho}_{n,t} = \rho_{n,t} + \xi_{n,t} + r_{nt}.$$

To obtain (4.7), it suffices to show

$$r_{nt} = O_P((H/n)^\gamma), \tag{8.9}$$
$$\xi_{n,t} = O_P(1/\sqrt{H}). \tag{8.10}$$

To prove (8.9), observe that by Lemma 8.1 below, $S_{n,2} = O_P\big((\frac{H}{n})^\gamma H\big)$, whereas by (8.24) of Lemma 8.2, $|\rho_{n,t}| \le \rho < 1$, and (8.8),

$$V_n^{-1} = O_P(B_{H,t}^{-2}) = O_P(H^{-1}), \tag{8.11}$$

which yields

$$r_{nt} = \frac{S_{n,2}}{V_n} = O_P((\frac{H}{n})^\gamma).$$

To prove (8.10), recall that $\xi_{n,t} = S_{n,1}/V_n$. We show that

$$ES^2_{n,1} \le CH, \tag{8.12}$$

which implies $S_{n,1} = O_P(H^{1/2})$ and with (8.11) proves that $\xi_{n,t} = O_P(H^{-1/2})$, which implies (8.10). To prove (8.12), recall that by Assumption 3.1, u_k, y_{k-1} are uncorrelated random variables. Therefore, by (3.9),

$$\begin{aligned}\mathrm{Var}(S_{n,1}) &= E\Big(\sum_{k=1}^n b_{tk}u_k y_{k-1}\Big)^2 \\ &= \sum_{k=1}^n b^2_{tk}E[u^2_k]E[y^2_{k-1}] \le C\sum_{k=1}^n b^2_{tk} = CB^2_{H,t} = O(H).\end{aligned}$$

To prove (4.8), note that by Lemma 8.3 and by (8.24) of Lemma 8.2 below,

$$\frac{T_{H,t}}{\sqrt{1-\rho^2_{n,t}}}\xi_{n,t} = \frac{\frac{\sqrt{1-\rho^2_{n,t}}}{B_{H,t}}S_{n,1}}{\frac{1-\rho^2_{n,t}}{\beta_{H,t}}V_n} \to_D \frac{N(0,\sigma^4_u)}{\sigma^2_u} = N(0,1).$$

(ii) Observe that (4.7)-(4.8) imply (4.9), which completes proof of Theorem 4.1. □

Next three lemmas contain auxiliary results.

Lemma 8.1. *Under Assumptions of Theorem 4.1,*

$$\sum_{k=1}^n |\rho_{n,k-1} - \rho_{n,t}|b_{t,k+i}y^2_{k-1} = O_P((\frac{H}{n})^\gamma H), \quad i = 0, 1. \tag{8.13}$$

Proof of Lemma 8.1. We prove (8.13) for $i = 0$. (Proof for $i = 1$ follows by the same argument). Let $t_n := \sum_{k=1}^n |a_{k-1} - a_t|b_{tk}y^2_{k-1}$. Since

$$\rho_{n,t} - \rho_{n,k} = \frac{a_t - a_k}{a_{\max}}, \quad a_{\max} := \max_{0\le k\le n} |a_k|,$$

then the left hand side of (8.13) can be written as $t_n/a_{\max}$. We show below that

$$n^{-\gamma}a_{max} \to_D \sup_{0\le\tau\le1} |W_\tau| > 0, \quad \text{in prob.,} \tag{8.14}$$

$$t_n = O_P(H^{\gamma+1}). \tag{8.15}$$

Then

$$\frac{t_n}{a_{max}} = (\frac{H}{n})^\gamma \frac{H^{-\gamma}t_n}{n^{-\gamma}a_{\max}} = (\frac{H}{n})^\gamma O_P(H),$$

which proves (8.13).

Proof of (8.14). By (3.3), $a_k = a_0 + \sum_{j=1}^{k} v_j = a_0 + S_{k,v}$, where, $S_{k,v} := \sum_{j=1}^{k} v_j$, $k = 1, \cdots, n$. Under the weak convergence assumption (3.4),

$$\max_{1\le k\le n} |n^{-\gamma} S_{k,v}| = \sup_{0\le\tau\le 1} |n^{-\gamma} S_{n+1}(\tau)| \tag{8.16}$$
$$\to_D \sup_{0\le\tau\le 1} |W_\tau|, \quad n\to\infty.$$

Since

$$-|a_0| + \max_{1\le k\le n} |S_{k,v}| \le a_{max} \le |a_0| + \max_{1\le k\le n} |S_{k,v}|,$$

and $a_0 = O_P(1)$, this proves (8.14).

Proof of (8.15). 1. Assume that K has a finite support. Then there exists $L > 0$, such that $b_{tk} = 0$, when $|t-k| > LH$. Without restriction of generality, assume that $L = 1$. Then

$$|t_n| \le \max_{k:|t-k|\le H} |a_{k-1} - a_t| \sum_{k=1}^{n} b_{tk} y_{k-1}^2.$$

By (8.23) of Lemma 8.2 below, $\sum_{k=1}^{n} b_{tk} y_{k-1}^2 = O_P(H)$. Therefore to obtain (8.15), it suffices to show that

$$\max_{k:|t-k|\le H} |a_{k-1} - a_t| = O_P(H^\gamma). \tag{8.17}$$

Recall that $\{v_j\}$ is a stationary process. Therefore

$$\max_{t<k\le t+H} |a_{k-1} - a_t| \le \max_{1\le l\le H} |\sum_{j=t+1}^{t+l} v_j| =_D \max_{1\le l\le H} |S_{l,v}| = O_P(H^\gamma),$$

by (8.16). Similarly, $\max_{t-H\le k\le t} |a_t - a_{k-1}| = O_P(H^\gamma)$, which completes proof of (8.17).

