Journal of Econometrics 235 (2023) 1447–1463

Contents lists available at ScienceDirect

Journal of Econometrics

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

The distribution of rolling regression estimators[☆]

Zongwu Cai [a], Ted Juhl [b,*]

[a] Department of Economics, University of Kansas, Lawrence, KS, 66045, USA
[b] School of Business, University of Kansas, Lawrence, KS 66045, USA

ARTICLE INFO

Article history:
Received 20 January 2021
Received in revised form 4 December 2022
Accepted 4 December 2022
Available online 17 December 2022

JEL classification:
C14
C22
C5

Keywords:
Parameter instability
Nonparametric estimation
Rolling regressions
Uniform confidence intervals
Nonstationary

ABSTRACT

We establish the asymptotic distribution for rolling linear regression models using various window widths. The limiting distribution depends on the width of the rolling window and on a "bias process" that is typically ignored in practice. Based on the asymptotic distribution, we tabulate critical values used to find uniform confidence intervals for the average values of regression parameters over the windows. We propose a corrected rolling regression technique that removes the bias process by rolling over smoothed parameter estimates. The procedure is illustrated using a series of Monte Carlo experiments. The paper includes an empirical example to show how the confidence bands suggest alternative conclusions about the persistence of inflation.

1. Introduction

Rolling regression is often employed in empirical studies as a method to characterize changing economic relationships over time. As a simple robustness check, regression parameters are estimated using some fraction of the data early in the sample. The fixed fraction is then "rolled" through the sample, so that the estimated regression parameters may vary over time. This intuitive procedure is one of conventional methods to examine the stability of statistical relationships over time. A cursory search should reveal that there are rolling regression routines written in languages or packages such as R, STATA, Matlab, RATS, Python, Eviews and Excel.

As part of the prototypical exercise of reporting rolling regression estimates, researchers often plot bands around the point estimates as a way to conduct a type of ocular inference about whether there are changes in the relationships over time. The regression bands are constructed using estimated standard errors from the regression parameters in the relevant time period for the rolling window. Then, these estimated standard errors are multiplied by critical values from the standard normal distribution. Recent papers using rolling regression with confidence bands include but not limited to the papers by Swanson and Williams (2014), Linnainmaa and Roberts (2018), Adrian et al. (2015), Blanchard and Gali (2009), Blanchard (2018), Georgiev et al. (2018), Jimeńez et al. (2017) and Løpez-Salido et al. (2017), among others.[1]

☆ We thank the Managing Editor (Professor Serena Ng), an Associate Editor and two anonymous referees for their comments and suggestions that improved significantly the content and presentation of the paper. We also thank Graham Elliott, Bruce Hansen and Ulrich Müller for their comments on earlier versions of this paper.

* Corresponding author.
E-mail addresses: caiz@ku.edu (Z. Cai), juhl@ku.edu (T. Juhl).

[1] Rolling regression is also used in forecasting in the paper by Clark and McCracken (2009) in a framework that allows for structural change in regression parameters.

https://doi.org/10.1016/j.jeconom.2022.12.001
0304-4076/

In this paper, we characterize the population parameters that rolling regression is attempting to estimate and find the distribution of the rolling estimator. We provide new critical values that are used to construct asymptotically correct confidence bands for the estimated function. As part of this exercise, we show that rolling regression contains a bias process that may inhibit inference about the true population parameters. We develop a new procedure to estimate rolling regression parameters that is not affected by the bias process.

The original idea of rolling regression is an intuitive one, in that one wants to use regression over different time intervals to examine how the relationship may have changed. The window width may appear to be ad-hoc, but it is often based on equating the window width with some number of observations that seems appropriate for time series estimation, or based on decades, or some other relevant time frame. However, the results of this paper suggest that rolling regression is a compromise of the usual bias variance tradeoff. In particular, one can obtain the usual parametric convergence rates for rolling regression estimates (rather than nonparametric ones), although with a different limiting distribution.

The remainder of the paper is structured as follows. In Section 2, we develop the model, assumptions and asymptotic distribution results for the existing rolling regression procedures. In Section 3, a new procedure is proposed to deal with the bias process. Our procedure is discussed. Section 4 provides Monte Carlo evidence for the competing procedures. An empirical example is treated in Section 5 to illustrate differences in results based on our techniques versus the traditional confidence bands for the persistence of inflation. Section 6 concludes the paper, together with the theoretical derivations relegated to Appendix.

2. Model and assumptions

2.1. Model

It is natural to ask what we are attempting to estimate by employing rolling regression. To answer this question, let us consider a standard regression model given by

$$y_t = x_t^\top \beta + \epsilon_t,$$

where β is a $p \times 1$ vector of parameters. Let λ be the fraction of the total sample of T observations that is used in the rolling sample of data. The rolling regression estimator uses the $[T\lambda]$ observations, where $[x]$ denotes the integer part of x, and each of the periods is indexed as r, so that we have

$$\hat{\beta}_\lambda(r) = \left(\frac{1}{T\lambda} \sum_{s=[rT-T\lambda+1]}^{[rT]} x_s x_s^\top \right)^{-1} \frac{1}{T\lambda} \sum_{s=[rT-T\lambda+1]}^{[rT]} x_s y_s.$$

If a constant coefficient regression model is correctly specified, then the rolling regression estimator would estimate the same parameter, β, in each of the sub-samples using $\lambda \times 100\%$ of the data. Such an exercise is not interesting as β is constant over time and rolling estimators should be inefficient relative to the full sample regression estimator. Now, consider the model

$$y_t = x_t^\top \beta\,(t/T) + \epsilon_t,$$

where $\beta(\cdot)$ is a $p \times 1$ vector of integrable functions defined in $[0, 1]$, so that the regression parameter changes over time. The data are assumed to become more dense around a point t as T increases, a device employed in many studies, beginning notably in Robinson (1989). Given that the regression parameter is potentially changing at each point in time, as T increases, we hope to estimate the "average" value of $\beta(t/T)$ in the rolling window. To be specific, define

$$\bar{\beta}_\lambda(r) = \frac{1}{\lambda} \int_{r-\lambda}^{r} \beta(u)du,$$

which is the population parameter of interest indexed by λ and r. This quantity represents the average of the coefficients at point r given a rolling window fraction λ.

