INFERENCE IN A STATIONARY/NONSTATIONARY AUTOREGRESSIVE TIME-VARYING-PARAMETER MODEL

By

Donald W. K. Andrews and Ming Li

September 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2389R1

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Inference in a Stationary/Nonstationary Autoregressive Time-Varying-Parameter Model

Donald W. K. Andrews and Ming Li*

September 23, 2024

Abstract

This paper considers nonparametric estimation and inference in first-order autoregressive (AR(1)) models with deterministically time-varying parameters. A key feature of the proposed approach is to allow for time-varying stationarity in some time periods, time-varying nonstationarity (i.e., unit root or local-to-unit root behavior) in other periods, and smooth transitions between the two. The estimation of the AR parameter at any time point is based on a local least squares regression method, where the relevant initial condition is endogenous. We obtain limit distributions for the AR parameter estimator and t-statistic at a given point τ in time when the parameter exhibits unit root, local-to-unity, or stationary/stationary-like behavior at time τ. These results are used to construct confidence intervals and median-unbiased interval estimators for the AR parameter at any specified point in time. The confidence intervals have correct asymptotic coverage probabilities with the coverage holding uniformly over stationary and nonstationary behavior of the observations.

Keywords: Autoregressive time-varying-parameter model, endogenous initial condition, nonparametric estimation, confidence interval.

*Andrews: Department of Economics, Yale University. Email: donald.andrews@yale.edu. Li: Department of Economics, National University of Singapore. Email: mli@nus.edu.sg. The authors thank the participants of the Yale econometrics seminar, the SMU econometrics seminar, the Stock and Watson conference, and the CUHK Econometrics Workshop for helpful comments.

1 Introduction

Autoregressive models — stationary or nonstationary — are workhorse models in econometric time series. In this paper, we consider a deterministically time-varying parameter (TVP) autoregressive model that allows for stationary and non-stationary behavior at different points in the time period of interest. Thus, the level of persistence of the time series can change over the time period. The motivation for considering such a model is that the economy is in continual transition due to technological, institutional, political, and demographic changes. The model considered allows for the intercept and error variance of the model also to be time-varying, not just the AR parameter.

The estimation method employed is local least squares, which depends on a tuning parameter h that determines the local neighborhood that is considered. We construct a confidence interval (CI) for the value of the AR parameter at a time point τ via the inversion of tests, as is common in the literature for constant parameter AR models, e.g., see Stock (1991), Andrews (1993), Hansen (1999), Mikusheva (2007), and Andrews and Guggenberger (2014). We show that the CI's have correct uniform asymptotic coverage probability for a parameter space that allows the time series to be stationary in parts of the time period and nonstationary in other parts, using the approach in Andrews, Cheng, and Guggenberger (2020). We also construct asymptotically median-unbiased interval estimators (MUE's) in an analogous fashion.

In the TVP case, the initial condition is endogenous due to the choice of the local neighborhood and depends on the potentially different behavior of the time series prior to the local neighborhood. For a given time point τ of interest, we find that the asymptotic distributions of the LS estimator and t-statistic depend on the endogenous initial condition in the local-to-unity case. This is analogous to the asymptotic effect of the initial condition–under certain assumptions on the initial condition–in constant parameter AR models, see Elliott (1999), Elliott and Stock (2001), and Müller and Elliott (2003). On the other hand, the endogenous initial condition does not affect the asymptotic distributions when the AR coefficient at time τ is more distant from one than local-to-unity.

We note that whether a time series is local-to-unity at time τ, or not, depends on τ and the chosen bandwidth h. It is important to provide asymptotic results that hold uniformly over a parameter space that does not depend on h, which we do.

We introduce a method for determining the tuning parameter h based on a forecast-error criterion. We provide conditions under which this data-dependent choice of h is asymptotically equivalent to an infeasible choice that minimizes the unobserved "empirical loss." These results are similar to results for i.i.d. models given in Li (1987) and Andrews (1991).

We provide Monte Carlo simulation results for the methods introduced in the paper. We consider true autoregressive functions whose shapes are sinusoidal, linear, partly flat/partly linear, flat, and kinked linear. We consider cases where the functions are close, or equal, to one in some regions, but different from one by varying amounts in other regions. We find that the proposed CI has reasonably good coverage probabilities and short average lengths for most of the data-generating processes considered. For example, nominal 95% CI's are found to have finite-sample coverage probabilities ranging from 92.5% to 96% in 88.8% of the cases, across 205 cases. The lowest coverage probabilities, in the range from 87.5% to 90%, occur only in 2.0% of the cases. The MUE's are found to have very small finite-sample median bias across the different cases considered. The magnitude of the data-dependent choice of h varies widely depending upon the shape of the autoregressive function, as desired.

We provide some empirical applications of the methods to monthly inflation and real exchange rates in several countries using data from the IMF International Financial Statistics database. We find that the inflation series exhibit noticeable time variation of the AR parameter across the time period considered. We discuss how this relates to the literature on the persistence properties of inflation. On the other hand, we find that the real exchange rate series have nearly constant AR values across time that are equal to, or close to, one. Hence, the TVP methods are capable of producing constant AR values when it is appropriate to do so. We discuss how these results relate to the literature on the persistence of exchange rates. In Section E of the Supplemental Material, we also report results for interest rates for several countries and results for a number of US macroeconomic time series using the Federal Reserve Economic Database (FRED).

The results of the paper apply to a TVP-AR(1) model. For some time series, a TVP-AR(p) model with $p > 1$ may be more appropriate than a TVP-AR(1) model. The methods introduced in the paper can be extended to a TVP-AR(p) model, see Section D of the Supplemental Material for details.

Relative to the literature, the contribution of this paper is to develop methods for a deterministically time-varying parameter (TVP) AR(1) model that allows time-varying stationarity in some time periods and time-varying nonstationarity in other periods. The resulting model is much more flexible than a constant parameter model. No paper in the existing literature does this. The methods we employ are quite similar to those employed in constant parameter AR(1) models that impose stationarity or nonstationarity across the whole time period, such as those referenced above. In particular, we invert tests of null hypotheses concerning the AR coefficient (at a particular point in time) and utilize the nonstandard asymptotic distributions of the t-statistics for such null hypotheses to obtain critical values. Our results differ from those obtained for constant parameter AR(1) models in that (i) we

consider a local neighborhood of the time point of interest τ, indexed by a bandwidth parameter h, and within that time period the AR coefficient changes with time, which causes biases that have to be accounted for, (ii) the initial condition of the local neighborhood depends on the choice of the local neighborhood and on the past behavior of the TVP-AR(1) process, which may differ from its current behavior, which needs to be accounted for, and (iii) we use the data to select a suitable local neighborhood using a forecast-error criterion function. None of these features arise in a constant parameter model.

The literature contains numerous papers that consider time-varying parameter AR models. Some of these papers consider deterministic TVP's, as in this paper, but they do not allow for stationarity in some time periods and nonstationarity in others. References for stationary (i.e., short-range dependent) TVP AR models include Subba Rao (1970), Grenier (1983), Dahlhaus (1996, 1997), Dahlhaus and Giraitis (1998), Dahlhaus, Neumann, and Sachs (1999), Moulines, Priouret, and Roueff (2005), Xu and Phillips (2008), Ding, Qiu, and Chen (2017), van Delft and Eichler (2018), and Karmakar, Richter, and Wu (2022). References for nonstationary (i.e., long-range dependent) TVP AR models include Bykhovskaya and Phillips (2018, 2020), which focus on tests of a unit root null hypothesis against functional local-to-unity alternatives, as opposed to CI's for an AR parameter, which is the focus of this paper. No papers in the literature consider CI's for an AR parameter in nonstationary TVP AR models.

The literature also includes papers on random coefficient (RC) AR models and functional coefficient (FC) AR models (in which the AR coefficient depends on observable variables). References for stationary RC AR models includes Nicholls and Quinn (1980), Quinn and Nicholls (1981), Doan, Litterman, and Sims (1984), and Cogley and Sargent (2005). References for papers on RC AR models with random coefficients that follow a nonstationary process include Föllmer and Schweizer (1993), Giraitis, Kapetanios, and Yates (2014, 2018), and Tao, Phillips, and Yu (2019). A reference for stationary FC AR models is Cai, Fan, and Yao (2000). References for nonstationary FC AR models includes Juhl (2005), Lieberman (2012), and Lieberman and Phillips (2014, 2017, 2018).

This paper is organized as follows. Section 2 introduces the TVP-AR(1) model. Section 3 introduces the CI and MUE for the AR parameter at time τ. Section 4 introduces the data-dependent method for choosing the bandwidth parameter h based on a forecast-error criterion. Section 5 presents the Monte Carlo simulation results. Section 6 presents the empirical results. Section 7 shows that the CI has correct uniform asymptotic coverage probability, the MUE is asymptotically median unbiased, and the data-dependent method for choosing the bandwidth parameter h has some some desirable asymptotic properties. It also provides the asymptotic behavior of the local least squares estimator and t-statistic under

a variety of drifting sequences of distributions, which are used in the proof of the uniform asymptotic coverage probability results and the asymptotic median unbiasedness results. The Supplemental Material includes the critical values of the limiting distribution of our t-statistic in Section A, the proofs of the results of the paper in Section B, additional simulation results in Section C, a description of how to extend the methods to TVP-AR(p) models for $p > 1$ in Section D, and information about the empirical applications and additional empirical results in Section E.

All limits in this paper are as $n \to \infty$. For notational simplicity, but with some abuse of notation, we let $\cdot/nh$ denote $\cdot/(nh)$ throughout the paper.

2 Model

The TVP-AR(1) model we consider is

$$\begin{aligned} Y_t &= \mu_t + Y_t^* \text{ and} \\ Y_t^* &= \rho_t Y_{t-1}^* + \sigma_t U_t, \text{ for } t = 1, ..., n, \end{aligned} \tag{2.1}$$

where $\rho_t \in [-1 + \varepsilon_1, 1]$ for some $0 < \varepsilon_1 < 2$. The autoregressive parameter ρ_t is allowed to vary with time t. A key feature of the model is that it allows for stationary, unit root, or local-to-unity behavior at different points in time. The errors $\{U_t : t = 0, 1, ..., n\}$ are a stationary martingale difference sequence under F with $E_F(U_t | \mathscr{G}_{t-1}) = 0$ a.s., $E_F(U_t^2 | \mathscr{G}_{t-1}) = 1$ a.s., and $E_F(U_t^4 | \mathscr{G}_{t-1}) < M$ a.s. for some $M \in (0, \infty)$, where $\mathscr{G}_t$ is some non-decreasing sequence of σ-fields for which $\sigma(U_0, ..., U_t, Y_0^*) \subseteq \mathscr{G}_t$ for $t = 1, ..., n$.

We assume ρ_t, μ_t, and σ_t^2 satisfy

$$\rho_t := \rho(t/n), \ \mu_t := \mu(t/n), \text{ and } \sigma_t^2 := \sigma^2(t/n), \tag{2.2}$$

respectively, where $\rho(\cdot)$ is a twice continuously differentiable function on $[0, 1]$ and $\mu(\cdot)$ and $\sigma^2(\cdot)$ are Lipschitz functions on $[0, 1]$. Given (2.2), Y_t, Y_t^*, ρ_t, μ_t, and σ_t depend implicitly on n.

Let $\tau \in (0, 1)$. We consider estimation and inference concerning

$$\rho(\tau), \tag{2.3}$$

which is the value of the autoregressive function $\rho(\cdot)$ at the τ fraction of the way through the sample.

For ease of reading, the definition of the parameter space of functions $\rho(\cdot)$, $\mu(\cdot)$, and

$\sigma^2(\cdot)$ that is considered, which includes some structure on the $\rho(\cdot)$ and $\mu(\cdot)$ functions, is given in Section 7.1 below.

3 Confidence Interval for the Autoregressive Parameter $\rho(\tau)$

3.1 Local Least Squares Estimator of $\rho(\tau)$

We employ a LS estimator of $\rho(\tau)$ based on the time periods $t = T_1, ..., T_2$, where

$$T_1 = \lfloor n\tau \rfloor - \lfloor nh/2 \rfloor \text{ and } T_2 = \lfloor n\tau \rfloor + \lfloor nh/2 \rfloor, \tag{3.1}$$

for a bandwidth parameter h. The total number of periods in $[T_1, T_2]$ is within one of nh. For this range of time periods, the initial condition time period is $T_0 := T_1 - 1$.

For the asymptotic results given below the bandwidth h satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$. In Section 4 below, we introduce a data-dependent bandwidth that is smaller or larger depending on how wiggly or flat the true $\rho(\cdot)$ and $\mu(\cdot)$ functions are.

Define

$$\overline{Y}_{nh} := \frac{1}{nh} \sum_{t=T_1}^{T_2} Y_t \text{ and } \overline{Y}_{nh,-1} := \frac{1}{nh} \sum_{t=T_1}^{T_2} Y_{t-1}. \tag{3.2}$$

To estimate $\rho(\tau)$, we regress Y_t on a constant and Y_{t-1}. The resulting local LS estimator $\widehat{\rho}_{n\tau}$ is

$$\widehat{\rho}_{n\tau} = \frac{\sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)\left(Y_t - \overline{Y}_{nh}\right)}{\sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)^2}. \tag{3.3}$$

3.2 Confidence Interval for $\rho(\tau)$

The CI for $\rho(\tau)$ that we consider is obtained by inverting tests of null hypotheses of the form $H_0 : \rho(\tau) = \rho_0$ for different values $\rho_0 \in [-1 + \varepsilon_1, 1]$.

The estimator of the time-varying variance $\sigma^2(\cdot)$ at $t/n = \tau$ is defined to be

$$\widehat{\sigma}^2_{n\tau} := (nh)^{-1} \sum_{t=T_1}^{T_2} \left[Y_t - \overline{Y}_{nh} - \widehat{\rho}_{n\tau}\left(Y_{t-1} - \overline{Y}_{nh,-1}\right)\right]^2. \tag{3.4}$$

For arbitrary $\rho_0 \in (-1, 1]$, the t-statistic that is used to construct the CI for $\rho(\tau)$ is

$$T_n(\rho_0) := \frac{(nh)^{1/2}(\widehat{\rho}_{n\tau} - \rho_0)}{\widehat{s}_{n\tau}}, \text{ where } \widehat{s}^2_{n\tau} := \widehat{\sigma}^2_{n\tau} / (nh)^{-1} \sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)^2. \tag{3.5}$$

Let $B(\cdot)$ denote a standard Brownian motion on $[0,1]$. Let Z_1 be a standard normal random variable that is independent of $B(\cdot)$. Define

$$\begin{aligned} I_\psi(s) &:= \int_0^s \exp\{-(s-r)\psi\}\, dB(r), \\ I_\psi^*(s) &:= \begin{cases} I_\psi(s) + \dfrac{1}{\sqrt{2\psi}} \exp(-\psi s) Z_1 & \text{for } \psi > 0 \\ B(s) & \text{for } \psi = 0, \text{ and} \end{cases} \\ I_{D,\psi}^*(s) &:= I_\psi^*(s) - \int_0^1 I_\psi^*(r)\, dr. \end{aligned} \tag{3.6}$$

The stochastic process $I_\psi(s)$ is an Ornstein-Uhlenbeck process on $[0,1]$ with parameter ψ.