2. Suppose that K has an infinite support. We shall show that

$$E|t_n| \le CH^{\gamma+1}, \tag{8.18}$$

which implies (8.15) and completes the proof of this lemma.

To prove (8.18), we shall need two facts:

$$\max_{1\le j\le n} Ey_j^4 \le C, \tag{8.19}$$
$$E(a_t - a_k)^2 \le C|t-k|^{2\gamma}, \quad k = 1, \cdots, n. \tag{8.20}$$

By Assumption 4.1, $Eu_1^4 < \infty$ and $Ey_0^4 < \infty$. Thus, by Assumption 3.1, (3.8) and $|c_{t,j}| \le |\rho|^j$,

$$Ey_t^4 = E\Big(\sum_{j=0}^{t-1} c_{t,j} u_{t-j} + c_{t,t} y_0\Big)^4$$

$$
\begin{aligned}
&\le 4E\Big(\sum_{j=0}^{t-1} c_{t,j}u_{t-j}\Big)^4 + 4Ec_{t,t}^4Ey_0^4\\
&\le \sum_{j_1,\cdots,j_4=0}^{t-1} |\rho|^{j_1+\cdots+j_4}(Eu_{t-j_1}^4\cdots Eu_{t-j_4}^4)^{1/4} + C|\rho|^t\\
&\le C(\sum_{j=0}^{\infty}|\rho|^j)^4 < \infty, \qquad (8.21)
\end{aligned}
$$

which proves (8.19). To show (8.20), without restriction of generality assume that $t > k$. Then by stationarity of $\{v_j\}$ and Assumption 4.1,

$$
E(a_t - a_k)^2 = E(\sum_{j=k+1}^{t} v_j)^2 = E(\sum_{j=1}^{t-k} v_j)^2 \le C|t-k|^{2\gamma}, \quad t \ge j,
$$

which proves (8.20).

Now applying (8.19)-(8.20), one obtains

$$
\begin{aligned}
H^{-\gamma-1}E|t_n| &\le H^{-\gamma-1}\sum_{k=1}^{n} b_{tk}(E(a_{k-1}-a_t)^2)^{1/2}(Ey_{k-1}^4)^{1/2}\\
&\le CH^{-1}\sum_{k=1}^{n}(\frac{|t-k|+1}{H})^{\gamma}K(\frac{t-k}{H})\\
&\to \int_{\mathbb{R}}|x|^{\gamma}K(x)dx < \infty, \quad n\to\infty, \qquad (8.22)
\end{aligned}
$$

by (4.3) and TDC, which proves (8.18) and completes proof of lemma. □

The next lemma deals with properties of the sum $V_n = \sum_{k=1}^n b_{tk}y_{k-1}^2$.

Lemma 8.2. *Under Assumptions 3.1 and 3.2,*

$$
E\big(\sum_{k=1}^{n} b_{tk}y_{k-1}^2\big) \le CH, \qquad (8.23)
$$

$$
\frac{1-\rho_{n,t}^2}{\beta_{H,t}}\sum_{k=1}^{n} b_{tk}y_{k-1}^2 \to_D \sigma_u^2, \qquad (8.24)
$$

with $\beta_{H,t}$ as in (8.8).

Proof of Lemma 8.2. To prove (8.23), note that by (3.10), $Ey_k^2 \le C$, $1\le k\le n$. Therefore

$$
EV_n = \sum_{k=1}^{n} b_{tk}Ey_k^2 \le C\sum_{k=1}^{n} b_{tk} = C\beta_{H,t} = O(H),
$$

by (8.8).

To show (8.24), let $V_n' := \sum_{k=2}^n b_{tk} y_{k-2}^2$.

We shall show that

$$V_n - \rho_{n,t}^2 V_n' = \beta_{H,t}\sigma_u^2 + o(\beta_{H,t}), \tag{8.25}$$

$$V_n - V_n' = o_P(H). \tag{8.26}$$

Since by (8.8), $\beta_{H,t} \sim CH$, applying (8.26) in (8.25) yields

$$(1-\rho_{n,t}^2)V_n = V_n - \rho_{n,t}^2 V_n' + o_P(\beta_{H,t}) = \sigma_u^2 \beta_{H,t} + o_P(\beta_{H,t}),$$

which proves (8.24).