Our first task is to ascertain whether rolling regression indeed provides a consistent estimate of the parameter $\bar{\beta}_\lambda(r)$. We make several assumptions about the data generating process for the results in this section. Our theory makes use of the characterizations of processes from Zhou and Wu (2010). First, we allow for time varying processes, a form of nonstationarity. In particular, given that we are attempting to estimate time-varying parameters, it is natural to allow for nonstationary processes. Changing values of $\beta(t/T)$ necessarily induces nonstationarity in y_t. Moreover, if the model is dynamic, then x_t is also nonstationary if $\beta(t/T)$ changes. To this end, consider processes depending on deterministic functions coupled with the identically independent distributed (iid) process υ_t. Let $\mathcal{F}_t = \sigma\{\dots, \upsilon_{t-1}, \upsilon_t\}$. The processes for the covariates and errors are given by

$$x_t = G\,(t/T, \mathcal{F}_t) \quad \text{and} \quad \epsilon_t = H\,(t/T, \mathcal{F}_t)\,,$$

where the functions G and H allow for nonstationary processes in that the moments may change over time. The index t is scaled by the sample size T so that the data is assumed to be observed more densely as we collect more observations. In particular, define the second moment matrix of the x_t process as

$$M(t) = E\left[G\,(t/T, \mathcal{F}_t)\, G\,(t/T, \mathcal{F}_t)^\top\right].$$

The process associated with $x_t\epsilon_t$ is denoted GH and the covariance matrix of this product process as

$$\Omega(t) = E\left[GH\left(t/T, \mathcal{F}_t\right) GH\left(t/T, \mathcal{F}_t\right)^\top\right].$$

Conventionally, econometricians characterize dependence in data in several ways; linear processes, α-mixing, β-mixing, etc. Recent papers by Cai (2007) and Chen and Hong (2012) both make use of β-mixing assumptions, but the data are assumed to be stationary in both papers. We follow Zhou and Wu (2010) and allow for nonstationary processes that might arise from dynamic models with time-varying parameters. To this end, let v_0' be an iid copy of the variable v_0 that is part of $\mathcal{F}_j$. Define $\mathcal{F}_j^* = \sigma\{\ldots, v_{-1}, v_0', v_1, \ldots, v_{j-1}, v_j\}$. We wish to characterize the dependence of a process by measuring the effects of a shock to the system. For the variable x_t, define

$$\delta_q(x, j) = \sup_t \left\{ \left\| G\left(t/T, \mathcal{F}_j\right) - G\left(t/T, \mathcal{F}_j^*\right) \right\|_q \right\},$$

which is a measure of the effect of a shock after j periods. Limiting the allowable dependence in a process amounts to specifying suitable rates of decay for $\delta_q(x, j)$ as j increases.

Characterizing the dependence in these processes over time is a separate issue from checking whether the process is stationary. One class of nonstationary processes is the unit root process, where the autoregressive parameter is related to both dependence of the process over time as well as whether the variance of the process is constant over time. In the framework of this paper, we allow the processes to be nonstationary in a way that is separate from the dependence over time. To this end, we denote a process G to be stochastically Lipschitz continuous if

$$\sup_{0\le s\le t\le T} \left\{ \|G\left(s/T, \mathcal{F}_0\right) - G\left(t/T, \mathcal{F}_0\right)\|_2 \right\} \le c_1 \parallel t - s|/T$$

for some constant $c_1 > 0$. Additionally, the following assumptions are needed.

Assumption 1. The true data generating process is given by

$$y_t = x_t^\top \beta\left(t/T\right) + \epsilon_t,$$

and $\beta(u)$ is Riemann integrable on [0, 1].

Assumption 2. The error process ϵ_t is a martingale difference with respect to $\mathcal{F}_t$, so that $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$.

Assumption 3. For $x_t\epsilon_t$, we have

$$N^{1/4+\gamma} \sum_{j=N}^{\infty} \delta_4(x\epsilon, j) < \infty$$

for some $\gamma > 0$ and for x_t^2, we have $\sum_{j=N}^{\infty} \delta_4(x, j) < \infty$.

Assumption 4. The functions G and GH are stochastically Lipschitz continuous processes.

Assumption 5. The smallest eigenvalue of $M(t)$ is bounded away from zero.

Assumption 1 allows for the regression parameters to vary over time, perhaps with discontinuities. The martingale difference assumption could be relaxed and would require a long-run variance estimator for inference procedures. In the present case, we can assume a dynamic model may be specified to remove any correlation in ϵ_t. Moreover, because of the Assumptions 3 and 4, we allow for dynamic models with changing parameters, so that we can assume the error process to be a martingale difference. In particular, autoregressive models with time varying coefficients are considered in the paper by Zhang and Wu (2012), which shows that, given (standard) conditions on the time varying roots of the characteristic function, the process is locally stationary and it characterizes the decay in dependence. These models are shown to satisfy similar conditions to the conditions in this paper. Assumption 5 is sufficient for the rolling regression estimator to exist in the limit for any of the rolling windows one might consider.