Below we consider sequences of functions $\{\rho_n(\cdot), \mu_n(\cdot), \sigma_n(\cdot)\}_{n\geq 1}$ and null hypotheses $H_0 : \rho_n(\tau) = \rho_{0,n}$, where the null hypotheses values $\rho_{0,n}$ depend on n for $n \geq 1$. For suitable sequences $\{\rho_{0,n}\}_{n\geq 1}$, we show that under H_0,

$$T_n(\rho_{0,n}) \to_d J_\psi \text{ for } \psi \in [0,\infty], \tag{3.7}$$

where $T_n(\rho_{0,n})$ is defined with $\rho_{0,n}$ in place of ρ_0 in (3.5), ψ depends on the sequence $\{\rho_{0,n}\}_{n\geq 1}$, and J_ψ is defined as follows. For $\psi = \infty$, which corresponds to a "stationary" sequence $\{\rho_{0,n}\}_{n\geq 1}$, J_ψ has a $N(0,1)$ distribution. For $\psi \in [0,\infty)$, which corresponds to a local-to-unity or unit root sequence $\{\rho_{0,n}\}_{n\geq 1}$,

$$J_\psi := \left(\int_0^1 I_{D,\psi}^{*2}(s)\, ds\right)^{-1/2} \int_0^1 I_{D,\psi}^*(s)\, dB(s). \tag{3.8}$$

For $\alpha \in (0,1)$, let $c_\psi(\alpha)$ denote the α quantile of the distribution of J_ψ. For given α, we compute $c_\psi(\alpha)$, the α-quantile of J_ψ in (3.8), by simulating the asymptotic distribution J_ψ. To do so, $B = 300,000$ independent constant coefficient AR(1) sequences are generated with innovations $U_t \sim_{iid} N(0,1)$, stationary start-up, $n = 25,000$, and $\rho = \exp(-\psi/n)$. For each sequence, the test statistic $T_n(\rho)$ defined in (3.5) is calculated. Then the simulated estimate of $c_\psi(\alpha)$ is the α-quantile of the empirical distribution of the $B = 300,000$ realizations of the test statistic $T_n(\rho)$.

The nominal $1-\alpha$ equal-tailed two-sided CI for $\rho(\tau)$ is

$$\begin{aligned} CI_{n,\tau} &:= \left\{\rho_0 \in [-1+\varepsilon_1, 1] : c_{\psi_{nh,\rho_0}}(\alpha/2) \leq T_n(\rho_0) \leq c_{\psi_{nh,\rho_0}}(1-\alpha/2)\right\}, \text{ where} \\ \psi_{nh,\rho_0} &:= -nh \ln(\rho_0) \text{ for } \rho_0 > 0 \text{ and } \psi_{nh,\rho_0} := \infty \text{ for } \rho_0 \leq 0. \end{aligned} \tag{3.9}$$

The CI $CI_{n,\tau}$ can be computed by taking a fine grid of values $\rho_0 \in [-1+\varepsilon_1, 1]$ and

comparing $T_n(\rho_0)$ to $c_{\psi_{nh,\rho_0}}(\alpha/2)$ and $c_{\psi_{nh,\rho_0}}(1-\alpha/2)$. Table SM.1 provides the critical values $c_{\psi_{nh,\rho_0}}(\alpha/2)$ and $c_{\psi_{nh,\rho_0}}(1-\alpha/2)$ for $\alpha = .05$, and .1. Given these critical values, computation of equal-tailed two-sided 90% and 95% CI's for $\rho(\tau)$ is fast.

The correct asymptotic size and asymptotic similarity of the CI $CI_{n,\tau}$ are established in Theorem 7.1 in Section 7 below.

3.3 Median-Unbiased Interval Estimator of $\boldsymbol{\rho(\tau)}$

By definition, an estimator $\widehat{\theta}_n$ of a parameter θ is median unbiased if $P\left(\widehat{\theta}_n \geq \theta\right) \geq 1/2$ and $P\left(\widehat{\theta}_n \leq \theta\right) \geq 1/2$. Here we introduce a median-unbiased interval estimator of $\rho(\tau)$ that satisfies an analogous condition. Also, with probability close to one, the estimator is a single point.[1] Let $CI^{up}_{n,\tau}(.5)$ and $CI^{low}_{n,\tau}(.5)$ denote level .5 one-sided upper-bound and lower-bound CIs for $\rho(\tau)$, respectively. By definition,

$$\begin{aligned} CI^{up}_{n,\tau}(.5) &:= \left\{\rho_0 \in [-1+\varepsilon_1, 1] : c_{\psi_{nh,\rho_0}}(.5) \leq T_n(\rho_0)\right\} \text{ and} \\ CI^{low}_{n,\tau}(.5) &:= \left\{\rho_0 \in [-1+\varepsilon_1, 1] : T_n(\rho_0) \leq c_{\psi_{nh,\rho_0}}(.5)\right\} \text{ for } \psi_{nh,\rho_0} \text{ as in (3.9).} \end{aligned} \tag{3.10}$$

The median-unbiased interval estimator $\widetilde{\rho}_{n\tau}$ of $\rho(\tau)$ is defined by

$$\begin{aligned} \widetilde{\rho}_{n\tau} &= [\widetilde{\rho}_{n\tau,low}, \widetilde{\rho}_{n\tau,up}], \text{ where} \\ \widetilde{\rho}_{n\tau,up} &= \max\left\{\rho_0 : \rho_0 \in CI^{up}_{n,\tau}(.5)\right\} \text{ and} \\ \widetilde{\rho}_{n\tau,low} &= \min\left\{\rho_0 : \rho_0 \in CI^{low}_{n,\tau}(.5)\right\}. \end{aligned} \tag{3.11}$$

By construction, $\widetilde{\rho}_{n\tau,low} \leq \widetilde{\rho}_{n\tau,up}$.[2] In addition, $\widetilde{\rho}_{n\tau}$ is a singleton whenever the set $\left\{\rho_0 \in [-1+\varepsilon_1, 1] : T_n(\rho_0) = c_{\psi_{nh,\rho_0}}(.5)\right\}$ contains a single point, in which case $\widetilde{\rho}_{n\tau}$ equals this point. Simulations show that $\widetilde{\rho}_{n\tau}$ is a singleton with probability close to one and a very short interval when it is not a singleton. Table SM.1 provides the critical values $c_{\psi_{nh,\rho_0}}(.5)$ for a wide range of ψ. Given these critical values, computation of $\widetilde{\rho}_{n\tau}$ is fast.

The estimator $\widetilde{\rho}_{n\tau}$ has the following median-unbiasedness property

$$\begin{aligned} &\liminf_{n\to\infty} P\left(\widetilde{\rho}_{n\tau,up} \geq \rho(\tau)\right) \geq 1/2 \text{ and} \\ &\liminf_{n\to\infty} P\left(\widetilde{\rho}_{n\tau,low} \leq \rho(\tau)\right) \geq 1/2. \end{aligned} \tag{3.12}$$

[1] If the .5 quantile of the asymptotic null distribution of the t-statistic was strictly decreasing in ρ_0, or equivalently, $c_\psi(.5)$ was strictly increasing in ψ, then the proposed interval estimator would be a point estimator with probability one. Because this condition fails to hold exactly, but almost holds, there is a very small probability that the estimator is a short interval, rather than a point.

[2] This holds because $\widetilde{\rho}_{n\tau,up} \geq \sup\left\{\rho_0 \in [-1+\varepsilon_1, 1] : c_{\psi_{nh,\rho_0}}(.5) = T_n(\rho_0)\right\}$ and $\widetilde{\rho}_{n\tau,low}$ is less than or equal to the infimum of the values in the same set.

Furthermore, as shown in Corollary 7.1 in Section 7 below, these asymptotic properties hold in a uniform sense over the parameter space. This property is important for ensuring that the asymptotic properties of $\widetilde{\rho}_{n\tau}$ reflect its finite-sample properties.

4 Data-dependent Bandwidth Parameter h

The CI for $\rho(\tau)$ proposed in Section 3.2 requires a tuning parameter h that determines the interval of observations used to construct the CI. This section introduces a data-dependent method of selecting h that minimizes a forecast-error criterion. We present its asymptotic properties in Section 7.6 under somewhat high-level conditions.

The forecast-error criterion is the average over $t = 1, ..., n$ of the squared errors from forecasting Y_t using the predicted value $\widehat{Y}_t$. Specifically, for $t > \lfloor nh \rfloor$, let $(\widehat{\mu}_{t-1}(h), \widehat{\rho}_{t-1}(h))$ denote the LS estimator of (μ_t, ρ_t) from the regression of Y_s on a constant and Y_{s-1} using the nh observations $s = t - \lfloor nh \rfloor, ..., t-1$. For $t \leq \lfloor nh \rfloor$, since there are fewer than $\lfloor nh \rfloor$ observations, (μ_t, ρ_t) is estimated using the LS estimator $(\widehat{\mu}_{t-1}(h), \widehat{\rho}_{t-1}(h))$ based on the $\lfloor nh \rfloor$ observations $s = 1, ..., t-1, t+1, ..., \lfloor nh \rfloor + 1$. The observation Y_t is predicted by $\widehat{\mu}_{t-1}(h) + Y_{t-1}\widehat{\rho}_{t-1}(h)$ for $t = 1, ..., n$. The forecast-error criterion is

$$FE_n(h) := n^{-1} \sum_{t=1}^{n} (Y_t - \widehat{\mu}_{t-1}(h) - Y_{t-1}\widehat{\rho}_{t-1}(h))^2. \tag{4.1}$$

We assume that h is chosen from a finite set $\mathcal{H}_n$, which depends on n and whose cardinality may depend on n. By definition, the data-dependent choice of h, $\widehat{h}$, minimizes $FE_n(h)$ over $\mathcal{H}_n$: [3]

$$\widehat{h} := \arg\min_{\mathcal{H}_n} FE_n(h). \tag{4.2}$$

Remark 4.1. The criterion $FE_n(h)$ has the desirable property of being an average of unbiased risk estimators for $t > \lfloor nh \rfloor$. This holds because

$$\begin{aligned}
&E(Y_t - \widehat{\mu}_{t-1}(h) - Y_{t-1}\widehat{\rho}_{t-1}(h))^2 \\
&= E(U_t - (\widehat{\mu}_{t-1}(h) - \mu_t) + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t)))^2 \\
&= EU_t^2 - 2EU_t(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t)) + E(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t))^2 \\
&= EU_t^2 + E(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t))^2,
\end{aligned} \tag{4.3}$$

where the third equality holds for $t > \lfloor nh \rfloor$ since $(\widehat{\mu}_{t-1}(h), \widehat{\rho}_{t-1}(h))$ is a function of the

[3] If the argmin is not unique, for specificity $\widehat{h}$, is taken to be the largest argmin. But, non-uniqueness occurs with probability zero provided the innovations have a continuous component.

innovations indexed by $s \leq t-1$ and $E(U_t|U_{t-1},...,U_1,Y_0^*) = 0$. For the relatively small number of initial observations with $t \leq \lfloor nh \rfloor$, the third equality does not hold because $\widehat{\rho}_{t-1}(h)$ depends on observations s with $s \geq t$.

In terms of computation, the criterion $FE_n(h)$ has the advantage that the same $\widehat{h}$ is employed for multiple $\lfloor nh \rfloor$ values of interest. It also avoids introducing additional tuning parameters, such as the length of a forecasting period.

Remark 4.2. In Section 7.6, we show that the data-dependent $\widehat{h}$ value achieves the asymptotically optimal trade-off between bias and variance obtained by the infeasible $\widehat{h}_{opt}$ which minimizes the "empirical loss" as a function of h defined in (7.22) below. In consequence, undersmoothing $\widehat{h}$ should yield a value of h for which the bias is dominated by the variance asymptotically, the CI's defined above have correct asymptotic size, and the median-unbiased interval estimator is asymptotically median unbiased. We employ a relatively sophisticated undersmoothing method. The objective of undersmoothing (i.e., making $\widehat{h}_{us}$ smaller than $\widehat{h}$) is to reduce the bias. The cost of undersmoothing is an increase in variance. We want to undersmooth more when the cost of doing so in terms of increasing the variance is relatively small, and we want to undersmooth less when the cost is larger. This means that we need to take account of the shape of the $FE_n(h)$ function. When it is flatter at its minimum, we want to undersmooth more. This leads to the following definition:

$$\widehat{h}_{us0} = \min\left\{h \in \mathcal{H}_n : FE_n(h) \leq Q_{c_1}(FE_n)\right\}, \tag{4.4}$$

where $Q_{c_1}(FE_n)$ is the c_1 quantile of FE_n among $h \in \mathcal{H}_n$ and we take $c_1 = .2$ in the simulations and applications. This definition does not guarantee that $\widehat{h}_{us0}$ is of a smaller order than $\widehat{h}$, which is necessary to ensure proper asymptotics. In consequence, we modify the definition as follows:

$$\widehat{h}_{us} = \min\left\{\widehat{h}_{us0}, \widehat{h}_{us1}\right\}, \tag{4.5}$$

where $\widehat{h}_{us1} = c_2 n^{-a}\widehat{h}$ for $c_2, a > 0$. In the simulations and applications, we use $c_2 = 1.5$ and $a = 1/10$.

5 Monte Carlo Simulations

In this section, we analyze the finite-sample performance of the methods introduced above using Monte Carlo simulations. First, we describe the data generating processes (DGP's) considered and how the CI and MUE are implemented. Then, we report the results. We find that the proposed CI has reasonably good coverage probabilities and short average lengths

for most of the DGP's considered. We also find that the MUE performs quite well both when the true DGP of ρ_t is time-varying and flat.

5.1 Simulation Setup and Methodology

We consider 21 DGP's for ρ_t. The graphs of the ρ_t functions are given in Figures 1–3 and Figures SM.1–SM.4, which appear in the Supplemental Material. There are five categories of ρ_t functions: sinusoidal, linear, flat linear, flat, and kinked linear. For the sinusoidal functions, we consider sinusoidal functions for ρ_t with 6 different ranges and shapes. For example, "sin 1.00-0.90-1.00" corresponds to the sinusoidal function where ρ_t first achieves its maximum of 1.00 at $t/n = .20$, then drops to 0.90 at $t/n = .60$ and finally increases to 1.00 at $t/n = 1$, with a frequency of 2.5π. For linear functions, we consider 4 different linear DGP's for ρ_t corresponding to different values of the slopes and intercepts. For flat linear functions, the first half of ρ_t is flat and the second half is linear, and 4 different cases are considered. For example, "flat-lin 0.90-0.99" means

$$\rho_t = \rho(t/n) = \begin{cases} .90 & \text{for } t/n \in [0, 1/2]\,, \\ .18t/n + .81 & \text{for } t/n \in (1/2, 1]. \end{cases}$$

For the flat functions, we consider 3 values .75, .90, and .99. In the Supplemental Material, we also report results for kinked linear functions, where the first and second halves of ρ_t are linear, but with slopes of different signs. For example, "kinked 1.00-0.80-1.00" is a DGP for ρ_t where the first half of ρ_t decreases linearly from 1 to .80 while the second half increases linearly from .80 to 1. We consider 4 different cases in this category.

For each of the 21 DGP's for ρ_t, we consider both constant and time-varying DGP's for μ_t and σ_t in (2.1) except for that of "sin 1.00-0.90-1.00" with constant μ_t and σ_t to present the legend for all the figures. Define $\mu_t^* = (1 - \rho_t)\mu_t$, which allows one to rewrite (2.1) as

$$Y_t = \mu_t^* + \rho_t Y_{t-1} + \sigma_t U_t, \text{ for } t = 1, 2, ..., n. \tag{5.1}$$

When μ_t and σ_t are constant, we take $\mu_t = 0$ and $\sigma_t = 1$, which implies $\mu_t^* = 0$. When μ_t and σ_t are time-varying, we generate μ_t such that μ_t^* increases linearly from -.1 to .1, and σ_t increases linearly from .95 to 1.05 as t/n goes from 0 to 1. In consequence, there are a total of 41 $(= 21 \times 2 - 1)$ DGP's for $(\rho_t, \mu_t^*, \sigma_t)$. We take U_t to be i.i.d. $N(0, 1)$, initialize Y_0 by drawing it from a $N(0, \overline{\sigma}_n^2)$ distribution, where $\overline{\sigma}_n = 1/(1 - \overline{\rho}_n^2)$ and $\overline{\rho}_n = n^{-1} \sum_{t=1}^n \rho_t$, and form the Y_t sequence as in (5.1).