Proof of (8.25). Note that $b_{t1}y_0^2 = O_P(1)$, because $E|b_{t1}y_0^2| \le E|y_0^2| < \infty$. Therefore, using

$$\begin{aligned} y_{k-1}^2 &= (\rho_{n,k-2}y_{k-2} + u_{k-1})^2 \\ &= \rho_{n,k-2}^2 y_{k-2}^2 + 2u_{k-1}\rho_{n,k-2}y_{k-2} + u_{k-1}^2, \end{aligned}$$

$$\begin{aligned} V_n - \rho_{n,t}^2 V_n' &= \sum_{k=2}^n b_{tk}(y_{k-1}^2 - \rho_{n,t}^2 y_{k-2}^2) + O_P(1) \\ &= \sum_{k=2}^n b_{tk}(\rho_{n,k-2}^2 - \rho_{n,t}^2)y_{k-2}^2 + 2\sum_{k=2}^n b_{tk}u_{k-1}\rho_{n,k-2}y_{k-2} + \sum_{k=2}^n b_{tk}u_{k-1}^2 + O_P(1) \\ &=: Q_{n,1} + Q_{n,2} + Q_{n,3} + O_P(1). \end{aligned}$$

We shall show that

$$Q_{n,1} = o_P(H), \tag{8.27}$$

$$Q_{n,2} = o_P(H), \tag{8.28}$$

$$Q_{n,3} = \beta_H^2 \sigma_u^2 + o_P(H), \tag{8.29}$$

which proves (8.25).

To evaluate $Q_{n,1}$, use $|\rho_{n,k}| \le 1$, to obtain

$$|\rho_{n,k-2}^2 - \rho_{n,t}^2| = |\rho_{n,k-2} - \rho_{n,t}||\rho_{n,k-2} + \rho_{n,t}| \le 2|\rho_{n,k-2} - \rho_{n,t}|$$

and Lemma 8.1, to obtain

$$\begin{aligned} |Q_{n,1}| &\le 2\sum_{k=2}^n b_{t,k}|\rho_{n,k-2} - \rho_{n,t}|y_{k-2}^2 \le 2\sum_{k=1}^n b_{t,k+1}|\rho_{n,k-1} - \rho_{n,t}|y_{k-1}^2 \\ &= O_P((\frac{H}{n})^\gamma H) = o_P(H), \end{aligned}$$

because $H = o(n)$, which proves (8.27).

To evaluate $Q_{n,2}$, note that $u_{k-1}\rho_{n,k-2}y_{k-2}$, $k = 2, \cdots, n$ are uncorrelated random variables. Therefore

$$\begin{aligned}\mathrm{Var}(Q_{n,2}) &= E\Big(2\sum_{k=2}^{n} b_{tk}u_{k-1}\rho_{n,k-2}y_{k-2}\Big)^2 \\ &= 4\sum_{k=2}^{n} b_{tk}^2 Eu_{k-1}^2 E[\rho_{n,k-2}^2 y_{k-2}^2] \\ &\le C\sum_{k=2}^{n} b_{tk}^2 \le CB_H^2 = O(H),\end{aligned}$$

by $|\rho_{n,k-2}| \le 1$, (3.9) and (8.8), which implies that $Q_{n,2} = O_P(H^{1/2}) = o_P(H)$ and proves (8.28).

Finally,

$$Q_{n,3} = \sum_{k=2}^{n} b_{tk}u_{k-1}^2 = \sum_{k=2}^{n} b_{tk}\sigma_u^2 + \sum_{k=2}^{n} b_{tk}(u_{k-1}^2 - \sigma_u^2).$$

Note, that $\sum_{k=2}^{n} b_{tk}\sigma_u^2 \sim \beta_{H,t}\sigma_u^2$. So, to show (8.29), it remains to prove that

$$\widetilde{Q}_{n,3} := \sum_{k=2}^{n} b_{tk}(u_{k-1}^2 - \sigma_u^2) = o_P(H).$$

For any $\epsilon > 0$, one can choose $L > 0$ such that $Eu_1^2 I(|u_1| > L) \le \varepsilon$. Write $u_k^2 - \sigma_u^2 = \eta_{k,1} + \eta_{k,2}$, where

$$\begin{aligned}\eta_{k,1} &= u_k^2 I(|u_k| \le L) - E[u_k^2 I(|u_k| \le L)], \\ \eta_{k,2} &= u_k^2 I(|u_k| > L) - E[u_k^2 I(|u_k| > L)].\end{aligned}$$

Then

$$\widetilde{Q}_{n,3} := \sum_{k=2}^{n} b_{tk}\eta_{k-1,1} + \sum_{k=2}^{n} b_{tk}\eta_{k-1,2} := q_{n,1} + q_{n,2}.$$

Since $\eta_{k,1}$ are i.i.d. variables with a finite variance, and $b_{tk} \le \sup_x K(x) < \infty$, then

$$\mathrm{Var}(q_{n,1}) = E\eta_{1,1}^2 \sum_{k=2}^{n} b_{tk}^2 \le C\sum_{k=2}^{n} b_{tk}^2 \le CB_{H,t}^2 = O(H).$$

Hence, $q_{n,1} = o_P(H)$, for any fixed L. On the other hand, $E|\eta_{k,2}| \le 2Eu_1^2 I(|u_1| > L) \le 2\varepsilon$, and

$$E|q_{n,2}| \le \sum_{k=2}^{n} b_{tk}E|\eta_{k-1,2}| \le 2\varepsilon\beta_{H,t} \le 2\varepsilon\, CH = o(H), \quad \varepsilon \to 0.$$

This completes proof of (8.29).