We now provide the limiting distribution of the rolling regression estimator with its proof given in Appendix.

Theorem 1. *Suppose that Assumptions 1–5 hold and that $r \in [\lambda, 1]$. Then,*

$$\sqrt{T}\left[\hat{\beta}_\lambda(r) - \bar{\beta}_\lambda(r) - B_T(r)\right] \Rightarrow \left(\int_{r-\lambda}^{r} M(s)\right)^{-1} \left[Q(r) - Q(r-\lambda)\right],$$

where $\Rightarrow$ indicates weak convergence, $Q(r)$ is a p dimensional Gaussian process with covariance $E[Q(r_1)Q(r_2)^\top] = \int_0^{\min(r_1,r_2)} \Omega(s)ds$, and

$$B_T(r) = \left(\sum_{s=[rT-T\lambda+1]}^{[rT]} x_s x_s^\top\right)^{-1} \sum_{s=[rT-T\lambda+1]}^{[rT]} x_s x_s^\top \left[\beta(s/T) - \frac{1}{T\lambda} \sum_{s=[rT-T\lambda]}^{[rT]} \beta(s/T)\right].$$

The result in Theorem 1 provides several insights toward the use of rolling regression. First, the limiting distribution involves $Q(r)$, a functional of Brownian motion. In the search for confidence bands of the "average" of the regression parameters, using critical values from a standard normal distribution is incorrect.[2] As we illustrate in Section 4, using standard normal critical values for confidence bands for the estimate of the average parameter vector, $\bar{\beta}_\lambda(r)$, should be too narrow, so that coverage probabilities are well below target levels. Intuitively, the smaller the rolling window, we expect that wider confidence bands are required. Such an effect stems from the new distribution based on the functional of $Q(r)$, but also of the scaling factor given by $\left(\int_{r-\lambda}^{r} M(s)\right)^{-1}$. As λ gets smaller, this scaling factor should increase since $M(s)$ is positive (definite).

In addition to the finding of the limiting distribution involving functionals of Brownian motion, the distribution is affected by a bias process denoted $B_T(r)$. If the parameter vector $\beta(s/T)$ is constant over the entire sample, then this process disappears. However, in such cases, rolling regression is not interesting. If $x_s x_s^\top$ is unrelated to $\beta(s/T)$, then, the process should have zero mean and hence, we can apply a functional central limit theorem to it. In this case, we can think of this term as generating an additional term in the variance process. If $x_s x_s^\top$ is related to $\beta(s/T)$, then, the bias process has nonzero mean. The intuition for the bias process is that regression over the rolling interval weights the data and hence, the parameter vector, with more weights to observations with larger values of $x_s x_s^\top$. If these quantities are related to the parameter vector, then we may find inconsistent estimates of the average, given that some parameter values are over-weighted. A related phenomenon arises for fixed effects regression in panel data models. In particular, Campello et al. (2019) show that if cross-sectional units are heterogeneous in slope, a bias may result if the heterogeneity is related to second moments of the regressors.

Theorem 1 allows for both x_t and ϵ_t to exhibit a form of nonstationary behavior, which complicates the limiting distribution of the rolling estimator. Also, Theorem 1 provides a simplification if x_t and ϵ_t are both stationary. First, the bias process $B_T(r)$ has mean zero. In addition, $M(s)$ and $\Omega(s)$ are constant, so that $\int_{r-\lambda}^{r} M(s) = \lambda M$, where $M = E(x_s x_s^\top)$ and $\int_0^{\min(r_1,r_2)} \Omega(s) = \min(r_1, r_2)\Omega$ with $\Omega = E(x_s x_s^\top \epsilon_s^2)$.

To illustrate the bias process, we consider a simple autoregressive model given by

$$y_t = \alpha(t/T) + \phi(t/T) y_{t-1} + \epsilon_t$$

with $\alpha(t/T) = k_1 \times t/T$ and $\phi(t/T) = \phi$, so that the autoregressive coefficient $\phi(t/T)$ is constant but the intercept term is changing over the sample. It is easy to show, regardless of the width of the rolling window, that the asymptotic bias for estimating the parameter ϕ is given by

$$k_1^2(1-\phi)(1-\phi^2)\left[k_1^2(1-\phi^2) + 12\sigma^2\right]^{-1},$$

where σ^2 is the variance of ϵ_t. To simplify further, if $\phi = 0$, bias is $k_1^2/\left[k_1^2 + 12\sigma^2\right]$. This simple case is much like the result of Perron (1989); an omitted (broken) trend in the data generating process causes a bias in the estimates of the autoregressive parameters. Even if there is no serial correlation, the estimated autoregressive parameter may be close to one. The larger the magnitude of the omitted trend, the more bias. In general, the bias process $B_T(r)$ should be larger in magnitude as there is more correlation in the second moment of x_t and the parameter vector $\beta(t/T)$.

2.2. Critical values

The limiting distribution is a functional of the process given by $Q(r)$. The usual (incorrect) procedure for constructing confidence bands is to calculate a standard error estimate for each of the sub-periods in the rolling regression and then, employ standard normal critical values. For a rolling regression indexed by r, we would use a variance estimator given by

$$\hat{V}\left(\hat{\beta}_\lambda(r)\right) = \left(\sum_{s=[T(r-\lambda)+1]}^{[Tr]} x_s x_s^\top\right)^{-1} \sum_{s=[T(r-\lambda)+1]}^{[Tr]} x_s x_s^\top \hat{\epsilon}_s^2 \left(\sum_{s=[T(r-\lambda)+1]}^{[Tr]} x_s x_s^\top\right)^{-1}.$$

If there is no bias process, it is easy to see that the standardized process converges to a Gaussian process, $\tilde{Q}(r)$ with covariance

$$E\left[\tilde{Q}(r_1)\tilde{Q}(r_2)^\top\right] = \left(\int_{r_1-\lambda}^{r_1} \Omega(s)\right)^{-1/2} \int_{r_2-\lambda}^{r_1} \Omega(s) \left(\int_{r_2-\lambda}^{r_2} \Omega(s)\right)^{-1/2}$$

for $r_1 < r_2$ (or 0 if $r_1 < r_2 - \lambda$). The process has variance I_k and the dependence arises from the nonzero covariance when $r_1 > r_2 - \lambda$.