We compute a nominal .95 two-sided CI and MUE for $\rho(\tau)$ for each of 5 time points

of interest, indexed by $\tau \in \{.2, .4, .6, .8, 1\}$. The CI and MUE are implemented as follows. First, we compute $\widehat{h}$ based on the method described in Section 4 and take the undersmoothed $\widehat{h}_{us}$ to be as defined in (4.5) with $c_1 = .2$, $c_2 = 1.5$, and $a = 1/10$. In this step, the set of nh values based on $\mathcal{H}_n$ is $\{140, 155, ..., 500, 650, ..., 1500\}$. Next, for each candidate value $\rho_0 \in \{-1, -.995, ..., .945, .95, .951, .952, ..., 1\}$, we calculate the t-statistics $T_n(\rho_0)$ defined in (3.5) from the regression of Y_t on a constant and Y_{t-1} using the $n\widehat{h}_{us}$ observations centered around $n\tau$. When $n\tau$ is close to the boundary, we use $n\widehat{h}_{us}/2$ observations to the side that has abundant data, and as many observations as available to the other side until the boundary is hit. In consequence, a different number of observations for different τ may be used even though the $n\widehat{h}_{us}$ value is the same. Comparing the $T_n(\rho_0)$ values with corresponding critical values $c_{\psi_{nh,\rho_0}}(\alpha/2)$ and $c_{\psi_{nh,\rho_0}}(1-\alpha/2)$ (for ψ_{nh,ρ_0} defined in (3.9)) at each candidate ρ_0 gives the nominal $1-\alpha$ equal-tailed two-sided CI for $\rho(\tau)$ as defined in (3.9). We compute coverage probabilities and average lengths of the CI's across all simulations for each point of interest for each DGP.

To calculate the MUE for $\rho(\tau)$, we follow the procedure described in Section 3.3 and use $\widetilde{\rho}_{n\tau}$ as the MUE of $\rho(\tau)$ when $\widetilde{\rho}_{n\tau}$ is a singleton. When $\widetilde{\rho}_{n\tau,up} \neq \widetilde{\rho}_{n\tau,low}$, we take $\widetilde{\rho}_{n\tau,up}$ as the MUE for $\rho(\tau)$ based on two considerations. First, using the upper bound $\widetilde{\rho}_{n\tau,up}$ has the theoretical property that it is not median-biased towards zero for positive $\rho(\tau)$ values because, by Corollary 7.1, $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda_n} P_\lambda(\widetilde{\rho}_{n\tau,up} \geq \rho(\tau)) \geq 1/2$. Second, in the simulations we find that $\widetilde{\rho}_{n\tau}$ is much more likely to be an interval when the true $\rho(\tau)$ is close to one. We report the absolute median biases of the MUE and the range of the MUE values across the simulations for each time point of interest for each DGP in the figures.

For each of the 41 DGP's for Y_t, we run $M = 5,000$ simulations. The length of the Y_t sequence is $n = 1,500$. Given the 41 DGP's and 5 time points of interest for each one, a total of 205 cases are considered.

5.2 Discussion of Results

First, Table 1 summarizes the CP results by reporting the number and percentage of times that the CI CP's lie in different ranges across the 205 cases considered. The nominal CP is .95. Table 1 shows that only in a small fraction of cases, 2.0%, is the CP quite low, i.e., in $[.875, .90)$. In the majority of cases, 92.2%, the CP is .925 or larger, which shows that the proposed method has reasonably good CP regardless of whether the underlying DGP is curvy or flat and whether the point of interest is in the middle or at the boundary of the time period.

Second, we consider summary statistics for the absolute values of the median biases

Table 1: Distribution of CP's Across 205 Cases

CP Range	[.875,.90)	[.90,.925)	[.925,.94)	[.94,.96)	[.96,.965]
Number	4	12	34	148	7
Percent	2.0%	5.9%	16.6%	72.2%	3.4%

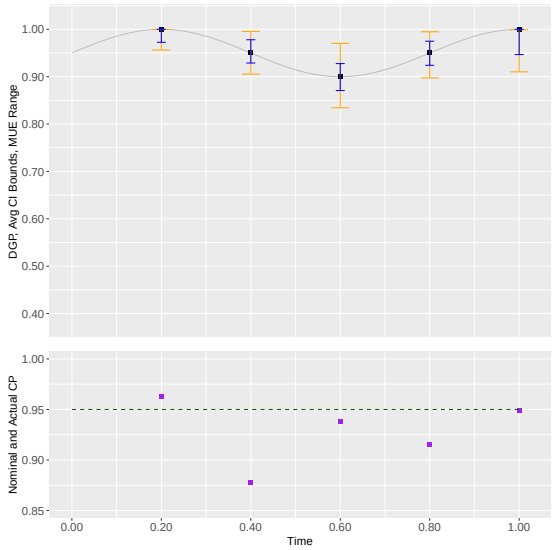
(AMB) of the MUE, i.e., $|\text{median}\,(\widetilde{\rho}_{n\tau,up} - \rho\,(\tau))|$, across the 205 cases considered. The mean and median of the absolute median biases across all cases are .004 and .003, respectively. The range is [.000, .023]. Thus, the magnitudes of the absolute median biases of the MUE are generally small.

Third, we consider the values of $n\widehat{h}_{us}$ selected by the data-dependent method. The mean, median, and range are 212, 207, and [56, 433], respectively. The $n\widehat{h}_{us}$ values vary depending on the shape of the ρ function considered. The flatter is the true ρ function, the larger are the $n\widehat{h}_{us}$ values.

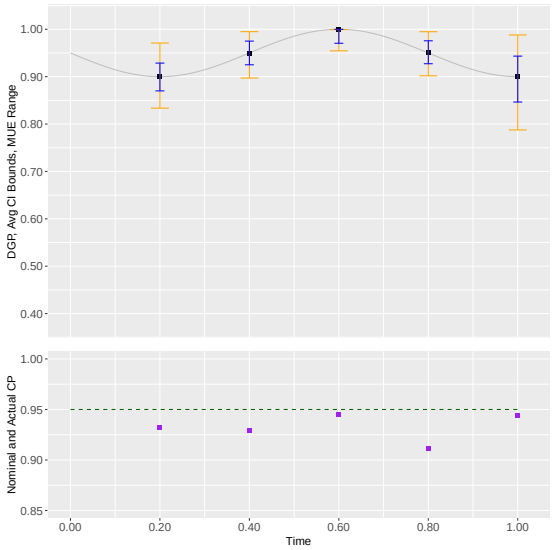
Now, we describe the simulation results for the different ρ functions considered. Figure 1 provides simulation results for five sinusoidal ρ functions with constant μ and σ^2 functions. The amplitudes of the sinusoidal functions increase as one moves down the rows. In the first column of graphs, the ρ function's maximum value occurs at $\tau = 1$, which corresponds to $t = n\tau = n$. In the second column of graphs, the ρ function's minimum value occurs at $\tau = 1$. Each graph consists of an "upper" and a "lower" graph. The "upper" graph shows the true ρ function, the average CI lower and upper bounds at five points of interest $\tau = .2$, .4, .6, .8, 1.0, where $t = n\tau$, and twice the average lower and upper absolute deviations of the MUE (i.e., $2E|\widehat{\rho}_t - \rho_t|\mathbb{1}(\widehat{\rho}_t < \rho_t)$ and $2E|\widehat{\rho}_t - \rho_t|\mathbb{1}(\widehat{\rho}_t > \rho_t)$, respectively, whose sum is somewhat analogous to two standard deviations) at each of these points of interest. The difference between the average upper and lower CI bounds gives the average lengths (AL's) of the CI's. The "lower" graphs report the CI coverage probabilities (CP's) at the five points of interest. The CI's all have nominal CP .95.

The sinusoidal graphs in Figure 1 show the following: (i) All but three of the CP's range from .925 to .965, which are close to the nominal CP of .95. (ii) The AL's are shortest for $\rho(\tau)$ values closest to one, and longest for $\rho(\tau)$ values farthest from one. This reflects the nh-consistency of the LS estimator in the (temporally local) unit root and local-to-unity scenarios, and the $(nh)^{1/2}$-consistency of the LS estimator in the (temporally local) stationary scenarios. (iii) The AL's are larger for the $\tau = 1$ than the $\tau < 1$ cases because fewer observations are used to construct the CI in the former cases, which reflects the need to reduce the boundary bias in order to obtain a good CP. Combining this fact with point (ii), we find the largest AL's occur when $\tau = 1$

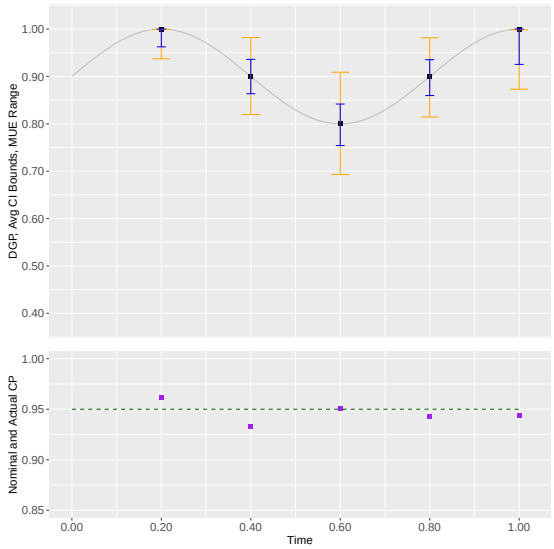
(a) sin 1.00-0.90-1.00, constant μ and σ



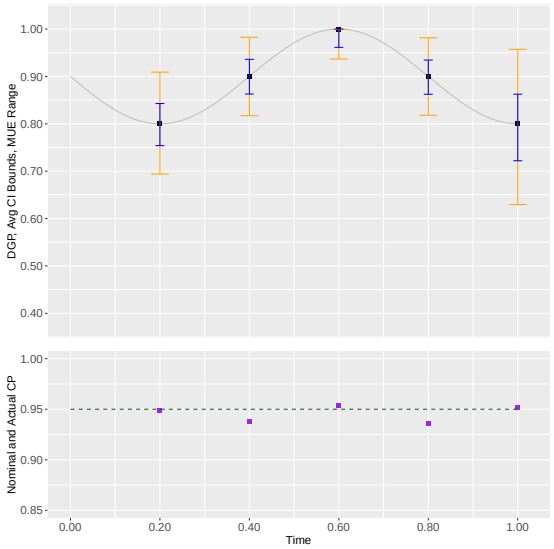
(b) sin 0.90-1.00-0.90, constant μ and σ



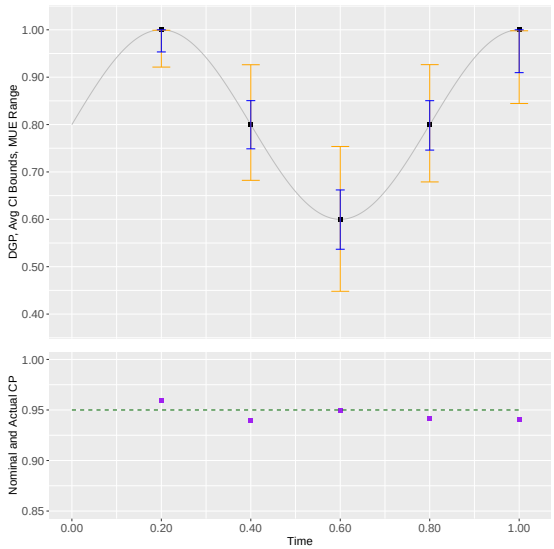
(c) sin 1.00-0.80-1.00, constant μ and σ



(d) sin 0.80-1.00-0.80, constant μ and σ



(e) sin 1.00-0.60-1.00, constant μ and σ



Legend for the Figures

Figure 1: CP's and AL's of CI's for $\rho(\tau)$ and MAD's of the MUE of $\rho(\tau)$

and $\rho(\tau)$ is far from one (which occurs in the two graphs in the second column). (iv) The magnitudes of the MUE average absolute deviations are roughly proportional to the AL's of the CI's. (v) The simulation results for the sinusoidal functions with time-varying μ and σ^2 functions, which are given in Figure SM.1 in the Supplemental Material, are quite similar to those in Figure 1.

Figure 2(a)-(d) provides results for linear ρ functions with constant μ and σ^2 functions. The results are similar to those for the sinusoidal functions, although all of the CP's lie between .924 and .957. Points (ii)-(v) above also apply to the linear functions in Figure 2. Results for time-varying μ and σ^2 are provided in Figure SM.2 in the Supplemental Material, and are similar to those in Figure 2.

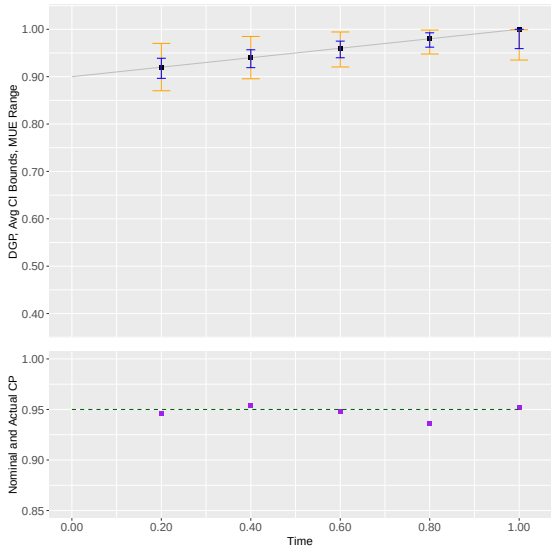
Figures 2(e)-(f) and 3(a)-(b) report results for flat-linear ρ functions with $\rho(t)$ varying between .90 and .99 in Figure 2(e)-(f) and between .80 and .99 in Figure 3(a)-(b). Figures 2(e) and 3(a) report results for constant μ and σ^2; while Figures 2(f) and 3(b) report results for time-varying μ and σ^2. The CP results in Figures 2(e)-(f) and 3(a)-(b) lie between .932 and .956.

Figures SM.2(e)-(f) and SM.3(a)-(b) provide analogous results to those just discussed for flat-linear ρ functions, but for the case where the flat part has value .99, rather than .80 or .90, and the linear part has a negative slope, rather than a positive slope. The CP results for these cases are lower than those for the flat-linear ρ functions in Figures 2 and 3. For example, Figure SM.2(e) shows a CP of .905 at $\tau = .60$. For the other four cases, the CP's are in the range of .917 to .957.

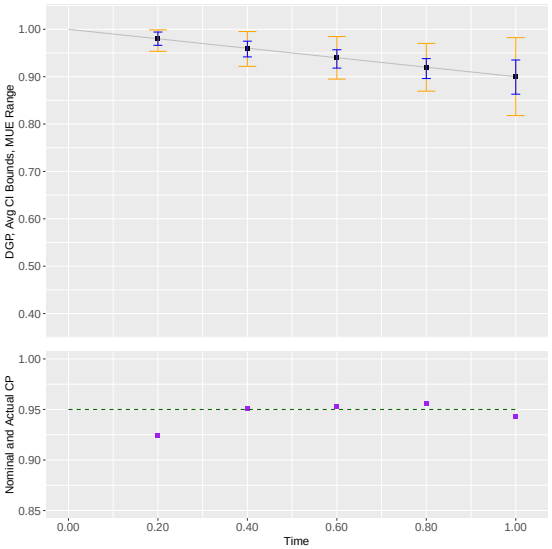
Figure 3(c)-(f) considers flat ρ functions. In Figure 3(c)-(d), $\rho = .99$ and the μ and σ^2 functions are constant and time-varying, respectively. Figure 3(e)-(f) is analogous, but with $\rho = .90$. Figure SM.3(c)-(d) also is analogous, but with $\rho = .75$. The CP's in the flat ρ function cases are all close to .95 and lie between .932 and .960. This occurs because there is no bias accruing due to a time-varying ρ function. However, in the cases of flat ρ, μ, and σ^2 functions, the CI's are not as short as the oracle CI's that rely on knowledge that the ρ, μ, and σ^2 functions are all flat and take $nh = n$. For the case of $\rho = .99$ and flat μ and σ^2 functions, the AL's of the TVP CI and the oracle CI are .022 and .012, respectively, at $\tau = .2$. For $\rho = .90$, they are .085 and .039 at the same τ. For $\rho = .75$, they are .124 and .062. There is a substantial increase in the average lengths in the flat cases using the methods of this paper, in order to ensure good CP's in the near flat cases.