Proof of (8.26). Changing summation $k \to k-1$ in V_n, one obtains:

$$\begin{aligned} V_n - V_n' &= \sum_{k=3}^{n+1} b_{t,k-1} y_{k-2}^2 - \sum_{k=2}^{n} b_{tk} y_{k-2}^2 \\ &= \sum_{k=2}^{n} (b_{t,k-1} - b_{tk}) y_{k-2}^2 + b_{t,n} y_{n-1}^2 - b_{t,1} y_0^2. \end{aligned}$$

Since by (3.9), $E|b_{t,n} y_{n-1}^2 - b_{t,1} y_0^2| \le E|y_{n-1}^2| + E|y_0^2| \le C$, and $E y_{k-2}^2 \le C$, $k = 2, \cdots, n$, then

$$E|V_n - V_n'| \le C \sum_{k=2}^{n} |b_{t,k-1} - b_{tk}| + C = o(H),$$

where the last bound $o(H)$ follows from the continuity of K and assumption (4.3), using theorem of dominated convergence. This completes proof of (8.26) and the lemma. □

Lemma 8.3. *Under assumptions of Theorem 4.1,*

$$\frac{\sqrt{1-\rho_{n,t}^2}}{B_{H,t}} \sum_{k=1}^{n} b_{tk} u_k y_{k-1} \to_D N(0, \sigma_u^4). \tag{8.30}$$

Proof of Lemma 8.3 1. First we prove convergence (8.30) in case when

$$E u_1^4 < \infty, \ \ E y_0^4 < \infty. \tag{8.31}$$

By definition, the random variables u_k are independent of $\rho_{n,t}$ and $y_{k-1}, \cdots, y_1$. Therefore,

$$\xi_k := u_k b_{tk} \frac{\sqrt{1-\rho_{n,t}^2}}{B_{H,t}} y_{k-1}, \quad k = 1, \cdots, n$$

is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_k = \sigma(u_k, \cdots, u_0, a_n, \cdots, a_0)$. By the central limit theorem for martingale differences, to show asymptotic normality (8.30), it suffices to prove that

$$\sum_{k=1}^{n} E[\xi_k^2 | \mathcal{F}_{k-1}] \to_p \sigma_u^4, \tag{8.32}$$

$$\sum_{k=1}^{n} E[\xi_k^2 I(|\xi_k| \ge \delta)] \to_p 0, \tag{8.33}$$

for any $\delta > 0$. Note that

$$\sum_{k=1}^{n} E[\xi_k^2 | \mathcal{F}_{k-1}] = E[u_1^2] \frac{1-\rho_{n,t}^2}{B_{H,t}^2} \sum_{k=1}^{n} b_{tk}^2 y_{k-1}^2 \to_p \sigma_u^4,$$

by (8.24). Next,

$$\sum_{k=1}^{n} E[\xi_k^2 I(|\xi_k| \geq \delta)] \leq \delta^{-1} \sum_{k=1}^{n} E[\xi_k^4]$$
$$\leq \delta^{-1} B_{H,t}^{-4} \sum_{k=1}^{n} b_{tk}^4 E[u_k^4 y_{k-1}^4].$$

Notice that $E[u_k^4 y_{k-1}^4] = E[u_k^4]E[y_{k-1}^4] \leq C$, $k = 1, \cdots, n$, because by (8.31) and (8.21), $Ey_t^4 \leq C(\sum_{j=0}^{\infty} |\rho|^j)^4 < \infty$. Thus, using $b_{tk}^4 \leq C b_{tk}^2$, one obtains

$$\sum_{k=1}^{n} E[\xi_k^2 I(|\xi_k| \geq \delta)] \leq \delta^{-1} C B_{H,t}^{-4} \sum_{k=1}^{n} b_{tk}^2 = \delta^{-1} C B_{H,t}^{-2} \to 0,$$

since $B_{H,t}^2 \sim CH$, by (8.8). This proves (8.33) and completes proof of (8.30).

2. In case, when Eu_1^4 and Ey_0^4 are not finite, we use the truncation argument. Let $L > 0$. Define

$$\begin{aligned}
\zeta_{k,1} &= u_k I(|u_k| \leq L) - E[u_k I(|u_k| \leq L)], \\
\zeta_{k,2} &= u_k I(|u_k| > L) - E[u_k I(|u_k| > L)], \\
y_{0,1} &= y_0 I(|y_0| \leq L) - E[y_0 I(|y_0| \leq L)], \\
y_{0,2} &= y_0 I(|y_0| > L) - E[y_0 I(|y_0| > L)].
\end{aligned}$$

Then $u_k = \zeta_{k,1} + \zeta_{k,2}$, and by (3.8), one can write

$$\begin{aligned}
y_t &= \sum_{j=0}^{t-1} c_{t,j} u_{t-j} + c_{t,t} y_0 \\
&= \Big(\sum_{j=0}^{t-1} c_{t,j} \zeta_{t-j,1} + c_{t,t} y_{0,1}\Big) + \Big(\sum_{j=0}^{t-1} c_{t,j} \zeta_{t-j,2} + c_{t,t} y_{0,2}\Big) \\
&=: y_{t,1} + y_{t,2}.
\end{aligned}$$

According to (3.8), $y_{t,1}$, $t = 1, 2, \cdots, n$ is a solution of equations $y_{t,1} = \rho_{n,t-1} y_{t-1,1} + \zeta_{t,1}$, $t = 1, \cdots, n$, with the initial condition $y_{0,1}$. Write the summand of (8.30) as $u_k y_{k-1} = \zeta_{k,1} y_{k-1,1} + (u_k y_{k-1} - \zeta_{k,1} y_{k-1,1})$.