[2] A large literature on testing for structural change with unknown change point illustrates the need for different critical values in hypothesis testing.

Table 1
Critical values.

λ	Confidence level		
	0.90	0.95	0.99
0.05	3.498	3.708	4.148
0.10	3.271	3.499	3.956
0.15	3.125	3.356	3.843
0.20	3.013	3.265	3.755
0.25	2.921	3.174	3.691
0.30	2.828	3.091	3.619
0.35	2.753	3.016	3.533
0.40	2.677	2.942	3.481
0.45	2.603	2.887	3.436
0.50	2.544	2.825	3.388
0.55	2.468	2.758	3.322
0.60	2.404	2.695	3.260
0.65	2.343	2.636	3.214
0.70	2.283	2.578	3.176
0.75	2.212	2.512	3.115
0.80	2.154	2.459	3.059
0.85	2.072	2.377	2.999
0.90	1.985	2.297	2.906
0.95	1.882	2.193	2.798

From the limiting distribution of rolling regression estimators, we see that if we are attempting to construct uniform confidence bands for $\bar{\beta}_\lambda(r)$, standard normal critical values are inappropriate. To this end, we want to find critical values θ_λ such that

$$P\left(\sup_{r\in[\lambda,1]}\left|\tilde{Q}(r)\right| \le \theta_\lambda\right) = 0.95.$$

If the variance process is constant over r, so that $\Omega(s) = \Omega$, then the distribution of $\tilde{Q}(r)$ is a function of standard Brownian motion, adjusted for different values of the fraction of data used in rolling, λ.[3] To illustrate the distribution under this case when $p = 1$, we simulate critical values by generating random variables u_t from a standard normal distribution. Then, the standard Brownian motion $W(r)$ is simulated with $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_t$ where $T = 10{,}000$. For values of λ, find $\sup_{r\in[\lambda,1]} \frac{1}{\sqrt{\lambda}}|W(r) - W(r-\lambda)|$ and repeat 100,000 times for values of λ from 0.05 to 0.95. The resulting critical values are analogous to those calculated in Andrews (1993) for tests of structural change in regression models. The 0.90, 0.95 and 0.95 quantiles of empirical distribution are provided in Table 1.[4]

The width of the confidence bands increases as the rolling window becomes smaller. For example, if we use 10% of the sample in the rolling window, a confidence band is 3.499 times the relevant standard error for 95% confidence. The factor for a 20% window is 3.265, while an 80% window width increases the factor to 2.377. If one increases the window width to 100% of the sample, we obtain the usual 1.96. These are the critical values that are appropriate for confidence bands of rolling regressions in the absence of the bias process. At a minimum, the table provides values to replace critical values from the standard normal tables when applying rolling regression.

In the more realistic case of a time varying variance process governed by $\Omega(s)$, critical values should depend on $\Omega(s)$ which enters in the Gaussian process. We follow Hansen (1996) and simulate the limiting distribution of the process so that we have the proper critical values for the process associated with $\Omega(s)$. To this end, define the following estimators,

$$\hat{\epsilon}_s = y_s - x_s^\top \hat{\beta}_\lambda(s/T)$$

and let v_s be a generated standard normal variable. Then, we can simulate the process $\tilde{Q}(r)$ via

$$\tilde{U}(r) = \left(\frac{1}{T}\sum_{s=[T(r-\lambda)+1]}^{[Tr]} x_s x_s^\top \hat{\epsilon}_s^2\right)^{-1/2} \frac{1}{\sqrt{T}}\sum_{s=[T(r-\lambda)+1]}^{[Tr]} x_s \hat{\epsilon}_s v_s.$$

The simulated (standardized) process should have the same covariance structure as the limiting distribution in $\tilde{Q}(r)$, so that we can construct bands using the maximum of the absolute value of the appropriate entry of the process, in combination with the element from the variance matrix to obtain standard errors. The procedure can be applied under the

[3] A similar distribution results in the work of Giacomini and Rossi (2010) for comparing out of sample forecasts. We thank a referee for suggesting this connection.

[4] Tables with finer gradations of λ are available from the authors upon request. Interpolation between given values of λ can be used as an approximation for the appropriate critical values.

assumption that the bias process is zero, as would be the case if second moments of x_t are unrelated to the time-varying regression parameters.

3. Estimating average slope

Given that the goal of rolling regression appears to be estimating the average slope of the regression function as we move the window through time, we propose to estimate that quantity directly. That is, we seek to estimate the integral of the regression coefficient over the subintervals of the sample data. Our intent in the direct estimation of $\hat{\beta}_\lambda(r)$ is to avoid the potential bias process given by $B_T(r)$.