Lastly, Figures SM.3(e)-(f) and SM.4(a)-(f) in the Supplemental Material provide results for kinked ρ functions that either linearly increase until $\tau = .5$ and then linearly decrease, or linearly decrease until $\tau = .5$ and then linearly increase. Both constant and time-varying μ and σ^2 functions are considered. In two cases, the CP's are .904 and .906. In all other cases,

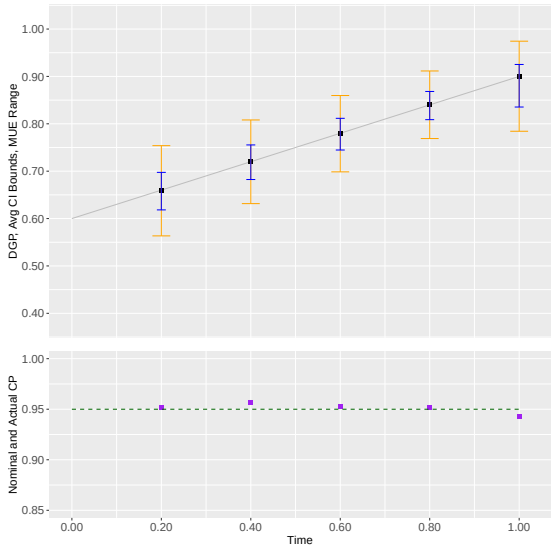
(a) linear 0.90-1.00, constant μ and σ



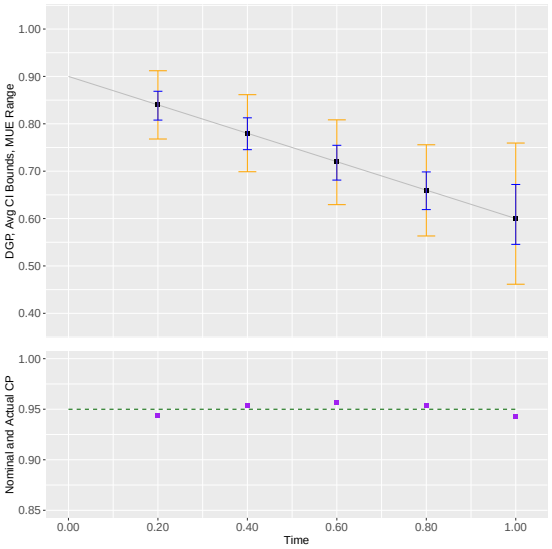
(b) linear 1.00-0.90, constant μ and σ



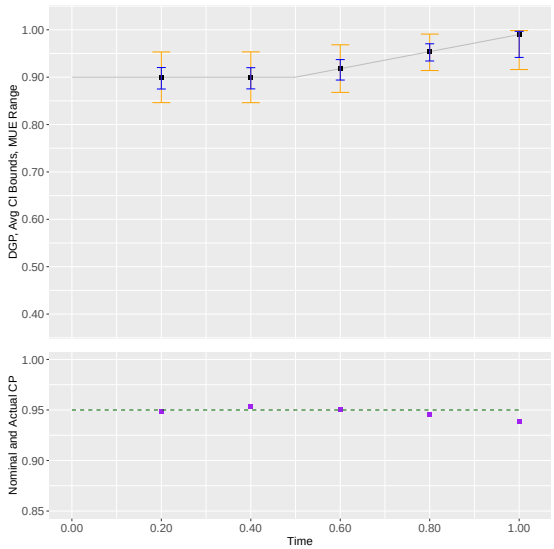
(c) linear 0.60-0.90, constant μ and σ



(d) linear 0.90-0.60, constant μ and σ



(e) flat-lin 0.90-0.99, constant μ and σ



(f) flat-lin 0.90-0.99, time-varying μ and σ

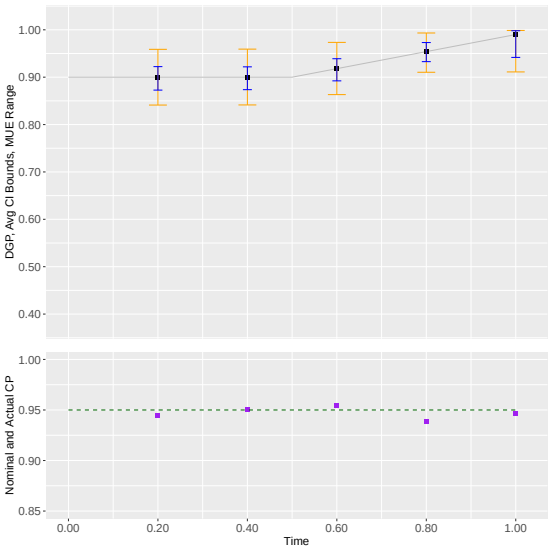
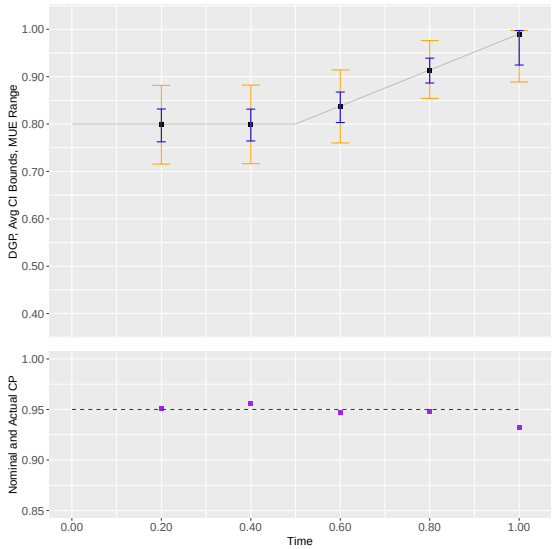
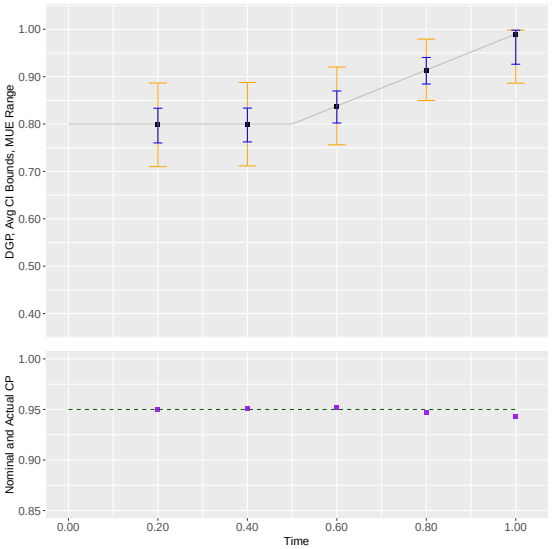


Figure 2: CP's and AL's of CI's for $\rho(\tau)$ and MAD's of the MUE of $\rho(\tau)$

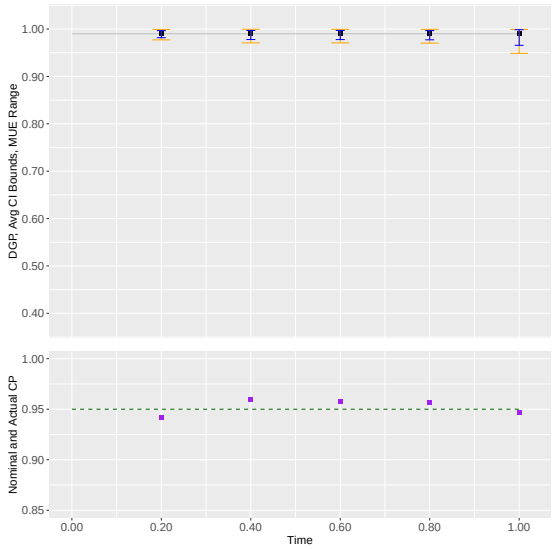
(a) flat-lin 0.80-0.99, constant μ and σ



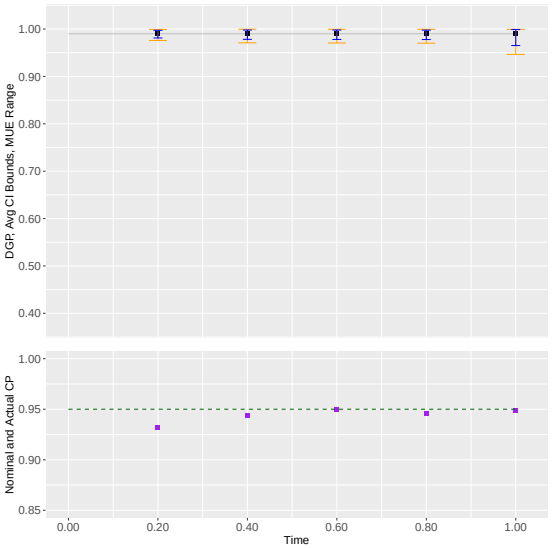
(b) flat-lin 0.80-0.99, time-varying μ and σ



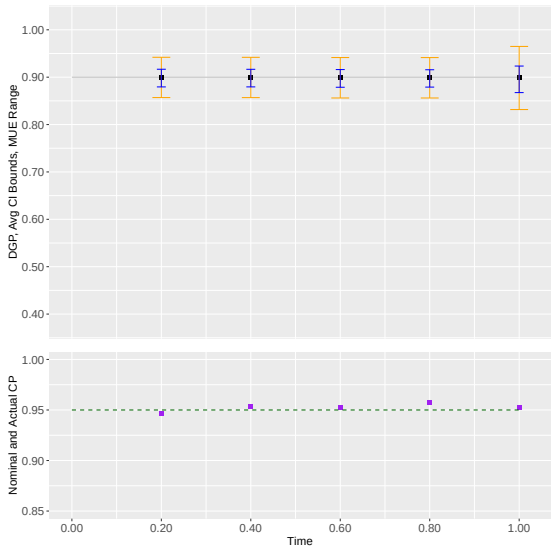
(c) flat 0.99, constant μ and σ



(d) flat 0.99, time-varying μ and σ



(e) flat 0.90, constant μ and σ



(f) flat 0.90, time-varying μ and σ

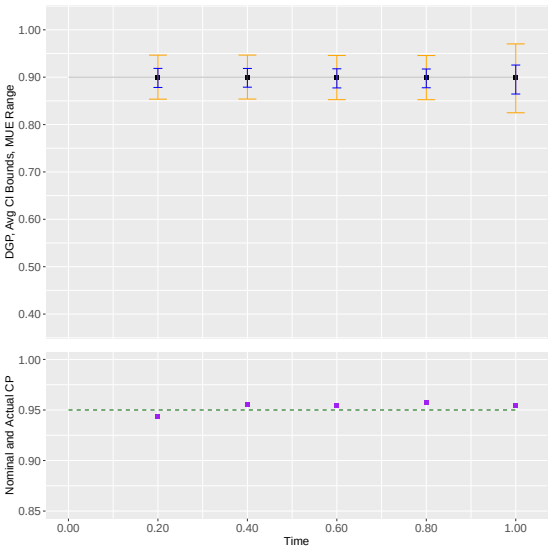


Figure 3: CP's and AL's of CI's for $\rho(\tau)$ and MAD's of the MUE of $\rho(\tau)$

the CP's are in the range of .930 to .953. When the difference between the minimum and maximum ρ value considered is large, viz., .4, the AL's of the CI's are large. This occurs because the data-dependent choice of h is small in order to avoid bias.

Overall, the CP's of the CI's are quite good with only a few being as low as .88 or .89 out of the 205 cases considered. The AL's vary substantially across scenarios depending on how close $\rho(\tau)$ is to one and how much the ρ function varies with time, as is to be expected.

6 Empirical Applications

This section presents applications of the proposed methods to time series of inflation and exchange rates in several countries. The data comes from the IMF (International Financial Statistics (IFS)) database.

In the Supplemental Material, results are provided for some additional countries and for interest rates. The Supplemental Material also provides applications to eight macroeconomic series for the US, using the Federal Reserve Economic Data (FRED).

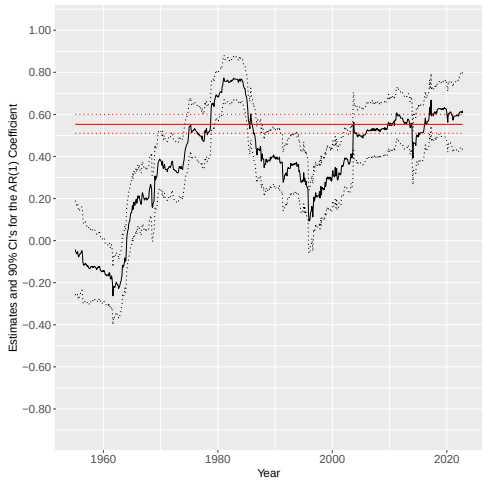
In terms of computation, we set $nh_{\min} = .2n$ and $nh_{\max} = 2n$, where n is the sample size. For nh values between $nh_{\min}$ and nh_{mid}, where $nh_{\text{mid}} \coloneqq .5n$, we use a grid size of $.02n$, while for nh values between nh_{mid} and $nh_{\max}$, we use a grid size of $.05n$ because the range of nh values is wider than between $nh_{\min}$ and nh_{mid}.

6.1 Inflation

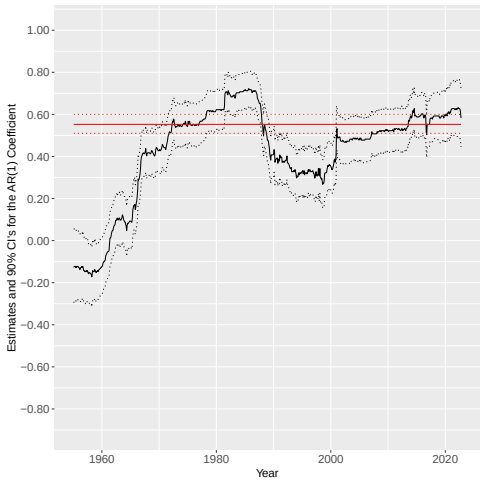
First, we apply our method to the monthly inflation data, defined as the percentage change in CPI over the previous month, see Figure 4. We consider three countries: the US, Canada, and Germany. The data span is Feb 1955 to Oct 2022, which contains $n = 813$ observations for each country. We compute the $n\widehat{h}_{us}$ values based on (4.5). Given these, we compute the MUE's and 90% CI's for $\rho(t)$ at each t and for each country. As a robustness check, we multiply the undersmoothed $n\widehat{h}_{us}$ by 1.5 and present the results in the right column of each figure. The final $n\widehat{h}_{us}$ values used for the computations are listed in the titles of the figures. For all of the inflation series, Ljung-Box tests based on six lags of the residuals from the AR(1) model fail to reject the null hypothesis of no autocorrelation at the 5% significance level, see Table SM.2.

For comparative purposes, we also fit a constant autoregression coefficient AR(1) model to the data and report the estimated $\widehat{\rho}$ and its 90% CI in the same graph. To obtain the constant parameter MUE's denoted by the flat red solid lines and constant parameter CI's denoted by the flat red dotted lines, we fix $n\widehat{h}_{us} = 2n$ and apply our method. Note that the

(a) US Inflation, $n\widehat{h}_{us} = 125$

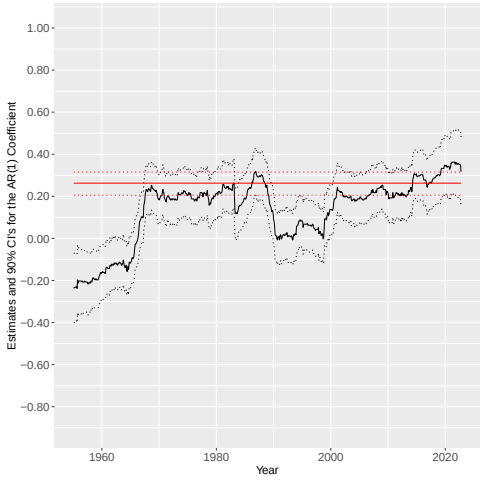


(b) US Inflation, $1.5n\widehat{h}_{us} = 188$

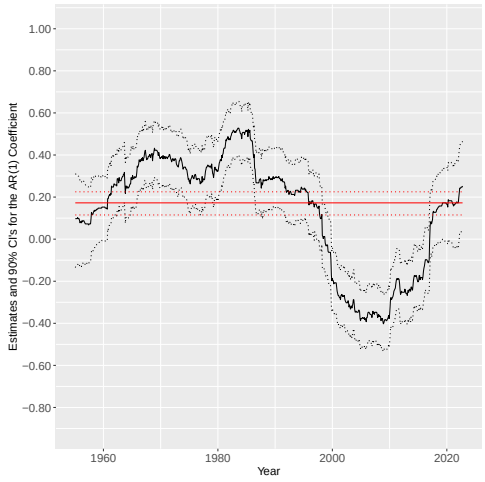


(c) Canada Inflation, $n\widehat{h}_{us} = 125$

(d) Canada Inflation, $1.5n\widehat{h}_{us} = 188$



(e) Germany Inflation, $n\widehat{h}_{us} = 125$



(f) Germany Inflation, $1.5n\widehat{h}_{us} = 188$

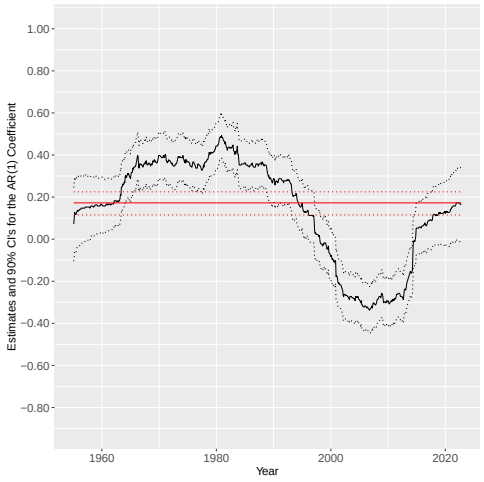


Figure 4: Estimates and 90% CI's for the AR(1) Coefficient in TVP-AR(1) Models: US, Canada, and Germany Inflation. The solid black line is the MUE's of $\rho(t)$ in a TVP-AR(1) model, with the 90% CI's denoted by the dotted black line. The solid red line is the MUE's of the AR coefficient in a constant parameter AR(1) model, with the 90% CI's denoted by the dotted red line.

method to derive the constant parameter estimates is equivalent to Mikusheva's (2007) modification of Stock's (1991) method. Mikusheva (2007) proves that these confidence sets are uniformly valid asymptotically for non-time-varying $\rho \in (0, 1]$.