Since $E\zeta_{1,1}^4 < \infty$, then as it was shown above in 1), for any $L > 0$,

$$\frac{\sqrt{1 - \rho_{n,t}^2}}{B_{H,t}} \sum_{k=1}^{n} b_{tk} \zeta_{k,1} y_{k-1,1} \to_D N(0, \sigma_{\zeta,1}^4), \tag{8.34}$$

where $\sigma_{\zeta,1}^2 = E\zeta_{k,1}^2 \to \sigma_u^2, \quad L \to \infty$. On the other hand, by (3.9),

$$Ey_{k-1,1}^2 \leq \frac{\sigma_u^2 + Ey_0^2}{1 - \rho^2}, \quad E\zeta_{k,1}^2 \leq 2\sigma_u^2,$$

$$Ey^2_{k-1,2} \le \frac{E\zeta^2_{k,2} + Ey^2_{0,2}}{1-\rho^2} \to 0, \quad E\zeta^2_{k,2} \to 0, \quad L \to \infty.$$

Note that the variables $z_k := u_k y_{k-1} - \zeta_{k,1} y_{k-1,1}$, are uncorrelated, and

$$\begin{aligned} Ez_k^2 &= E(\zeta_{k,2} y_{k-1,1} + \zeta_{k,1} y_{k-1,2} + \zeta_{k,2} y_{k-1,2})^2 \\ &\le C(E\zeta^2_{k,2} + Ey^2_{k-1,2}) \to 0, \quad L \to \infty. \end{aligned}$$

Therefore, as $n \to \infty$, $L \to \infty$,

$$\mathrm{Var}(\sum_{k=1}^n b_{tk} z_k) = \sum_{k=1}^n b^2_{tk} E z_k^2 = o(B^2_{H,t}),$$
$$\frac{\sqrt{1-\rho^2_{n,t}}}{B_{H,t}} \sum_{k=1}^n b_{tk} z_k = o_P(1),$$

which together with (8.34) proves (8.30) and completes the proof of lemma. $\square$

Proof of Theorem 4.2. Proof of Theorem 4.2 follows the same line as that of Theorem 4.1. It is based on Theorem 3.1 and Lemmas 8.1-8.3. Observe, that under assumption of Theorem 4.2, conditions of Theorem 3.1 are satisfied. Proof of Lemmas 8.1-8.3 indicates, that under assumption of Theorem 4.2, Lemmas 8.2-8.3 holds true, whereas Lemma 8.1 needs to be replaced by

Lemma 8.4. *Under Assumptions of Theorem 4.2,*

$$\sum_{k=1}^n |\rho_{n,k-1} - \rho_{n,t}| b_{t,k+i} y^2_{k-1} = O_P((\frac{H}{n})H), \quad i = 0, 1. \tag{8.35}$$

Proof of Lemma 8.4. Let $i = 0$. (Proof in the case $i = 1$ follows using the same argument). Then the left hand side of (8.35) can be written as $t_n / a_{\max}$, where t_n and a_{max} are same as in proof of Lemma 8.1.

(i) Assume that a_t satisfies (4.12). We show that

$$\lim_n n^{-1} a_{max} > 0, \quad \text{in prob.,} \tag{8.36}$$
$$t_n = O_P(H^2), \tag{8.37}$$

which yields (8.35).

By (4.13) and (8.14),

$$\begin{aligned} n^{-1} a_{max} &= \max_{k=1,\cdots,n} n^{-1} \left| \sum_{j=1}^k \mu(j/n) \right| + o_P(1) \\ &\to_D \max_{0 \le \tau \le 1} \left| \int_0^\tau \mu(x) dx \right| > 0, \end{aligned}$$

which proves (8.36). By (4.13)

$$\begin{aligned} t_n &\le \sum_{k=1}^{n} \left| \sum_{j=1}^{k-1} \mu(j/n) - \sum_{j=1}^{t} \mu(j/n) \right| b_{tk} y_{k-1}^2 \\ &+ \sum_{k=1}^{n} \left| \sum_{j=1}^{k-1} v_j - \sum_{j=1}^{t} v_j \right| b_{tk} y_{k-1}^2 =: t_{n,1} + t_{n,2}. \end{aligned}$$

By (8.15), $t_{n,2} = O_P(H^{\gamma+1})$. On the other hand, since $|\sum_{j=1}^{k-1} \mu(j/n) - \sum_{j=1}^{t} \mu(j/n)| \le C|t-k|$, this together with $Ey_t^2 \le C$, $1 \le j \le n$ of (3.9) yields

$$\begin{aligned} H^{-2} t_{n,1} &\le C H^{-1} \sum_{k=1}^{n} \frac{|t-k|}{H} b_{tk} y_{k-1}^2 \\ &= O_P(1) H^{-1} \sum_{k=1}^{n} \frac{|t-k|}{H} b_{tk} = O_P(1), \end{aligned}$$

by the same argument as in (8.22), which proves (8.37).