The local linear (nonparametric) estimator for $\beta(t/T)$, proposed by Cai (2007) for stationary α-mixing data, is given by $\tilde{\beta}(t/T)$,

$$\tilde{\beta}(t/T) = \begin{pmatrix} I_p & 0_p \end{pmatrix} \begin{pmatrix} S_{T,0} & S_{T,1}^\top \\ S_{T,1} & S_{T,2} \end{pmatrix}^{-1} \begin{pmatrix} V_{T,0} \\ V_{T,1} \end{pmatrix},$$

where for $\ell = 0$, 1 and 2,

$$S_{T,\ell}(t/T) = \frac{1}{Th} \sum_{s=1}^{T} x_s x_s^\top \left(\frac{t-s}{Th}\right)^\ell K\left(\frac{t-s}{Th}\right),$$

$K(\cdot)$ is a kernel function, h is the bandwidth, satisfying that $h \to 0$ and $T\,h \to \infty$, and for $\ell = 0$ and 1,

$$V_{T,\ell}(t/T) = \frac{1}{Th} \sum_{s=1}^{T} x_s y_s \left(\frac{t-s}{Th}\right)^\ell K\left(\frac{t-s}{Th}\right).$$

Then, for stationary α-mixing $\{x_t\}$, Cai (2007) shows that the nonparametric estimator is $\sqrt{Th}$-consistent and point-wise normally distributed. Furthermore, Zhou and Wu (2010) extend the work of Johnston (1982) to include uniform confidence bands in a time series setting for the function over the entire range of the data. The uniform confidence bands converge at an even slower rate than the usual nonparametric estimators. In addition, there is an additional bias term which depends on the second derivative of $\beta(t/T)$ at each point, which adds another obstacle to the construction of confidence bands.[5]

Now, we propose following estimator for $\bar{\beta}_\lambda(r)$;

$$\hat{\beta}^*_\lambda(r) = \frac{1}{[Tr]} \sum_{s=[rT-T\lambda+1]}^{T\lambda} \tilde{\beta}\,(s/T)\,,$$

which is the smoothed rolling (SR) estimator. The intuition for our estimator is that we hope to combine each estimator of $\beta(s/T)$ for the relevant range. By choosing an appropriate bandwidth parameter, we can eliminate the bias process. Next, we list three more assumptions for the data generating process and the bandwidth as well as the kernel function.

Assumption 6. The function $\beta(s)$ has three continuous derivatives.

Assumption 7. The bandwidth is chosen such that $h = c_2 T^{-\delta}$ with $1/4 < \delta < 1/3$ for some $c_2 > 0$.

Assumption 8. The kernel function $K(z)$ is a second-order kernel and takes the value 0 outside of $[-1, 1]$.

Assumption 6 imposes smoothness conditions on the behavior of the regression coefficients over time. Since we are estimating the average of the coefficients, this assumption simplifies the results. Assumption 7 is required so that the estimate of the average slope should converge at the usual parametric rate, which is exactly the same as Assumption B3 in Cai and Xiao (2012) for a semiparametric model,[6] and Assumption 8 is useful to limit the observations that enter the smoothed estimators. The Epanechnikov kernel, $K(z) = 0.75(1 - z^2)I(|z| \le 1)$, is a popular example of such a kernel and indeed, we employ this kernel in our Monte Carlo and empirical examples. We state the theorem for the modified procedure below with its detailed proof provided in Appendix.

Theorem 2. *Suppose that Assumptions 1–8 hold with $r \in (\lambda, 1)$. Then,*

$$\sqrt{T}\left[\hat{\beta}^*_\lambda(r) - \bar{\beta}_\lambda(r)\right] \Rightarrow \left[Q_2(r) - Q_2(r-\lambda)\right],$$

[5] One could employ other estimators for the regression coefficients at each point. Each method introduces its own challenges and advantages. Uniform convergence rates for splines are provided in Wang and Yang (2009) and uniform rates for series estimators are derived in Belloni et al. (2015) and Li and Liao (2020) for series estimators. An alternative framework such as Müller and Petalas (2010) may be employed in future studies.

[6] Clearly, Assumption 7 is about the under-smoothing at the first step for the nonparametric estimate and it is slightly stronger than $T\,h^4 \to 0$. As for how to choose h in practice, the reader is referred to the paper by Cai and Xiao (2012) for details.

where $Q_2(r)$ is a p dimensional Gaussian process with covariance matrix $E\left[Q_2(r_1)Q_2(r_2)^\top\right] = \int_0^{\min(r_1,r_2)} \Lambda(s)$ and

$$\Lambda(s) = \frac{1}{\lambda^2} M(s)^{-1}\Omega(s)M(s)^{-1}.$$

Theorem 2 implies that the new estimator is $\sqrt{T}$-consistent for estimating $\bar{\beta}_\lambda(r)$. Moreover, like the naive rolling estimator, the limiting distribution of our new statistic is also a function of $Q(r)$. Hence, we can employ the generated critical values to construct uniform confidence bands for $\bar{\beta}_\lambda(r)$. This rolling average smoothed estimator can be viewed as a bias corrected version of rolling regression and the limiting distribution still involves a similar Gaussian process. The bias is absent because we are directly estimating the average of the parameter values through the two step process. We first estimate the time-varying parameters directly via local linear estimation and then, we average those estimates. Similar results are often obtained in semiparametric models involving averages of nonparametric estimates; see, for instance, the paper by Cai and Xiao (2012) for details. The advantage of this procedure is that the averaging operation provides a faster rate of convergence that does not depend on the nonparametric bandwidth rate. In this way, the rolling estimator provides a computationally tractable procedure with the same interpretation as the traditional rolling regression procedure. However, our method now has the correct uniform size, where the traditional rolling procedure would have bands that are too narrow, resulting in incorrect coverage.