Inflation persistence in the US has been extensively studied with mixed results regarding its extent and causes. Studies like those by Cogley and Sargent (2001) have noted a decrease in inflation persistence following the early 1980s, attributing this change to shifts in monetary policy, especially the Federal Reserve's (Fed) increased focus on inflation targeting. This view is supported by research that points to the Volcker disinflation period as a pivotal time when the Fed's credibility was enhanced, leading to a stabilization of inflation expectations, see Sims (1992). We also find a sharp decline from .78 to .12 in inflation persistence in the US measured by the MUE of the TVP-AR coefficient between 1983 and 1995 in Figure 4(a), which is consistent with the above empirical results.

Entering the era of 2000s, inflation persistence remains a central topic in macroeconomic research. One line of research (e.g., Eggertsson and Woodford (2003), Chung, Laforte, Reifschneider, and Williams (2012)) concerns zero lower bound (ZLB) effects when the Fed set the interest rates close to zero. The theory predicts that at ZLB, monetary policy could be less effective in controlling inflation, thereby potentially increasing inflation persistence if not coupled with assertive non-traditional interventions. To overcome the challenges of the ZLB effects on the effectiveness of its monetary policies, the Fed initiated a new practice called "forward guidance," where the future course of monetary policy is communicated to the public by the central bank. Numerous studies (e.g., Campbell, Evans, Fisher, Justiniano, Calomiris, and Woodford (2012), Del Negro, Giannoni, and Patterson (2023)) have suggested that forward guidance is effective in enabling the Fed to better control inflation, which could lead to lower inflation persistence. Along this line, Bernanke (2020) argues that forward guidance combined with quantitative easing (QE) granted the Fed significantly more space to provide accommodation when its standard policy rate was near zero. More recently, Cole, Martinez-Garcia, and Sims (2023) argue that despite the significant efforts to make their policy credible, the credibility of most central banks including the Fed has been generally declining, making monetary policies aimed at controlling inflation less effective. A concrete manifestation of their claim would be an increase in the inflation persistence over a longer horizon.

We find an increase in inflation persistence in the US from .3 to .6 measured by the MUE of TVP-AR coefficient between 2000 and 2008 in Figure 4(a), which seems to be mostly driven by the ZLB effects despite the introduction of forward guidance during this period. Later on, when more specific monetary policies (e.g., QE) were introduced to stimulate the economy following the 2008 financial crisis, inflation persistence drops from .6 to .4 between 2010 and

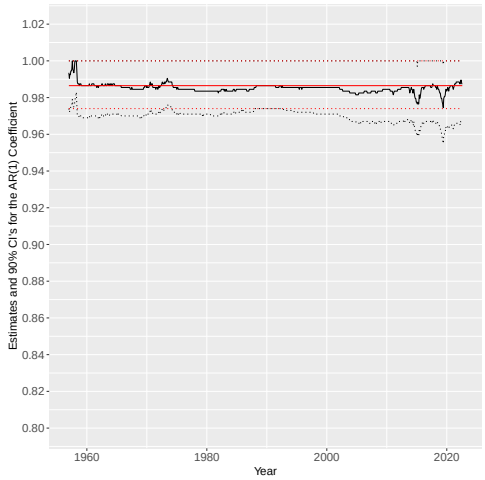
2013, in line with the result of Bernanke (2020). With that said, we observe in Figure 4(a) an upward trend in inflation persistence in the US over a longer horizon between 2000 and 2022, which could be caused by a decline in the credibility of the Fed as claimed by Cole, Martinez-Garcia, and Sims (2023). Overall, we conclude that while the Fed's measures have been effective in controlling inflation expectations and persistence, more nuanced policies are needed to handle complications arising from the ZLB effects, heterogeneous beliefs (Andrade, Gaballo, Mengus, and Mojon (2019)), structural changes (e.g., higher volatility and shifts in consumer behavior and supply chains) caused by the Covid-19 and other factors.

We also apply our method to study inflation persistence in Canada and Germany and obtain the following results. First, we find significant time variation in inflation persistence as measured by the MUE's of $\rho(t)$ for both countries since 1980s. Given the magnitude of the time-variation, our findings show that inflation persistence is more likely to be driven by changes in the regime of monetary policy and credibility of central banks, rather than by nominal or real frictions in the economy. Second, there seems to be a universal upward trend since 2000 in inflation persistence for all countries under consideration, echoing the findings of Cole, Martinez-Garcia, and Sims (2023), who argue that there is a decline in credibility in central bank policies. Third, the introduction of the Euro in 1999 marked a significant shift in inflation persistence in Germany, with the European Central Bank (ECB) taking over monetary policy. Inflation rates remained relatively stable but were influenced by broader Eurozone policies and economic conditions. The early 2000s saw moderate inflation, which aligned closely with the ECB's target. Our method captures this regime change by showing drastically different patterns of inflation persistence before and after 1999, supporting the conclusion above that inflation persistence is more likely to be driven by changes in the regime of monetary policy and the credibility of central banks.

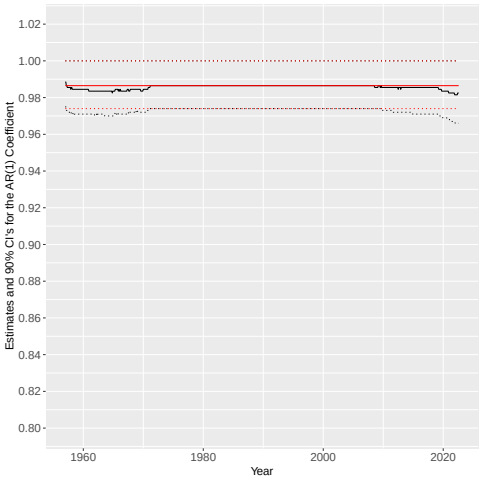
6.2 Real Exchange Rate

The second application concerns the real exchange rate. The results are given in Figure 5. We use US dollars (USD) as the benchmark currency, and calculate the bilateral real exchange rate re_{it} of country i at time t as $rex_{it} = nex_{it} \times \frac{CPI_{it}}{CPI_{0t}}$, where nex_{it} is the nominal exchange rate (USD per domestic currency) at time t, CPI_{it} is the price level of country i at time t, and CPI_{0t} is the price level of the US at time t. Therefore, an increase in rex_{it} represents an appreciation of country i's currency against USD. We report results for the UK, Sweden, and Switzerland. The data span for the monthly real exchange rate dataset is Jan 1957 to Aug 2022. Thus, $n = 788$ for each country. Similar to the inflation series, we compute the $n\widehat{h}_{us}$ values based on (4.5). Then, we compute the MUE's and 90% CI's for the

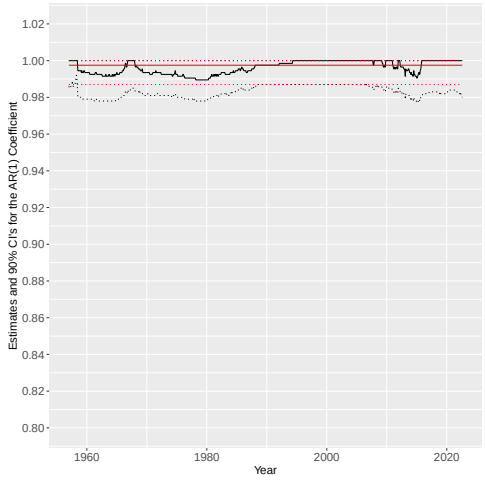
(a) UK Real Exchange Rate, $n\widehat{h}_{us} = 823$



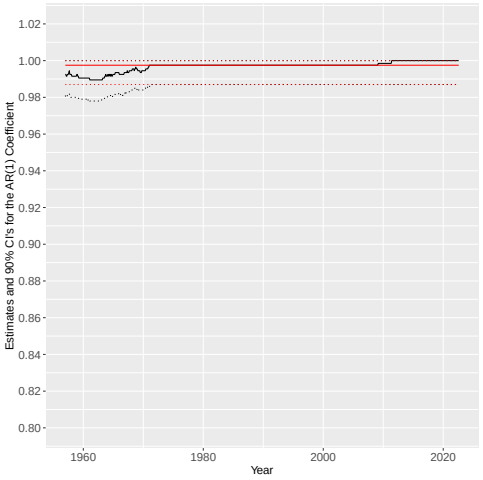
(b) UK Real Exchange Rate, $1.5n\widehat{h}_{us} = 1{,}234$



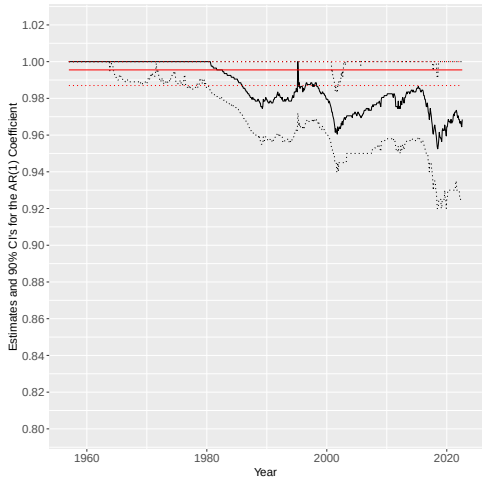
(c) Sweden Real Exchange Rate, $n\widehat{h}_{us} = 823$



(d) Sweden Real Exchange Rate, $1.5n\widehat{h}_{us} = 1{,}234$



(e) Switzerland Real Exchange Rate, $n\widehat{h}_{us} = 393$



(f) Switzerland Real Exchange Rate, $1.5n\widehat{h}_{us} = 590$

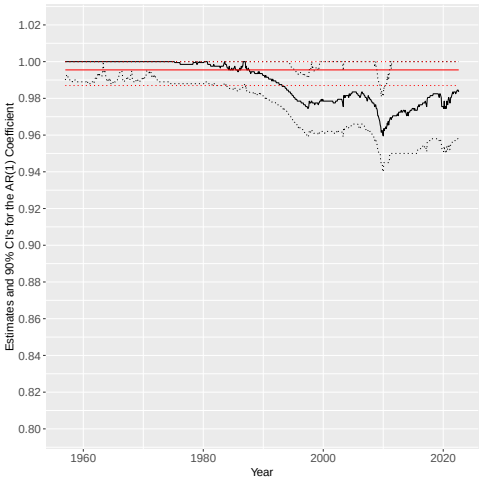


Figure 5: Estimates and 90% CI's for the AR(1) Coefficient in TVP-AR(1) Models: UK, Sweden, and Switzerland Real Exchange Rate

AR coefficient, $\rho(t)$, at each t and for each country. For robustness purposes, we multiply the undersmoothed $n\widehat{h}_{us}$ by 1.5 and report the results in the right column of each figure. For the real exchange rate series considered in this section, Ljung-Box tests based on six lags of the residuals from the AR(1) model fail to reject the null hypothesis of no autocorrelation at the 5% significance level, see Table SM.2.

It has been known in the literature that real exchange rates in developed countries tend to be highly persistent, with deviations from the PPP level taking a long time to disappear (Rogoff (1996), Engel (2014)). There are several explanations in the literature for the real exchange rate persistence, including nominal price rigidities, interest rate inertia in monetary policy (Benigno (2004)), heterogeneous dynamics in subcomponents (Imbs, Mumtaz, Ravn, and Rey (2005)), Balassa–Samuelson effects (Balassa (1964), Samuelson (1964)), to name a few. Figure 5 presents the results on real exchange rate persistence measured by the MUE's of $\rho(t)$ for the UK, Sweden, and Switzerland. Across all three countries the MUE's are close to 1, showing that our method yields near constant graphs in scenarios where that seems to be suitable. The 90% CI's for $\rho(t)$ are also very short. Chari, Kehoe, and McGrattan (2002) report estimates from a constant parameter autoregressive process to lie between .76 and .87 for US bilateral real exchange rates against nine developed European countries using data between 1972 and 1994. They developed a general equilibrium model where the firms can only set price once per year, which implies a very high level of price stickiness in the economy. Yet, their model cannot generate the level of persistence observed in the data. Allowing for a time-varying parameter autoregressive process and using data from a longer time horizon, our MUE's of $\rho(t)$ for the three currencies are even higher at close to one, supporting their claim that nominal price rigidities are not enough to explain the high real exchange rate persistence. Practitioners would need to seek other possible explanations such as the Balassa–Samuelson effects or heterogeneous dynamics in subcomponents.

While high real exchange rate persistence seems prevalent, there is heterogeneity in the pattern of persistence across the countries we consider. In particular, the MUE's of $\rho(t)$ for Switzerland demonstrate a downward trend since early 1980s in Figure 5(e). The Swiss National Bank (SNB) has adopted several significant monetary policy changes since the 1990s, including the shift to a three-fold target (price stability, 3-year inflation forecast, and a range for the 3M Libor, see Jordan, Peytrignet, and Rossi (2010)) in the late 1990s and the introduction of negative interest rates in 2015 when it was forced to abandon a policy of defending the Swiss franc with a peg to the euro. These policies aim to stabilize price levels and influence interest rates, which can affect exchange rate dynamics and potentially reduce persistence by promoting quicker adjustments to shocks. In Figure 5(e) we observe a sharp decline in the MUE of AR(1) coefficient between 1995 and 2002 and between 2015

and 2019, consistent with the timing of SNB's monetary policy changes. Meanwhile, the European debt crisis and subsequent economic turmoil in the Eurozone led to significant safe-haven flows into the Swiss franc, prompting the SNB to implement a cap on the franc's value against the euro in 2011. This cap was removed in 2015. Such a cap on the franc's value would increase its real exchange rate persistence since the price of the franc could have fluctuated above the cap for the duration of the policy. We also find the real exchange rate persistence to be slightly increasing between 2011 and 2015 in Figure 5(e). As shown by these results, our method can capture the major events and policy changes that affect real exchange rate persistence reasonably accurately.

7 Asymptotics

This section establishes the correct uniform asymptotic size and asymptotic similarity of the confidence interval $CI_{n,\tau}$ for $\rho(\tau)$, the median-unbiased property of $\widetilde{\rho}_{n\tau}$, and some asymptotic properties of the data-dependent bandwidth parameter $\widehat{h}$. To prove these results, this section provides asymptotic results for the LS estimator $\widehat{\rho}_{n\tau}$ and t-statistic $T_n(\rho_{0,n})$ under drifting sequences of parameter values. All proofs of the results stated below are given in Section B of the Supplemental Material.

7.1 Parameter Space

Let $I_{a,r} := [a-r, a+r]$ for $a \in \mathbb{R}$ and $r > 0$. Let $\lfloor x \rfloor$ denote the integer part of x.