(ii) Assume that a_t satisfies (4.14). We show that

$$\lim_n a_{max} > 0, \tag{8.38}$$

$$t_n = O_P(H^2 n^{-1}), \tag{8.39}$$

which yields (8.35).

Note, that (8.38) holds true, since

$$a_{max} = \max_{1\le k\le n} |\varphi(k/n)| > 0, \quad n \to \infty.$$

Next, by the mean value theorem,

$$|\varphi((k-1)/n) - \varphi(t/n)| \le \sup_{0\le x\le 1} |\varphi'(x)|\, |t-k|/n \le C|t-k|/n,$$

one obtains

$$\begin{aligned} H^{-2} t_n &\le C H^{-1} \sum_{k=1}^{n} |\varphi((k-1)/n) - \varphi(t/n)| b_{tk} y_{k-1}^2 \\ &\le C H^{-1} \sum_{k=1}^{n} \frac{|t-k|}{H} b_{tk} y_{k-1}^2 \\ &= O_P(1) H^{-1} \sum_{k=1}^{n} \frac{|t-k|}{H} b_{tk} = O_P(1), \end{aligned}$$

as above in 1. This completes proof of lemma and of Theorem 4.2. □

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.

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 (2010): *Large Sample Inference for Long memory Processes*. Singapore: World Scientific.

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. (2007): "Estimating Deterministically Time Varying Variances in Regression Models," *Economics Letters*, 97, 97–104.

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 (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74(4), 763–789.

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.

MUMTAZ, H., AND P. SURICO (2009): "Time varying yield curve dynamics and monetary policy," *Journal of Applied Econometrics*, 24, 895–913.

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 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.

PERRON, P. (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153–162.

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.

PLOBERGER, W., AND W. KRAMER (1992): "The CUSUM Test with OLS Residuals," *Econometrica*, 60, 271–285.

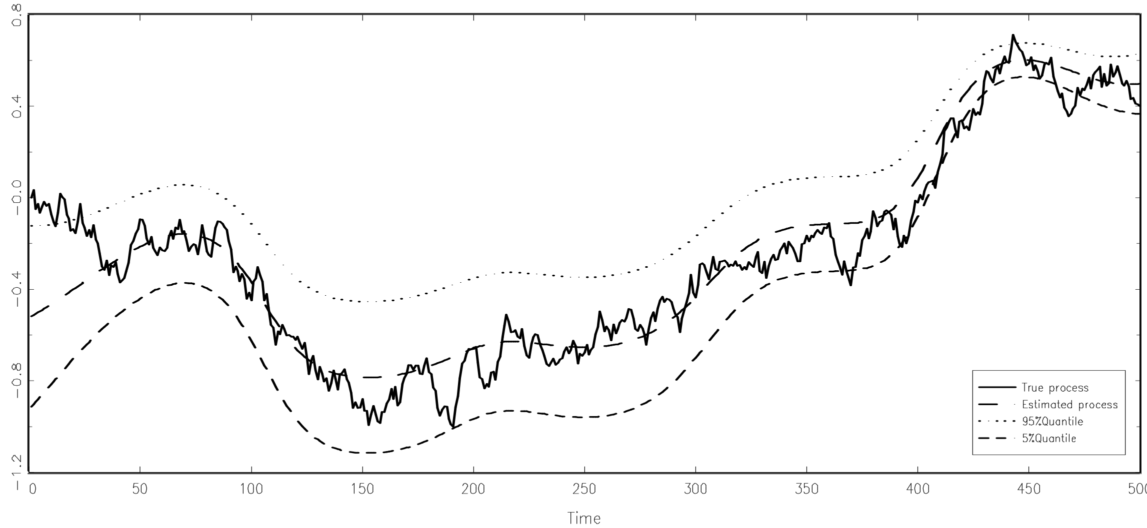
PRIESTLEY, M. (1965): "Evolutionary Spectra and Nonstationary Processes," *Journal of Royal Statistical Society, Series B*, 27, 204–237.

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

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.

Figure 1: Realisation of $\rho_{n,t}$, its estimate and confidence bands for the normal (Panel 1) and flat (Panel 2) kernels.