As in the case of Theorem 1, the above result also highlights the effect of using a smaller fraction of data in each of the rolling windows. Notice that as λ decreases, the variance of the Gaussian process at each point increases. This is intuitive, as we are effectively decreasing the sample size and increasing the variance. In the limit as $\lambda \to 0$, we obtain the nonparametric estimator at each point and the convergence rate is even slower than the nonparametric rate, as the results of Zhou and Wu (2010) apply. That is; as $\lambda \to 0$, the variance of the Gaussian process becomes infinite and the limiting distribution no longer converges at the parametric rate of $\sqrt{T}$. Theorem 2 highlights the challenging transition from estimating the average parameter over the rolling window to estimating the time varying parameter at a point.

Construction of the appropriate confidence bands is similar to the method used in Zhang and Wu (2012) along with our tabulated critical values. We estimate the standard errors of the modified rolling estimators. To this end, let

$$\tilde{S}(s) = \begin{pmatrix} S_{T,0}(s) & S_{T,1}^\top(s) \\ S_{T,1}(s) & S_{T,2}(s) \end{pmatrix} \quad \text{and} \quad \tilde{\Omega}(s) = \begin{pmatrix} \tilde{\Omega}_0(s) & \tilde{\Omega}_1(s) \\ \tilde{\Omega}_1(s) & \tilde{\Omega}_2(s) \end{pmatrix},$$

where with $\tilde{\epsilon}_r = y_r - x_r^\top \tilde{\beta}(r)$,

$$\tilde{\Omega}_\ell(s) = \frac{1}{Th}\sum_{r=1}^{T} x_r x_r^\top \tilde{\epsilon}_r^2 \left(\frac{s-r}{Th}\right)^\ell K\left(\frac{s-r}{Th}\right),$$

so that the standard errors are estimated via the variance matrix as

$$\frac{1}{T^2\lambda^2}\sum_{s=[T(r-\lambda)]}^{[Tr]} \left[I_k \vdots \quad 0_k\right] \tilde{S}(s)^{-1}\tilde{\Omega}(s)\tilde{S}(s)^{-1}\left[I_k \vdots \quad 0_k\right]^\top = \frac{1}{T^2}\sum_{s=[T(r-\lambda)]}^{[Tr]} \tilde{\Lambda}(s).$$

The standardized estimator is governed by the process $\tilde{Q}_2(r)$, which is Gaussian with covariance process

$$E\left[\tilde{Q}_2(r_1)\tilde{Q}_2(r_2)^\top\right] = \left(\int_{r_1-\lambda}^{r_1} \Lambda(s)\right)^{-1/2} \int_{r_2-\lambda}^{r_1} \Lambda(s) \left(\int_{r_2-\lambda}^{r_2} \Lambda(s)\right)^{-1/2}$$

for $r_1 < r_2$ (or 0 if $r_1 < r_2 - \lambda$). Finally, we note that the $B_T(r)$ process is removed from the limiting distribution due to the restrictions on the smoothing parameter h. In addition, if both x_t and ϵ_t are stationary, the distribution simplifies further since $\Omega(s) = E(x_s x_s^\top \epsilon_s^2)$ is constant.

Finally, we discuss on how to tabulate critical values. As in the case of the traditional rolling regression, the appropriate critical values for the rolling average slope estimator are not the usual values associated with the standard normal distribution. In the most restrictive case where $\Omega(s)$ and the second moment matrix of the regressors $M(s)$ are constant over s, we can employ the critical values appearing in Table 1. If $\Omega(s)$ and $M(s)$ are time-varying, we can use the estimated components of $\Lambda(s)$ so that we can simulate using

$$\tilde{U}_2(r) = \left(\frac{1}{T}\sum_{s=[T(r-\lambda)+1]}^{[Tr]} \tilde{\Lambda}(s)\right)^{-1/2} \frac{1}{\sqrt{T}}\sum_{s=[T(r-\lambda)+1]}^{[Tr]} \tilde{\Lambda}(s)^{1/2} v_s.$$

4. Monte Carlo simulation studies

The theorems from Sections 2 and 3 show theoretically that rolling regression estimators have a limiting distribution that depends on functionals of Brownian motion and those new critical values are provided in Table 1. The purpose of Theorem 2 is to provide a bias corrected estimate of the average slope. Now, we explore the empirical performance of