We impose the following structure on the ρ function: for some $\varepsilon_2, \varepsilon_3 > 0$,

$$\rho(s) = 1 - \kappa(s)/b \tag{7.1}$$

for $s \in I_{\tau,\varepsilon_2}$ and some $b \in [\varepsilon_3, \infty]$, where $\kappa(\cdot)$ is a nonnegative twice continuously differentiable function on I_{τ,ε_2}.

The parameter space for $(\rho, \mu, \sigma^2, \kappa, b, F)$ is given by

$\Lambda_n = \{\lambda = (\rho, \mu, \sigma^2, \kappa, b, F)$:

(i) ρ, μ, and σ^2 are Lipschitz functions from $[0,1]$ to $[-1+\varepsilon_1, 1]$, $[C_{2,L}, C_{2,U}]$, and $[C_{3,L}, C_{3,U}]$, respectively, with Lipschitz constants bounded by L_1, L_2, and L_3, respectively;

(ii) $\rho(s) = 1 - \kappa(s)/b$ for $s \in I_{\tau,\varepsilon_2}$ and $b \in [\varepsilon_3, \infty]$, where $\kappa(\cdot)$ is a twice continuously differentiable function from I_{τ,ε_2} to $[0, C_4]$ with Lipschitz constant bounded by L_4 and $\kappa(\tau) \geq \varepsilon_4$;

(iii) $\mu(s) = C_{\mu 1} \exp\{-\eta(s)/b\} + C_{\mu 2}$ for $s \in I_{\tau,\varepsilon_2}$ and $b \in [\varepsilon_3, \infty]$, where $\eta(\cdot)$ is a nonnegative Lipschitz function with Lipschitz constant bounded by L_5;

(iv) $\{U_t : t = 0, 1, ..., n\}$ is a stationary martingale difference sequence under F with $E_F(U_t|\mathscr{G}_{t-1}) = 0$ a.s., $E_F(U_t^2|\mathscr{G}_{t-1}) = 1$ a.s., and $E_F[U_t^4|\mathscr{G}_{t-1}] < M$ a.s., where $\mathscr{G}_t$ is some non-decreasing sequence of σ-fields for which $\sigma(U_0, ..., U_t) \subseteq \mathscr{G}_t$ and $\sigma(Y_0^*) \in \mathscr{G}_t$ for $t = 1, ..., n$;

(v) $E_F(Y_0^*)^2 \leq C_5 n$;

for some constants $0 < \varepsilon_1 < 2$, $\varepsilon_2, \varepsilon_3, \varepsilon_4 > 0$, $-\infty < C_{j,L} \leq C_{j,U} < \infty$ for $j = 2, 3$, $C_{3,L} > 0$, $0 \leq C_4 < \infty$, $0 \leq C_5 < \infty$, $L_j < \infty$ $\forall j \leq 5$, $C_{\mu 1}, C_{\mu 2} \in (-\infty, \infty)$, and $M \in (0, \infty)\}$.

Note that the dependence on n of Λ_n is only through part (v), which concerns the initial condition Y_0^*.[4] Also, $\mu(s)$ is defined in such a way that it is allowed to vary smoothly on $\mathbb{R}$. It is used in the proof of Lemma 7.1(d), which is further used to bound the absolute difference between $\mu(t)$ and $\mu(s)$ for $t, s \in [T_1, T_2]$ in obtaining the asymptotic distribution of the t-statistic $T_n(\rho_{0,n})$.

To establish the correct asymptotic size of the CI for $\rho(\tau)$, we need to consider sequences of parameters $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n)\}_{n \geq 1}$ in Λ_n. The parameter space Λ_n is defined to allow the sequence $\{\rho_n(\tau)\}_{n \geq 1}$ to equal one for all $n \geq 1$, converge to one at any rate, or be bounded away from one. More specifically, by the definition of $\rho_n(\cdot)$,

$$\rho_n(\tau) = 1 - \kappa_n(\tau)/b_n \text{ for some } b_n \in [0, \infty]. \tag{7.2}$$

If $b_n \to \infty$, then $\rho_n(\tau)$ is local to one. If $\limsup_{n\to\infty} b_n < \infty$, then $\rho_n(\tau)$ is bounded away from one. Furthermore, Λ_n is defined so that, for points $s \in [0, 1]$ other than τ, $\rho_n(s)$ can converge to one at any rate or be bounded away from one, and the distance-from-unity behavior may differ between different s outside of I_{τ,ε_2}.

7.2 Correct Asymptotic Size of the Confidence Interval for $\rho(\tau)$

The bandwidth h is assumed to satisfy the following assumptions.

Assumption 1 (Bandwidth h). *$h \to 0$ and $nh \to \infty$ as $n \to \infty$.*

Assumption 2 (Order of h). *$nh^5 \to 0$.*[5]

[4] The parameter space Λ_n could be made independent of n by specifying that $E(Y_0^*)^2 \leq C_5$. But, allowing the bound on $E(Y_0^*)^2$ to be $C_5 n$, allows the parameter space to expand with n and is in accord with assumptions in the literature for AR models with a possible unit root or near unit root.

[5] Assumption 2 is used in the proofs of Lemmas 7.2(b), 7.2(c), 7.7, and 7.8(a) below.

Let $P_\lambda(\cdot)$ denote probability under $\lambda \in \Lambda_n$. The correct asymptotic size and asymptotic similarity of the CI $CI_{n,\tau}$ are established in the following theorem. The theorem's proof relies on the "drifting pointwise" asymptotic results given in Sections 7.3-7.5 below and the generic asymptotic size results in Andrews, Cheng, and Guggenberger (2020).

Theorem 7.1 (**Correct Asymptotic Size of $CI_{n,\tau}$**). *Under Assumptions 1 and 2,*

$$\liminf_{n\to\infty} \inf_{\lambda\in\Lambda_n} P_\lambda(\rho(\tau) \in CI_{n,\tau}) = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda_n} P_\lambda(\rho(\tau) \in CI_{n,\tau}) = 1-\alpha.$$

The median-unbiased interval estimator $\widetilde{\rho}_{n\tau}$ has the following median-unbiasedness property. This result is a corollary to one-sided versions of Theorem 7.1.[6]

Corollary 7.1 (**Asymptotic Median-Unbiasedness of $\widetilde{\rho}_{n\tau}$**). *Under Assumptions 1 and 2,*

$$\liminf_{n\to\infty} \inf_{\lambda\in\Lambda_n} P_\lambda(\widetilde{\rho}_{n\tau,up} \geq \rho(\tau)) \geq 1/2 \text{ and}$$
$$\liminf_{n\to\infty} \inf_{\lambda\in\Lambda_n} P_\lambda(\widetilde{\rho}_{n\tau,low} \leq \rho(\tau)) \geq 1/2.$$

7.3 Preliminaries

Define

$$c_{t,j} := \prod_{k=0}^{j-1} \rho_{t-k} \text{ for } 1 \leq j \leq t,\ 1 \leq t \leq n, \quad \text{and } c_{t,0} := 1. \tag{7.3}$$

We consider the interval $I_{n\tau,nh/2}$ and decompose Y_t^* into the sum of two parts

$$Y_t^* = Y_t^0 + c_{t,t-T_0} Y_{T_0}^*, \text{ for } t \in I_{n\tau,nh/2}, \tag{7.4}$$

where Y_t^0 is an AR(1) process with the same time-varying parameters ρ_t as Y_t^*, but with a zero initial condition at $T_0 := T_1 - 1$, and $c_{t,t-T_0}$ is defined in (7.3). By recursive substitution in (2.1), we have

$$Y_t^0 = \sum_{j=0}^{t-T_1} c_{t,j}\sigma_{t-j}U_{t-j} \text{ and } Y_t^* = \sum_{j=0}^{t-T_1} c_{t,j}\sigma_{t-j}U_{t-j} + c_{t,t-T_0}Y_{T_0}^* \text{ for } t = T_1, ..., T_2. \tag{7.5}$$

[6]Corollary 7.1 holds because $CI_{n,\tau}^{up}(.5)$ and $CI_{n,\tau}^{low}(.5)$ both have coverage probabilities of 1/2 or greater by the proof of Theorem 7.1 applied to these one-sided CIs and $(\widetilde{\rho}_{n\tau} \geq \rho(\tau)) \supset CI_{n,\tau}^{up}(.5)$ and $(\widetilde{\rho}_{n\tau} \leq \rho(\tau)) \supset CI_{n,\tau}^{low}(.5)$.

Then, from (2.1) and (7.4), we have

$$Y_t = \mu_t + \sum_{j=0}^{t-T_1} c_{t,j}\sigma_{t-j}U_{t-j} + c_{t,t-T_0}Y^*_{T_0} \text{ for } t = T_1, ..., T_2. \tag{7.6}$$

We consider sequences $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$. The estimand of interest is $\rho_n(\tau)$ for $n \geq 1$.

Let

$$\begin{aligned} \rho_{n,t} &:= \rho_n(t/n),\ \mu_{n,t} := \mu_n(t/n),\ \sigma^2_{n,t} := \sigma^2_n(t/n), \\ \rho_{n\tau} &:= \rho_n(\tau),\ \mu_{n\tau} := \mu_n(\tau),\ \text{and } \sigma^2_{n\tau} := \sigma^2_n(\tau). \end{aligned} \tag{7.7}$$

By part (ii) of the definition of Λ_n, for $t \in I_{n\tau,nh/2}$,

$$\rho_{n,t} := \rho_n(t/n) = 1 - \kappa_n(t/n)/b_n \tag{7.8}$$

for n sufficiently large that $h/2+1/n \leq \varepsilon_2$. For notational simplicity, we drop the n subscripts on $\rho_{n,t}$, $\mu_{n,t}$, and $\sigma^2_{n,t}$ below.

We consider sequences of null hypotheses $H_0 : \rho_{n\tau} = \rho_{0,n}$ for $n \geq 1$. As noted above, the CI $CI_{n,\tau}$ is defined by the inversion of tests of such null hypotheses.

For some results, we assume the functions $\kappa_n(\cdot)$, $\mu_n(\cdot)$, and $\sigma^2_n(\cdot)$ converge.

Assumption 3 (**Limits of $\boldsymbol{\kappa_n(\cdot)}$, $\boldsymbol{\mu_n(\cdot)}$, and $\boldsymbol{\sigma^2_n(\cdot)}$**). *$\kappa_n(\cdot)$, $\mu_n(\cdot)$, and $\sigma^2_n(\cdot)$ restricted to I_{τ,ε_2} converge uniformly to some functions $\kappa_0(\cdot)$, $\mu_0(\cdot)$, and $\sigma^2_0(\cdot)$ on I_{τ,ε_2}, respectively.*

For some results, we assume that $\left\{b_n/\left(nh^{1/2}\right)\right\}_{n\geq 1}$ converges.

Assumption 4 (**Convergence of $\boldsymbol{b_n/\left(nh^{1/2}\right)}$**). *$b_n/\left(nh^{1/2}\right) \to w_0$ for some $w_0 \in [0, \infty]$.*

Assumptions 3 and 4 are innocuous because, to establish the uniform inference results in Theorem 7.1, we show that it suffices to consider sequences $\{\lambda_n \in \Lambda_n\}_{n\geq 1}$ that satisfy these assumptions. Note that Theorem 7.1 and Corollary 7.1 do not impose Assumption 3 or 4.

The following lemma bounds the maximum intertemporal difference between the TVP ρ_t and $\rho_{n\tau}$ for $t \in [T_1, T_2]$.

Lemma 7.1 (**Maximum Intertemporal Differences on $[\boldsymbol{T_1}, \boldsymbol{T_2}]$**). *Under Assumptions 1 and 3, for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma^2_n, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$,*

(a) $\max_{t\in[T_1,T_2]} |\rho_t - \rho_{n\tau}| = O\left(h/b_n\right)$,

(b) $\max_{t\in[T_1,T_2]} |\sigma_t^2 - \sigma_{n\tau}^2| = O\left(h\right)$,

(c) $\max_{t\in[T_1,T_2]} \max_{0\le j\le t-T_1} |c_{t,j} - \rho_{n\tau}^j| = O\left(nh^2/b_n\right)$, *and*

(d) $\max_{t\in[T_1,T_2]} |\mu_t - \mu_{n\tau}| = O\left(h/b_n\right)$.

The next lemma provides several bounds on the endogenous initial condition $Y_{T_0}^*$.

Lemma 7.2 (**Order of $Y_{T_0}^*$**). *For a sequence* $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\ge1}$, *we have*

(a) $Y_{T_0}^* = O_p\left(n^{1/2}\right)$ *under Assumption 1,*

(b) $Y_{T_0}^* = o_p\left(b_n/\left(nh\right)^{1/2}\right)$ *under Assumptions 1, 2, 3, and 4 and* $nh/b_n \to r_0 = 0$, *and*

(c) $Y_{T_0}^* = O_p\left(b_n^{1/2}\right)$ *under Assumptions 1 and 2 and* $nh/b_n \to r_0 = \infty$.

Remark 7.1. Lemma 7.2(a) shows that part (v) of Λ_n implies that the endogenous initial condition $Y_{T_0}^*$ is $O_p\left(n^{1/2}\right)$. Part (b) of the lemma is used in the "very local-to-unity" case in which $nh/b_n \to r_0 = 0$. It provides a better bound than part (a) when $b_n/\left(nh^{1/2}\right) \to w_0 < \infty$ because, in that case, the bound $o_p\left(b_n/\left(nh\right)^{1/2}\right)$ is $o_p\left(n^{1/2}\right)$. Part (c) of the lemma provides a better bound than part (a) in the "stationary" case in which $nh/b_n \to r_0 = \infty$.

7.4 Asymptotics in the Local-to-Unity Case

The local-to-unity case is characterized by $nh/b_n \to r_0 \in [0,\infty)$. In this section, we determine the asymptotic distributions of the local LS estimator $\widehat{\rho}_{n\tau}$ and corresponding t-statistic $T_n\left(\rho_{0,n}\right)$ in the local-to-unity case.

Define

$$t\left(s\right) := t_{n,\tau}\left(s\right) = T_1 + \lfloor nhs \rfloor = \lfloor n\tau \rfloor - \lfloor nh/2 \rfloor + \lfloor nhs \rfloor \tag{7.9}$$

as a function of $s \in [0,1]$ for any fixed $\tau \in [0,1]$. First, we obtain the asymptotic distribution of the zero-initial condition process $\left(nh\right)^{-1/2} Y_{n,t(s)}^0$ for $s \in [0,1]$. Then, we obtain the asymptotic distribution of the normalized initial condition $Y_{T_0}^*$ in the $r_0 \in (0,\infty)$ case. Lastly, we obtain the asymptotic distributions of $\widehat{\rho}_{n\tau}$ and $T_n\left(\rho_{0,n}\right)$. The last step includes dealing with the initial condition $Y_{T_0}^*$ in the $r_0 = 0$ case.

Lemma 7.3 (**Asymptotic Distribution of $Y_{n,t(s)}^0$**). *Under Assumptions 1 and 3, for a sequence* $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\ge1}$ *and* $nh/b_n \to r_0 \in [0,\infty)$ (*local-to-unity case*), *we have*

$$\left(nh\right)^{-1/2} Y_{n,t(s)}^0/\sigma_0\left(\tau\right) \Rightarrow I_\psi\left(s\right) \;\; for \;\; \psi = r_0\kappa_0(\tau),$$

where $t(s)$ is defined in (7.9), $I_\psi(s)$ is defined in (3.6), and "$\Rightarrow$" denotes weak convergence with respect to the Skorohod metric.

The following lemma is used to determine the effect of the initial condition on the LS estimator and t-statistic when $r_0 \in (0, \infty)$.

Lemma 7.4 (**Asymptotic Distribution of the Initial Condition $Y^*_{T_0}$**). *Under Assumptions 1 and 3, for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$ and $nh/b_n \to r_0 \in (0, \infty)$ (local-to-unity case), we have*

$$(2\psi/nh)^{1/2} Y^*_{T_0}/\sigma_0(\tau) \to_d Z_1 \sim N(0,1),$$

where $\psi := r_0\kappa_0(\tau)$, Z_1 is independent of $B(\cdot)$, $I_\psi(\cdot)$ defined in (3.6), and the convergence holds jointly with the convergence in Lemma 7.3.

The following lemma is useful in determining the asymptotic properties of $\widehat{\rho}_{n\tau}$ in the local-to-unity case.