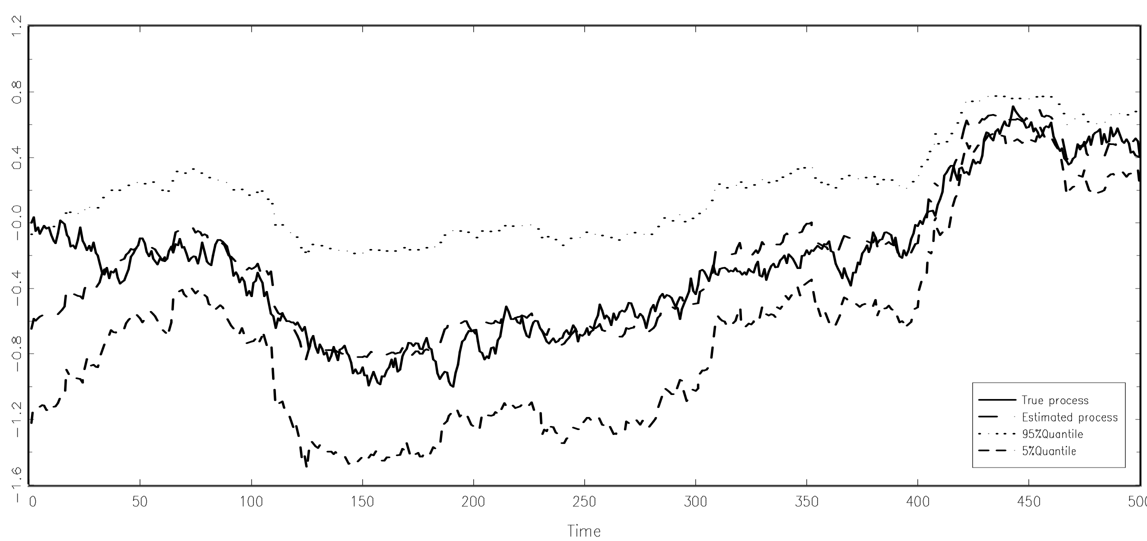


Table 8.1: MSE results for $\rho_{n,t} = 0.9\frac{a_t}{max_{i\leq n}|a_t|}$ and short memory for v_t. The model is $y_{n,t} = \rho_{n,t}y_{n,t-1} + u_t$, $a_t = a_{t-1} + v_t$, $u_t = \epsilon_{1t}$, $v_t = \phi v_{t-1} + \epsilon_{2t}$.

Normal kernel, bandwidth n^{α}							
ϕ	α/n	50	100	200	400	800	1000
	0.2	0.096	0.078	0.069	0.060	0.051	0.049
	0.4	0.047	0.036	0.027	0.020	0.016	0.015
0	0.5	0.036	0.025	0.018	0.014	0.011	0.011
	0.6	0.027	0.019	0.014	0.012	0.013	0.012
	0.8	0.021	0.017	0.017	0.023	0.029	0.033
	0.2	0.098	0.080	0.067	0.059	0.052	0.049
	0.4	0.047	0.034	0.026	0.021	0.016	0.014
0.2	0.5	0.034	0.024	0.018	0.014	0.012	0.011
	0.6	0.028	0.018	0.014	0.013	0.012	0.012
	0.8	0.022	0.015	0.019	0.023	0.031	0.035
	0.2	0.096	0.082	0.069	0.058	0.051	0.048
	0.4	0.048	0.034	0.025	0.020	0.015	0.014
0.5	0.5	0.035	0.023	0.017	0.013	0.011	0.011
	0.6	0.027	0.018	0.014	0.012	0.012	0.012
	0.8	0.021	0.015	0.018	0.022	0.030	0.035
	0.2	0.094	0.080	0.066	0.057	0.049	0.048
	0.4	0.046	0.032	0.024	0.018	0.014	0.013
0.9	0.5	0.032	0.022	0.015	0.012	0.009	0.009
	0.6	0.026	0.016	0.012	0.010	0.011	0.011
	0.8	0.018	0.013	0.016	0.023	0.033	0.037

Table 8.2: MSE results for $\rho_{n,t} = \rho \frac{a_t}{max_{i \leq n}|a_t|}$ and long memory for v_t. The model is $y_{n,t} = \rho_{n,t} y_{n,t-1} + u_t$, $a_t = a_{t-1} + v_t$, $u_t = \epsilon_{1t}$, $v_t \sim ARFIMA(0, d-1, 0)$.

Normal kernel, bandwidth n^α							
		ρ=0.9					
d	α/n	50	100	200	400	800	1000
	0.2	0.171	0.137	0.120	0.102	0.088	0.085
	0.4	0.135	0.107	0.086	0.070	0.059	0.055
0.51	0.5	0.130	0.102	0.079	0.066	0.056	0.052
	0.6	0.133	0.100	0.082	0.066	0.058	0.054
	0.8	0.134	0.106	0.088	0.073	0.065	0.062
	0.2	0.148	0.119	0.099	0.083	0.071	0.068
	0.4	0.117	0.085	0.064	0.048	0.037	0.034
0.75	0.5	0.112	0.082	0.060	0.045	0.034	0.032
	0.6	0.115	0.085	0.064	0.049	0.039	0.038
	0.8	0.131	0.107	0.085	0.073	0.064	0.062
	0.2	0.121	0.099	0.084	0.072	0.064	0.061
	0.4	0.078	0.053	0.039	0.029	0.021	0.019
1.25	0.5	0.075	0.050	0.032	0.021	0.015	0.013
	0.6	0.080	0.054	0.034	0.023	0.016	0.014
	0.8	0.115	0.093	0.074	0.061	0.051	0.047
	0.2	0.118	0.097	0.084	0.072	0.063	0.061
	0.4	0.072	0.048	0.035	0.026	0.020	0.019
1.49	0.5	0.068	0.041	0.027	0.018	0.013	0.012
	0.6	0.071	0.045	0.027	0.017	0.011	0.010
	0.8	0.111	0.085	0.064	0.054	0.043	0.041

Table 8.3: MSE results for $\rho_{n,t} = 0.9$. The model is $y_{n,t} = 0.9 y_{n,t-1} + u_t$, $u_t = \epsilon_{1t}$.