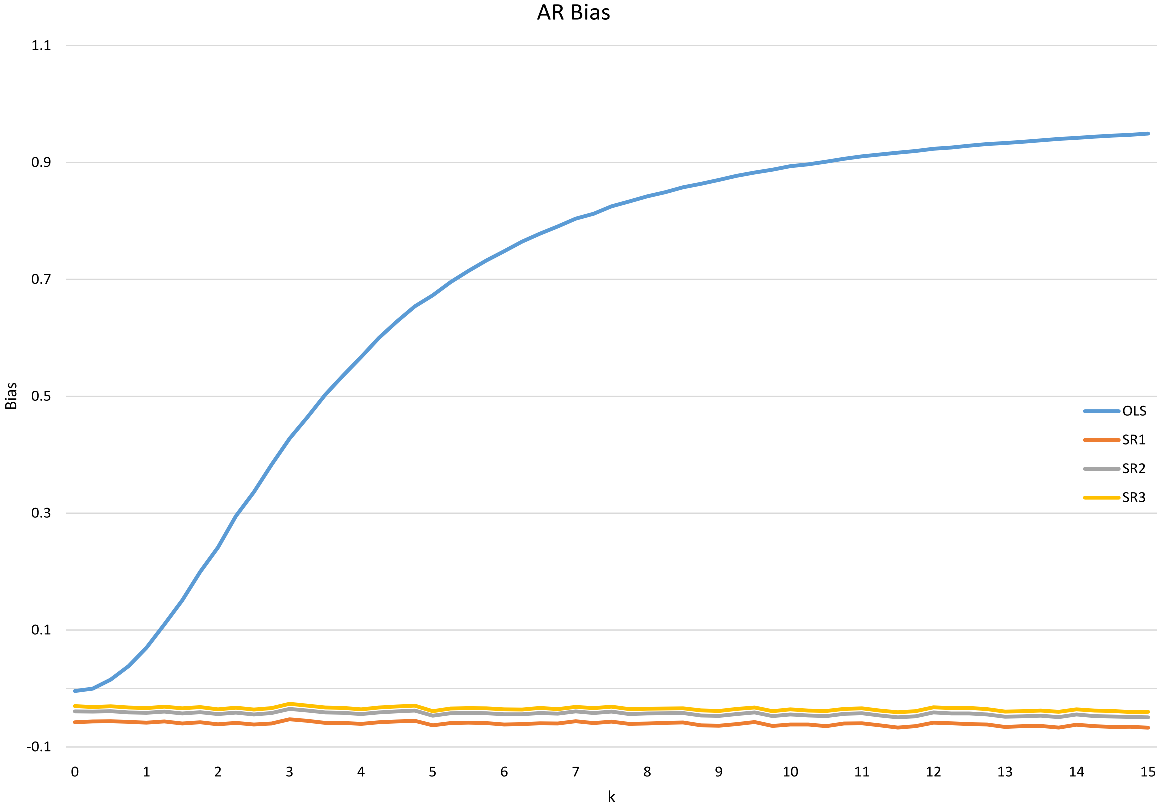

Fig. 1. The resulting bias versus k_1 for OLS and SR1, SR2, SR3.
Note: The data generating process is $y_t = k_1 t + \epsilon_t$, but the estimated model is $y_t = \alpha + \phi y_{t-1} + \epsilon_t$. The estimate of ϕ is labeled as OLS. The SR procedures estimate $y_t = \alpha(t/T) + \phi(t/T)y_{t-1} + \epsilon_t$ using bandwidth parameters $c_2 T^{-0.30}$ with c_2 = 0.75, 1.0, 1.25, respectively. The average value of the estimated $\phi(t/T)$ parameter for the different bandwidths corresponds to SR1, SR2 and SR3. The true value of the autoregressive parameter is zero, so that OLS, SR1, SR2 and SR3 measure the bias as a function of k_1.

various rolling regression estimators in this section. In particular, the naive rolling regression estimator that uses standard normal critical values is denoted by RO for rolling ordinary least squares (OLS) regression. A second estimator is considered but uses the new critical values and it is referred to as cRO. This rolling estimator accounts for $Q(r)$ in the limiting distribution, but does not correct for the possible bias process $B(r)$. Finally, we include three versions of the corrected rolling regression estimator proposed in Section 3. The estimator depends on a bandwidth parameter for the time varying parameter regression at the first stage. The form of the allowable bandwidth is $h = c_2 T^{-\delta}$, where $1/4 < \delta < 1/3$, which satisfies Assumption 7. We use $\delta = 0.30$ and set $c_2 = 0.75$, 1.0 and 1.25, and the corresponding estimators are denoted by SR1, SR2 and SR3, respectively.[7]

Before considering rolling regression, we illustrate the bias that one encounters if an autoregressive process has an omitted trend, which could be considered the case where $\lambda = 1$. To this end, we simulate the process discussed in Section 2 with a missing trend. The bias of the estimated parameter in an autoregressive (AR) model is indexed by k_1, the magnitude of the missing trend. We simulate an AR process with $\phi = 0$ and consider values of k_1 ranging from 0 to 15, with 200 and 1000 replications for each value of k_1. The resulting bias for OLS and the smoothed estimates (SR1, SR2 and SR3) is represented in Fig. 1, respectively. We note that there is a small negative bias for the smoothed regression estimators. However, it does not change as the magnitude of the missing trend grows, which illustrates that this procedure removes the bias process from the estimator. Moreover, the bias in OLS grows as the omitted trend indexed via k_1 grows larger, consistent with the results in Perron (1989).

Given the potential for bias in rolling estimators, we illustrate several data generating processes using rolling estimators. For each experiment, we allow for time series dimensions (sample size) T = 200, 400 and 600. The rolling window is set for $\lambda = 0.20$ so that there are 40, 80 and 120 observations used in each of the regimes in the estimation window. The number of replications for each experiment is set to 10,000. The parameter of interest is $\bar{\beta}_\lambda(r)$, the average

[7] From the simulation results in our Monte Carlo studies conducted in this section, one can see from Fig. 1 and Tables 2 and 3 that the differences in the finite sample performances for SR1, SR2 and SR3 are not so significant, so that the choice of h is not so sensitive as long as it satisfies Assumption 7. This observation is consistent with that in Cai and Xiao (2012).

value of $\beta(r)$ over each relevant subset of time. Our experiments report the estimated coverage probability for a 95% uniform confidence band for $\bar{\beta}_\lambda(r)$. In addition, we list the average width (denoted by Av. width in Tables 2 and 3) of the uniform confidence band. For example, if two competing procedures both have 95% coverage, we would prefer the method generating narrower bands.