Lemma 7.5 (**Convergence of Components in the Local-to-Unity Case**). *Under the null hypothesis $H_0 : \rho_{n\tau} = \rho_{0,n}$, Assumptions 1 and 3, and $nh/b_n \to r_0 \in (0, \infty)$ (local-to-unity case), for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$ and $\psi = r_0\kappa_0(\tau)$, the following results hold jointly*

(a) $(nh)^{-1/2} Y_{t(s)}/\sigma_0(\tau) \Rightarrow I^*_\psi(s)$, *where* $t(s) = T_1 + \lfloor nhs \rfloor$ *for* $s \in [0,1]$,

(b) $(nh)^{-3/2} \sum_{t=T_1}^{T_2} Y_{t-1}/\sigma_0(\tau) \to_d \int_0^1 I^*_\psi(s)\, ds$,

(c) $(nh)^{-2} \sum_{t=T_1}^{T_2} Y^2_{t-1}/\sigma_0^2(\tau) \to_d \int_0^1 I^{*2}_\psi(s)\, ds$,

(d) $(nh)^{-1/2} \sum_{t=T_1}^{T_2} U_t\sigma_t/\sigma_0(\tau) \to_d \int_0^1 dB(s)$,

(e) $(nh)^{-1} \sum_{t=T_1}^{T_2} Y_{t-1}U_t\sigma_t/\sigma_0^2(\tau) \to_d \int_0^1 I^*_\psi(s)\, dB(s)$,

(f) $(nh)^{-2} \sum_{t=T_1}^{T_2} \left(Y^0_{t-1}/\sigma_0(\tau)\right)^2 \to_d \int_0^1 I^2_\psi(s)\, ds$, *and*

(g) *when $r_0 = 0$, parts* (a)–(c) *and* (e) *hold with $Y_{t(s)}$ $(= \mu_{t(s)} + Y^0_{t(s)} + c_{t(s),t(s)-T_0} Y^*_{T_0})$ and Y_{t-1} replaced by $\mu_{t(s)} + Y^0_{t(s)}$ and $\mu_{t-1} + Y^0_{t-1}$, respectively.*

After proper re-scaling, we have

$$nh(\widehat{\rho}_{n\tau} - \rho_{0,n}) = \frac{(nh)^{-1} \sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)(Y_t - \rho_{0,n}Y_{t-1})}{(nh)^{-2} \sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)^2}. \tag{7.10}$$

The next theorem gives the asymptotic distribution of $nh\left(\widehat{\rho}_{n\tau}-\rho_{0,n}\right)$ and the t-statistic $T_n\left(\rho_{0,n}\right)$ in the local-to-unity case.

Theorem 7.2 (**Asymptotic Distribution of Normalized $\widehat{\rho}_{n\tau}$ and t-Statistic in the Local-to-Unity Case**). *Under the null hypothesis $H_0:\rho_{n\tau}=\rho_{0,n}$, Assumptions 1, 2, and 3, $nh/b_n\to r_0\in[0,\infty)$ (local-to-unity case), and Assumption 4 if $r_0=0$, for a sequence $\{\lambda_n=(\rho_n,\mu_n,\sigma_n^2,\kappa_n,b_n,F_n)\in\Lambda_n\}_{n\geq 1}$ and $\psi=r_0\kappa_0(\tau)$, we have*

$$nh\left(\widehat{\rho}_{n\tau}-\rho_{0,n}\right)\to_d\left(\int_0^1 I_{D,\psi}^{*2}(s)\,ds\right)^{-1}\int_0^1 I_{D,\psi}^{*}(s)\,dB(s)$$

and

$$T_n\left(\rho_{0,n}\right)\to_d\left(\int_0^1 I_{D,\psi}^{*2}(s)\,ds\right)^{-1/2}\int_0^1 I_{D,\psi}^{*}(s)\,dB(s).$$

Remark 7.2. For any subsequence $\{p_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Lemmas 7.1–7.5 and Theorem 7.2 hold with p_n in place of n throughout and h_{p_n} in place of $h=h_n$.

7.5 Asymptotics in the Stationary Case

The "stationary" case is characterized by $b_n=o(nh)$, or equivalently, $nh/b_n\to r_0=\infty$. The stationary case is defined such that the t-statistic has a standard normal distribution under $H_0:\rho_{n\tau}=\rho_{0,n}$ in the stationary case. If b_n is bounded, then $\rho_{n\tau}\leq C_\rho$ for some constant $C_\rho<1$ for all $n\geq 1$. If b_n diverges to infinity, then $\rho_{n\tau}$ goes to one at a rate slower than $1/nh$. Thus, the stationary case includes some scenarios where $\rho_{n\tau}$ goes to one.

Define

$$\overline{\rho}_n:=\max\left\{\exp\left\{-\kappa_0(\tau)/(2b_n)\right\},-1+\varepsilon_1\right\},\tag{7.11}$$

where $-1+\varepsilon_1$ is a lower bound on ρ_t by the definition of Λ_n. We can bound $|\rho_t|$ for $t\in[T_1,T_2]$ using $\overline{\rho}_n$. First, suppose $\rho_t\leq 0$, then $-1+\varepsilon_1\leq\rho_t\leq 0$, which implies that $|\rho_t|\leq\overline{\rho}_n$, as desired. Given this, in the following calculations we suppose $\rho_t\geq 0$ for all $t\in[0,1]$ without loss of generality. Then, for n sufficiently large, we have

$$\begin{aligned}\max_{t\in[T_1,T_2]}|\rho_t|&\leq\max_{t\in[T_1,T_2]}|\rho_t-\rho_{n\tau}|+\left|\rho_{n\tau}-\exp\left\{-\kappa_0(\tau)/b_n\right\}\right|+\exp\left\{-\kappa_0(\tau)/b_n\right\},\\&\leq O(h/b_n)+o(1)/b_n+\exp\left\{-\kappa_0(\tau)/b_n\right\}\leq\overline{\rho}_n,\end{aligned}\tag{7.12}$$

where $b_n\geq\varepsilon_3>0$, the second inequality uses Lemma 7.1(a), (7.8), Assumption 3, and a mean value expansion of the $\exp\{\cdot\}$ function, and the last inequality holds using $\kappa_0(\tau)>0$ and the fact that when $b_n\to\infty$, $\overline{\rho}_n-\exp\{-\kappa_0(\tau)/b_n\}\geq K/b_n$ for some constant $K>0$.

Equation (7.12) implies

$$|c_{t,j}| = \left| \prod_{k=0}^{j-1} \rho_{t-k} \right| \leq \overline{\rho}_n^j \text{ for } j = 1, ..., t - T_1 + 1 \text{ and } t = T_1, ..., T_2. \tag{7.13}$$

Furthermore, using Lemma A.1 of Giraitis, Kapetanios, and Yates (2014) and the definition of the parameter space Λ_n, under $H_0 : \rho_{n\tau} = \rho_{0,n}$, we have

$$\left| c_{t,j} - \rho_{0,n}^j \right| \leq j\overline{\rho}_n^{j-1} \max_{k \in [T_1, T_2]} |\rho_k - \rho_{0,n}| = j\overline{\rho}_n^{j-1} O\left(h/b_n\right) \text{ for } j = 0, ..., t - T_0 \text{ and } t = T_1, ..., T_2, \tag{7.14}$$

where the equality holds by Lemma 7.1(a).

The following results are used in the analysis:

$$\sum_{t=0}^{\infty} \overline{\rho}_n^t = \left(1 - \overline{\rho}_n\right)^{-1} = O\left(b_n\right), \tag{7.15}$$

$$\sum_{t=0}^{\infty} \overline{\rho}_n^{2t} = \left(1 - \overline{\rho}_n^2\right)^{-1} = O\left(b_n\right), \tag{7.16}$$

$$\sum_{t=0}^{\infty} t\overline{\rho}_n^t = \overline{\rho}_n \left(1 - \overline{\rho}_n\right)^{-2} = O\left(b_n^2\right), \tag{7.17}$$

$$\sum_{t=0}^{\infty} t^2\overline{\rho}_n^{2t} = \overline{\rho}_n^2 \left(1 + \overline{\rho}_n^2\right) \left(1 - \overline{\rho}_n^2\right)^{-3} = O\left(b_n^3\right), \text{ and} \tag{7.18}$$

$$\sum_{t>s=0}^{nh} \overline{\rho}_n^{t-s} = \sum_{t=1}^{nh} \sum_{s=0}^{t-1} \overline{\rho}_n^{t-s} = O\left(nhb_n\right). \tag{7.19}$$

When $\rho_{0,n} = 1 - \kappa_n(\tau)/b_n$ with $b_n = o(nh)$, one can replace $\overline{\rho}_n$ with $\rho_{0,n}$ in (7.15)–(7.19) and the results still hold.

We now establish the $N(0,1)$ asymptotic distribution of $\widehat{\rho}_{n\tau}$ in the stationary case with the following normalization:

$$\begin{aligned} &\left(1 - \rho_{0,n}^2\right)^{-1/2} (nh)^{1/2} \left(\widehat{\rho}_{n\tau} - \rho_{0,n}\right) \\ &= \frac{\left(1 - \rho_{0,n}^2\right)^{1/2} (nh)^{-1/2} \sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right) \left(Y_t - \rho_{0,n} Y_{t-1}\right)}{\left(1 - \rho_{0,n}^2\right) (nh)^{-1} \sum_{t=T_1}^{T_2} \left(Y_{t-1} - \overline{Y}_{nh,-1}\right)^2}. \end{aligned} \tag{7.20}$$

Next, we prove a lemma on the asymptotic properties of the zero-initial condition process Y_t^0 in the stationary case.

Lemma 7.6 (**Asymptotics in the Stationary Case**). *Under the null hypothesis* H_0 :

$\rho_{n\tau} = \rho_{0,n}$ and Assumptions 1 and 3, for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$ where $nh/b_n \to r_0 = \infty$ and $\kappa_0(\tau) > 0$ (stationary case), the following results hold jointly

(a) $\left(1-\rho_{0,n}^2\right)^{1/2}(nh)^{-1}\sum_{t=T_1}^{T_2} Y_{t-1}^0/\sigma_0(\tau) \to_p 0$,

(b) $\left(1-\rho_{0,n}^2\right)(nh)^{-1}\sum_{t=T_1}^{T_2}\left(Y_{t-1}^0/\sigma_0(\tau)\right)^2 \to_p 1$, *and*

(c) $\left(1-\rho_{0,n}^2\right)^{1/2}(nh)^{-1/2}\sum_{t=T_1}^{T_2} Y_{t-1}^0\sigma_t U_t/\sigma_0^2(\tau) \to_d N(0,1)$,

where Y_t^0 is defined in (7.5).

Lemma 7.6 concerns the zero-initial condition process Y_t^0 with time-varying autoregressive parameter ρ_t, which is the foundation of our asymptotic analysis of $\widehat{\rho}_{n\tau}$. Essentially Lemma 7.6 says that when $nh/b_n \to r_0 = \infty$, the asymptotic distributions of normalized sums of Y_t^0 behave in the same way as in an AR process with constant $\rho < 1$. The proof involves applications of appropriate weak laws of large numbers and central limit theorems for martingale difference triangular arrays and approximations of TVPs. One can extend the results to allow the U_t process to be conditionally heteroskedastic, but this requires a different definition of $\widehat{\sigma}_{n\tau}^2$ and complicates the analysis, see Andrews and Guggenberger (2014). For simplicity, we do not consider this extension here.

Define

$$\overline{\mu}_{nh} = (nh)^{-1}\sum_{t=T_1}^{T_2}\mu_t \text{ and } \overline{\mu}_{nh,-1} = (nh)^{-1}\sum_{t=T_1}^{T_2}\mu_{t-1}. \tag{7.21}$$

Lemma 7.7 (**Asymptotics of the Denominator in the Stationary Case**). *Under the null hypothesis $H_0 : \rho_{n\tau} = \rho_{0,n}$ and Assumptions 1, 2, and 3, for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$ where $nh/b_n \to r_0 = \infty$ and $\kappa_0(\tau) > 0$ (stationary case), the following results hold*

(a) $\left(1-\rho_{0,n}^2\right)(nh)^{-1}\sum_{t=T_1}^{T_2}\left(Y_{t-1}-\overline{\mu}_{nh,-1}\right)^2/\sigma_0^2(\tau) \to_p 1$, *and*

(b) $\left(1-\rho_{0,n}^2\right)\left(\overline{Y}_{nh,-1}-\overline{\mu}_{nh,-1}\right)^2/\sigma_0^2(\tau) \to_p 0$.

Lemma 7.7 concerns the asymptotic distribution of the denominator of the normalized $\widehat{\rho}_{n\tau}$ in (7.20). To bound the difference between Y_t^0 and Y_t and control for TVPs asymptotically, we use various inequalities and approximations in the proof of Lemma 7.7.

Lemma 7.8 (**Asymptotics of the Numerator in the Stationary Case**). *Under the null hypothesis $H_0 : \rho_{n\tau} = \rho_{0,n}$ and Assumptions 1, 2, and 3, for a sequence $\{\lambda_n = (\rho_n, \mu_n, \sigma_n^2, \kappa_n, b_n, F_n) \in \Lambda_n\}_{n\geq 1}$ where $nh/b_n \to r_0 = \infty$ and $\kappa_0(\tau) > 0$ (stationary case), the following results hold*

(a) $\left(1-\rho_{0,n}^2\right)^{1/2}(nh)^{-1/2}\sum_{t=T_1}^{T_2}\left(Y_{t-1}-\overline{\mu}_{nh,-1}\right)\left[\ Y_t-\overline{\mu}_{nh}-\rho_{0,n}\left(Y_{t-1}-\overline{\mu}_{nh,-1}\right)\ \right]/\sigma_0^2(\tau)$
$\to_d N(0,1)$, *and*

(b) $\left(1-\rho_{0,n}^2\right)^{1/2}(nh)^{-1/2}\sum_{t=T_1}^{T_2}\left(\overline{Y}_{nh,-1}-\overline{\mu}_{nh,-1}\right)\left[\ Y_t-\overline{\mu}_{nh}-\rho_{0,n}\left(Y_{t-1}-\overline{\mu}_{nh,-1}\right)\ \right]/\sigma_0^2(\tau)$
$\to_p 0$.

Lemma 7.8 concerns the asymptotic distribution of the numerator of the normalized $\widehat{\rho}_{n\tau}$ in (7.20). The proof of this lemma uses the results in Lemmas 7.2(c) and 7.6.

The next theorem provides the limit distributions of $\left(1-\rho_{0,n}^2\right)^{-1/2}(nh)^{1/2}(\widehat{\rho}_{n\tau}-\rho_{0,n})$ and $T_n(\rho_{0,n})$ in the stationary $\rho_{n\tau}$ case.

Theorem 7.3 (**Asymptotic Distribution of Normalized $\widehat{\boldsymbol{\rho}}_{\boldsymbol{n\tau}}$ and $\boldsymbol{t}$-Statistic in the Stationary Case**). *Under the null hypothesis* $H_0:\rho_{n\tau}=\rho_{0,n}$ *and Assumptions 1, 2, and 3, for a sequence* $\{\lambda_n=(\rho_n,\mu_n,\sigma_n^2,\kappa_n,b_n,F_n)\in\Lambda_n\}_{n\geq 1}$ *where* $nh/b_n\to r_0=\infty$ *and* $\kappa_0(\tau)>0$ (*stationary case*), *we have*

$$\left(1-\rho_{0,n}^2\right)^{-1/2}(nh)^{1/2}(\widehat{\rho}_{n\tau}-\rho_{0,n})\to_d N(0,1)$$

and

$$T_n(\rho_{0,n})\to_d N(0,1).$$

Remark 7.3. For any subsequence $\{p_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Lemmas 7.6–7.8 and Theorem 7.3 hold with p_n in place of n throughout and h_{p_n} in place of $h=h_n$.