Normal kernel, bandwidth n^α						
α/n	50	100	200	400	800	1000
0.2	0.077	0.062	0.050	0.041	0.034	0.032
0.4	0.034	0.019	0.014	0.009	0.006	0.005
0.5	0.020	0.012	0.007	0.004	0.003	0.002
0.6	0.016	0.007	0.004	0.002	0.001	0.001
0.8	0.009	0.004	0.002	0.001	0.000	0.000

Figure 2: Time-Varying AR Coefficient and 95% confidence bands from an $AR(1)$ model, with a time-varying 'constant' term, for CPI inflation using a standard normal kernel for 6 countries: Australia, Canada, Japan, Switzerland, US and UK. Every panel also reports the value of the autoregressive coefficient estimated in a fixed coefficient $AR(1)$ together with 95% confidence bands.

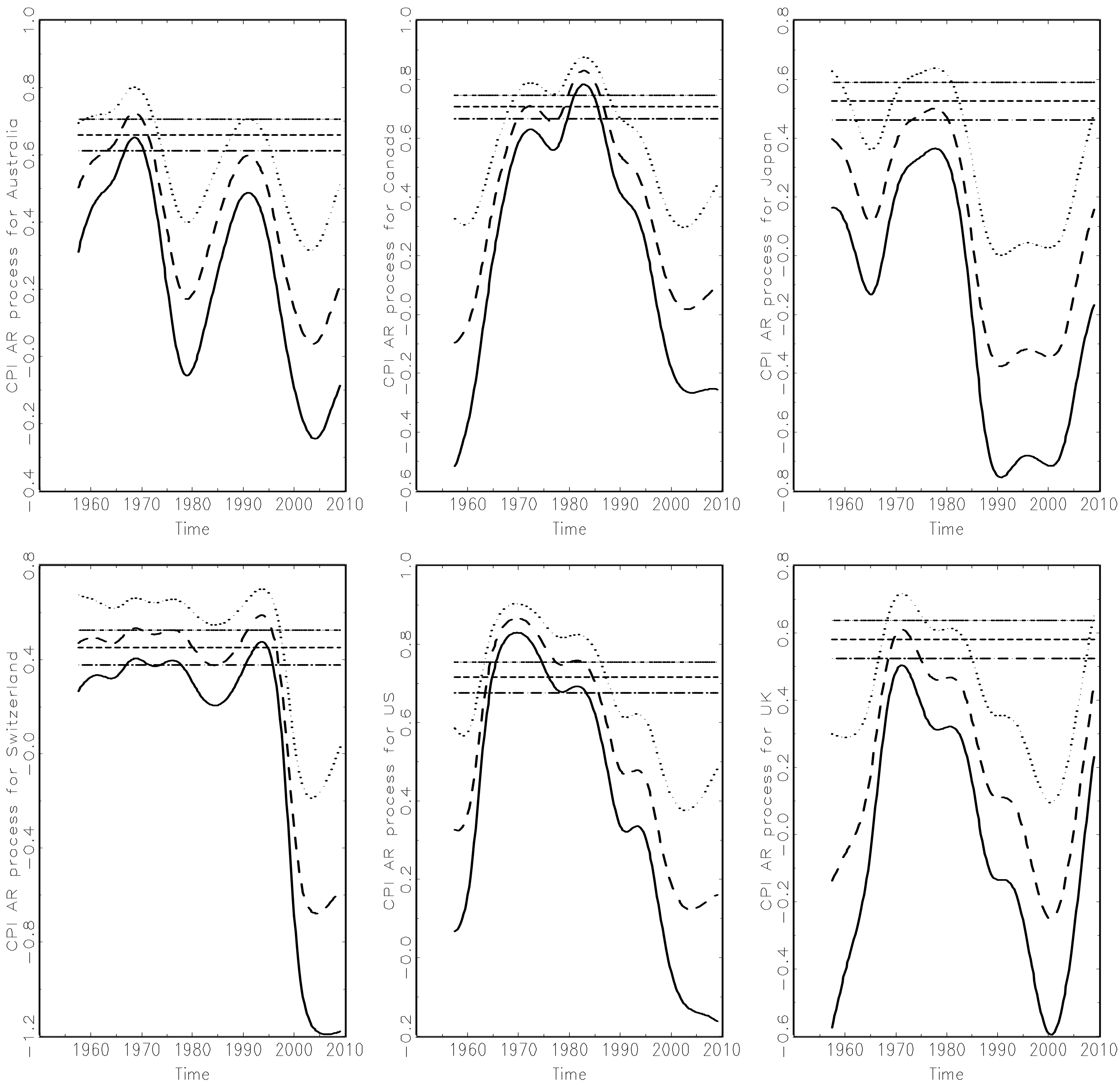


Figure 3: Time-Varying AR Coefficient from an $AR(1)$ model, with a time-varying 'constant' term, for real exchange rates using a standard normal kernel for 6 countries: Australia, Canada, Japan, Norway, Switzerland and UK. Every panel also reports the value of the autoregressive coefficient estimated in a fixed coefficient $AR(1)$ together with 95% confidence bands.