7.6 Asymptotic Results for $\widehat{h}$

Here we give conditions under which $\widehat{h}$ defined in (4.2) is asymptotically equivalent to the value $\widehat{h}_{opt}$ that minimizes the "empirical loss," which is unobserved. See Li (1987) and Andrews (1991) for analogous results in i.i.d. models. The empirical loss, $L_n(h)$, is

$$L_n(h):=n^{-1}\sum_{t=1}^{n}(\widehat{\mu}_{t-1}(h)-\mu_t+Y_{t-1}(\widehat{\rho}_{t-1}(h)-\rho_t))^2. \tag{7.22}$$

We give a simple high-level condition under which

$$\frac{L_n(\widehat{h})}{L_n(\widehat{h}_{opt})}=\frac{L_n(\widehat{h})}{\inf_{h\in\mathcal{H}_n}L_n(h)}\to_p 1. \tag{7.23}$$

The $FE_n(h)$ criterion can be written as follows:

$$FE_n(h) = L_n(h) - 2C_n(h) + E_n, \text{ where } E_n := n^{-1}\sum_{t=1}^{n}\sigma_t^2 U_t^2 \text{ and}$$
$$C_n(h) := n^{-1}\sum_{t=1}^{n}\sigma_t U_t(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t)). \tag{7.24}$$

The empirical loss $L_n(h)$ and the cross-product term $C_n(h)$ depend on h, but the average squared error E_n does not. Since E_n does not depend on h, $\widehat{h}$ minimizes $L_n(h) - 2C_n(h)$ over $\mathcal{H}_n$.

Under the following condition, $\widehat{h}$ is asymptotically equivalent to the infeasible value $\widehat{h}_{opt}$ that minimizes $L_n(h)$.

Assumption 5. $\sup_{h\in\mathcal{H}_n}\frac{|C_n(h)|}{L_n(h)} \to_p 0$.

Lemma 7.9. *Under* Assumption 5, $\frac{L_n(\widehat{h})}{L_n(\widehat{h}_{opt})} \to_p 1$.

Now, we give a set of sufficient conditions for Assumption 5. Let $h_{\min} = \min\{h : h \in \mathcal{H}_n\}$ and $h_{\max} = \max\{h : h \in \mathcal{H}_n\}$. Typically, $h_{\min}$ and $h_{\max}$ depend on n and decrease to 0 as $n \to \infty$. Let ξ_n denote the cardinality of $\mathcal{H}_n$. We decompose $L_n(h)$ and $C_n(h)$ into the main components $L_{2n}(h)$ and $C_{2n}(h)$, respectively, which depend on $t = nh_{\max} + 1, ..., n$ and for which $(\widehat{\mu}_{t-1}(h), \widehat{\rho}_{t-1}(h))$ depend only on random variables in $\mathcal{G}_{t-1}$, and the "small t" boundary components $L_{1n}(h)$ and $C_{1n}(h)$, respectively, which depend on $t = 1, ..., nh_{\max}$ for which $(\widehat{\mu}_{t-1}(h), \widehat{\rho}_{t-1}(h))$ depend on some random variables that are not in $\mathcal{G}_{t-1}$. Define

$$L_{1n}(h) := n^{-1}\sum_{t=1}^{nh_{\max}}(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t))^2,$$
$$L_{2n}(h) := n^{-1}\sum_{t=nh_{\max}+1}^{n}(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t))^2,$$
$$C_{1n}(h) := n^{-1}\sum_{t=1}^{nh_{\max}}\sigma_t U_t(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t)), \text{ and}$$
$$C_{2n}(h) := n^{-1}\sum_{t=nh_{\max}+1}^{n}\sigma_t U_t(\widehat{\mu}_{t-1}(h) - \mu_t + Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t)). \tag{7.25}$$

The risk as a function of h is $R_n(h) = EL_n(h)$. Let $R_{2n}(h) = EL_{2n}(h)$.

The following assumption is sufficient for Assumption 5.

Assumption 6 (Sufficient Conditions for Assumption 5).

(a) The true sequence of distributions is from the parameter spaces $\{\Lambda_n : n \geq 1\}$.

(b) $\sup_{h\in\mathcal{H}_n} \frac{|C_{1n}(h)|}{L_n(h)} \to_p 0$.

(c) $\sup_{h\in\mathcal{H}_n} |\frac{L_{2n}(h)}{R_{2n}(h)} - 1| \to_p 0$.

(d) $h_{\max} \leq 1 - \varepsilon$ for n large for some $\varepsilon > 0$.

(e) $\frac{n \inf_{h\in\mathcal{H}_n} R_{2n}(h)}{\xi_n} \to \infty$.

Assumption 6(b) implies that the "small t" boundary component of $C_n(h)$, which only depends on $t \leq nh_{\max}$, is asymptotically dominated by $L_n(h)$, which is based on all $t \leq n$. Assumption 6(c) requires that the variability of $L_{2n}(h)$ is small relative to its mean. In particular, Assumption 6(c) holds if $StdDev(L_{2n}(h))/EL_{2n}(h) = o(1)$. Assumption 6(d) implies that the elements of $\mathcal{H}_n$ are bounded away from one, which implies that $nh = n$ is not a feasible choice.

To interpret Assumption 6(e), we give an intuitive discussion of the magnitudes of $\inf_{h\in\mathcal{H}_n} L_{2n}(h)$ and $\inf_{h\in\mathcal{H}_n} R_{2n}(h)$. First, consider the case where $\{Y_t\}_{t\leq n}$ displays unit root or local-to-unity behavior across the whole time series. Then, $\widehat{\mu}_{t-1}(h) - \mu_t \approx (nh)^{-1/2}$ and $Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t) \approx n^{1/2}(nh)^{-1} = (nh^2)^{-1/2}$, $L_{2n}(h) \approx (nh^2)^{-1}$, and under suitable conditions, $R_{2n}(h) \approx (nh^2)^{-1}$. Second, in the case where $\{Y_t\}_{t\leq n}$ displays stationary behavior across the whole time series, $\widehat{\mu}_{t-1}(h) - \mu_t \approx (nh)^{-1/2}$, $Y_{t-1}(\widehat{\rho}_{t-1}(h) - \rho_t) \approx (nh)^{-1/2}$, $L_{2n}(h) \approx (nh)^{-1}$, and under suitable conditions $R_{2n}(h) \approx (nh)^{-1}$. Third, in the case where $\{Y_t\}_{t\leq n}$ displays behavior that varies between unit root and stationarity across the time series, the order of magnitude of $R_{2n}(h)$ is between $(nh^2)^{-1}$ and $(nh)^{-1}$, which is bounded below by $(nh_{\max})^{-1}$. Hence, Assumption 6(e) requires $n\cdot(nh_{\max})^{-1}/\xi_n \to \infty$ or $\xi_n = o(h_{\max}^{-1})$. That is, the number ξ_n of values h in $\mathcal{H}_n$ needs to be of smaller order than the reciprocal of the maximum value $h_{\max}$ in $\mathcal{H}_n$.

Lemma 7.10. Assumption *6 implies* Assumption *5*.

Remark 7.4. The proof of Lemma 7.10 shows that $EC_{2n}(h) = 0$ and $Var(C_{2n}(h)) \leq C_{3U} \times (n - nh)^{-1} EL_{2n}(h)$ $\forall h \in \mathcal{H}_n, n \geq 1$, where $C_{3U} < \infty$ is the bound on the variance function σ^2 in the definition of Λ_n.

Remark 7.5. Let $R_{1n}(h) = EL_{1n}(h)$. Under Assumption 6 and $\sup_{h\in\mathcal{H}_n} \frac{R_{1n}(h)+E|C_{1n}(h)|}{R_n(h)} \to 0$, we also have: $\frac{R_n(\widehat{h})}{R_n(\widehat{h}_{opt})} \to_p 1$.

References

Andrade, Philippe, Gaetano Gaballo, Eric Mengus, and Benoit Mojon (2019), "Forward guidance and heterogeneous beliefs." *American Economic Journal: Macroeconomics*, 11 (3), 1–29.

Andrews, Donald W. K. (1991), "Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors." *Journal of Econometrics*, 47 (2-3), 359–377.

Andrews, Donald W. K. (1993), "Exactly median-unbiased estimation of first order autoregressive/unit root models." *Econometrica*, 61 (1), 139–165.

Andrews, Donald W. K., Xu Cheng, and Patrik Guggenberger (2020), "Generic results for establishing the asymptotic size of confidence sets and tests." *Journal of Econometrics*, 218 (2), 496–531.

Andrews, Donald W. K. and Patrik Guggenberger (2014), "A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter." *Review of Economics and Statistics*, 96 (2), 376–381.

Balassa, Bela (1964), "The purchasing-power parity doctrine: A reappraisal." *Journal of Political Economy*, 72 (6), 584–596.

Benigno, Gianluca (2004), "Real exchange rate persistence and monetary policy rules." *Journal of Monetary Economics*, 51 (3), 473–502.

Bernanke, Ben S. (2020), "The new tools of monetary policy." *American Economic Review*, 110 (4), 943–983.

Bykhovskaya, Anna and Peter C. B. Phillips (2018), "Boundary limit theory for functional local to unity regression." *Journal of Time Series Analysis*, 39 (4), 523–562.

Bykhovskaya, Anna and Peter C. B. Phillips (2020), "Point optimal testing with roots that are functionally local to unity." *Journal of Econometrics*, 219 (2), 231–259.

Cai, Zongwu, Jianqing Fan, and Qiwei Yao (2000), "Functional-coefficient regression models for nonlinear time series." *Journal of the American Statistical Association*, 95 (451), 941–956.

Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher, Alejandro Justiniano, Charles W. Calomiris, and Michael Woodford (2012), "Macroeconomic effects of federal reserve forward guidance." *Brookings Papers on Economic Activity*, 1–80.

Chari, Varadarajan V., Patrick J. Kehoe, and Ellen R. McGrattan (2002), "Can sticky price models generate volatile and persistent real exchange rates?" *The Review of Economic Studies*, 69 (3), 533–563.

Chung, Hess, Jean-Philippe Laforte, David Reifschneider, and John C. Williams (2012), "Have we underestimated the likelihood and severity of zero lower bound events?" *Journal of Money, Credit and Banking*, 44, 47–82.

Cogley, Timothy and Thomas J. Sargent (2001), "Evolving post-world war II U.S. inflation dynamics." *NBER Macroeconomics Annual*, 16, 331–373.

Cogley, Timothy and Thomas J. Sargent (2005), "Drifts and volatilities: Monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics*, 8 (2), 262–302.

Cole, Stephen, Enrique Martinez-Garcia, and Eric R. Sims (2023), "Living up to expectations: Central bank credibility, the effectiveness of forward guidance, and inflation dynamics post-global financial crisis." Working Paper.

Dahlhaus, Rainer (1996), "On the Kullback-Leibler information divergence of locally stationary processes." *Stochastic Processes and Their Applications*, 62 (1), 139–168.

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

Dahlhaus, Rainer and Liudas Giraitis (1998), "On the optimal segment length for parameter estimates for locally stationary time series." *Journal of Time Series Analysis*, 19 (6), 629–655.

Dahlhaus, Rainer, Michael H. Neumann, and Rainer Von Sachs (1999), "Nonlinear wavelet estimation of time-varying autoregressive processes." *Bernoulli*, 5 (5), 873–906.

Del Negro, Marco, Marc P. Giannoni, and Christina Patterson (2023), "The forward guidance puzzle." *Journal of Political Economy Macroeconomics*, 1 (1), 43–79.

Ding, Xin, Ziyi Qiu, and Xiaohui Chen (2017), "Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models." *Electronic Journal of Statistics*, 11 (2), 3871–3902.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984), "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, 3 (1), 1–100.

Eggertsson, Egauti B. and Michael Woodford (2003), "The zero bound on interest rates and optimal monetary policy." *Brookings Papers on Economic Activity*, (1), 139–233.

Elliott, Graham (1999), "Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution." *International Economic Review*, 40 (3), 767–784.

Elliott, Graham and James H. Stock (2001), "Confidence intervals for autoregressive coefficients near one." *Journal of Econometrics*, 103 (1-2), 155–181.

Engel, Charles (2014), "Exchange rates and interest parity." *Handbook of International Economics*, 4, 453–522.

Föllmer, Hans and Martin Schweizer (1993), "A microeconomic approach to diffusion models for stock prices." *Mathematical Finance*, 3 (1), 1–23.

Giraitis, Liudas, George Kapetanios, and Anthony Yates (2014), "Inference on stochastic time-varying coefficient models." *Journal of Econometrics*, 179 (1), 46–65.

Giraitis, Liudas, George Kapetanios, and Anthony Yates (2018), "Inference on multivariate heteroscedastic time varying random coefficient models." *Journal of Time Series Analysis*, 39, 129–149.

Grenier, Yves (1983), "Time-dependent ARMA modeling of nonstationary signals." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31 (4), 899–911.

Hansen, Bruce E. (1999), "The grid bootstrap and the autoregressive model." *Review of Economics and Statistics*, 81 (4), 594–607.

Imbs, Jean, Haroon Mumtaz, Morten O Ravn, and Helene Rey (2005), "PPP strikes back: Aggregation and the real exchange rate." *The Quarterly Journal of Economics*, 120 (1), 1–43.

Jordan, Thomas J., Michel Peytrignet, and Enzo Rossi (2010), "Ten years' experience with the Swiss National Bank's monetary policy strategy." *Swiss Journal of Economics and Statistics*, 146, 9–90.

Juhl, Ted (2005), "Functional-coefficient models under unit root behaviour." *The Econometrics Journal*, 8 (2), 197–213.

Karmakar, Sayar, Stefan Richter, and Wei Biao Wu (2022), "Simultaneous inference for time-varying models." *Journal of Econometrics*, 227 (2), 408–428.

Li, Ker-Chau (1987), "Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: discrete index set." *Annals of Statistics*, 958–975.

Lieberman, Offer (2012), "A similarity-based approach to time-varying coefficient non-stationary autoregression." *Journal of Time Series Analysis*, 33 (3), 484–502.

Lieberman, Offer and Peter C. B. Phillips (2014), "Norming rates and limit theory for some time-varying coefficient autoregressions." *Journal of Time Series Analysis*, 35 (6), 592–623.

Lieberman, Offer and Peter C. B. Phillips (2017), "A multivariate stochastic unit root model with an application to derivative pricing." *Journal of Econometrics*, 196 (1), 99–110.

Lieberman, Offer and Peter C. B. Phillips (2018), "IV and GMM inference in endogenous stochastic unit root models." *Econometric Theory*, 34 (5), 1065–1100.

Mikusheva, Anna (2007), "Uniform inference in autoregressive models." *Econometrica*, 75 (5), 1411–1452.

Moulines, Eric, Pierre Priouret, and François Roueff (2005), "On recursive estimation for time varying autoregressive processes." *The Annals of Statistics*, 33 (6), 2610–2654.

Müller, Ulrich K. and Graham Elliott (2003), "Tests for unit roots and the initial condition." *Econometrica*, 71 (4), 1269–1286.

Nicholls, Desmond F. and Barry G. Quinn (1980), "The estimation of random coefficient autoregressive models. I." *Journal of Time Series Analysis*, 1 (1), 37–46.

Quinn, Barry G. and Desmond F. Nicholls (1981), "The estimation of random coefficient autogressive models. II." *Journal of Time Series Analysis*, 2 (3), 185–203.

Rogoff, Kenneth (1996), "The purchasing power parity puzzle." *Journal of Economic Literature*, 34 (2), 647–668.

Samuelson, Paul A. (1964), "Theoretical notes on trade problems." *The Review of Economics and Statistics*, 145–154.

Sims, Christopher A. (1992), "Interpreting the macroeconomic time series facts: The effects of monetary policy." *European Economic Review*, 36 (5), 975–1000.

Stock, James H. (1991), "Confidence intervals for the largest autoregressive root in US macroeconomic time series." *Journal of Monetary Economics*, 28 (3), 435–459.

Subba Rao, Tata (1970), "The fitting of non-stationary time-series models with time-dependent parameters." *Journal of the Royal Statistical Society. Series B (Methodological)*, 32 (2), 312–322.

Tao, Yubo, Peter C. B. Phillips, and Jun Yu (2019), "Random coefficient continuous systems: Testing for extreme sample path behavior." *Journal of Econometrics*, 209 (2), 208–237.

van Delft, Anne and Michael Eichler (2018), "Locally stationary functional time series." *Electronic Journal of Statistics*, 12 (1), 107–170.

Xu, Ke-Li and Peter C.B. Phillips (2008), "Adaptive estimation of autoregressive models with time-varying variances." *Journal of Econometrics*, 142 (1), 265–280.