Denote the current naive point-wise technology of using rolling OLS and employing critical values from the standard normal distribution as RO. If we use the adjusted critical values tabulated in Table 1 with rolled OLS. Our smoothed rolling procedure using the adjusted critical values is indexed by the bandwidths as SR1, SR2 and SR3.[8] In addition, we also consider conservative bands construction via a Bonferroni procedure, much like the methods suggested in Eubank and Speckman (1993). The procedure adjusts critical values according to the number of points that the rolling regression is estimated in a given sample, so that the critical values should change with T in our experiment. We list these results as Bon in Tables 2 and 3.[9]

Our first set of 4 experiments use a simple regression model and we attempt to estimate bands around the regression slope parameter. In these first four models, we examine static regression models and the bias process does not affect the distribution. Hence, we expect that the adjusted rolling procedure should be adequate to obtain improved coverage relative to rolling OLS that is currently employed in empirical research. The first data generating process is given by

$$y_t = 2x_t + \epsilon_t, \tag{M1}$$

where $x_t \sim N(1, 1)$ and $\epsilon_t \sim N(0, 0.25)$. Hence, in this model, the average parameter is constant throughout the sample. The results are depicted in Panel (a) of Table 2. We first examine the rolling OLS estimator using the (incorrect) standard normal distribution critical values. The coverage for the RO procedure is below the nominal 95% and ranges from around 19% to 27% depending on the sample size. The rolling OLS with adjusted critical values from Table 1 is much closer to the intended coverage, reaching as high as 95% when the sample size is 600. The smoothed rolling procedures have much better coverage than RO for all the considered bandwidth parameters. As expected, the Bonferroni procedures obtain their conservative coverage from their wide intervals. The smoothed procedures has the narrowest uniform bands (average width) of the procedures with adequate coverage.

Now, we consider a static regression model with a change in slope in the middle of the sample as

$$y_t = 2 + \beta(t/T)x_t + \epsilon_t, \tag{M2}$$

where $x_t \sim N(1, 1)$, $\epsilon_t \sim N(0, 0.25)$ and $\beta(t/T)$ changes halfway through the sample as $\beta(t/T) = 1$ if $t \le T/2$ and it is $3/2$ if $t > T/2$. The results appear in Panel (b) of Table 2. We see that the coverage is poor for rolling OLS, but improves as the sample size increases. The adjusted rolling regression has good coverage as the sample size increases, reaching as high as 91% for 600 (effectively $T = 120$ in each rolling window) and the smoothed procedures are not as accurate in coverage. Given that the bias process is not relevant, here, we recommend using cRO. Again, the Bonferroni procedure has good coverage, but wide intervals.

Next, by adding two more independent variables to the model in (M2), we consider the following data generating process

$$y_t = 2 + \beta(t/T)x_t + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t, \tag{M3}$$

where $\beta_2 = \beta_3 = 1$ and $x_{jt} \sim N(1, 1)$ for $j = 2$ and 3. We include this model to examine the properties of the statistics if the models have more parameters (4 to estimate), with a similarly structured data generating process. Our focus is on the confidence bands for $\bar{\beta}$. The results are shown in Panel (c) of Table 2. We note that the current standard RO again performs very poorly, but the adjusted critical values improve coverage while maintaining narrow intervals. The results are similar even with more parameters to estimate. The rolling regression procedure always performs poorly since it is a point-wise procedure. The proposed corrections are effective relative to rolling regression in this model with more parameters.

Finally, we simulate a random coefficient model where the intercept and slope are both random. In particular, we have

$$y_t = \beta_{0t} + \beta_{1t}x_t + \epsilon_t, \tag{M4}$$

where $\beta_{jt} = \beta_{j,t-1} + v_{jt}$ with $v_{jt} \sim N(0, 0.01)$ for $j = 0$ and 1, $x_t \sim N(1, 1)$ and $\epsilon_t \sim N(0, 0.25)$. The simulation results are displayed in Panel (d) of Table 2. The coefficients in this model are time varying and rolling OLS attempts to estimate the average slope coefficient over the rolling windows. The smoothed estimators are not expected to work well here, as the local averages involve parameters that are smooth as a function of t/T. That is, for the random coefficient model data generating process, our assumptions for Theorem 2 do not hold. The results again show that point-wise bands from rolling regression are not appropriate. The adjusted rolling regression procedure works as a tradeoff between coverage

[8] In addition to the procedures we evaluated in our Monte Carlo experiment, we also used simulated critical values to account for the possibility of time varying variances, which were present in the experiments. However, the case specific simulated critical values for the time varying variance models were similar to those in Table 1. We attempted to construct pathological examples of time varying variances but the results were similar to data generating processes with fixed variances.

[9] We thank a referee for suggesting this procedure as a baseline for comparison. These critical values are wider than our new values from Table 1 for all of our experiments.

Table 2
Regression models from model (M1) to model (M4).

Model		T	RO	cRO	SR1	SR2	SR3	Bon
(a) M1	Coverage	200	0.19	0.84	0.96	0.98	0.99	0.91
		400	0.24	0.90	0.96	0.98	0.99	0.97
		600	0.27	0.95	0.97	0.99	0.99	1.00
	Av. width	200	0.30	0.50	0.40	0.41	0.42	0.55
		400	0.22	0.36	0.28	0.29	0.29	0.42
		600	0.17	0.29	0.23	0.23	0.24	0.34
(b) M2	Coverage	200	0.18	0.83	0.91	0.91	0.88	0.90
		400	0.22	0.90	0.89	0.86	0.79	0.96
		600	0.26	0.91	0.87	0.82	0.71	0.98
	Av. width	200	0.32	0.53	0.41	0.43	0.44	0.58
		400	0.23	0.38	0.29	0.30	0.31	0.44
		600	0.19	0.31	0.24	0.24	0.25	0.37
(c) M3	Coverage	200	0.14	0.79	0.90	0.90	0.87	0.88
		400	0.20	0.88	0.89	0.86	0.78	0.96
		600	0.23	0.90	0.86	0.81	0.70	0.98
	Av. width	200	0.32	0.53	0.42	0.43	0.44	0.58
		400	0.23	0.38	0.29	0.30	0.31	0.44
		600	0.19	0.31	0.24	0.24	0.25	0.37
(d) M4	Coverage	200	0.14	0.81	0.85	0.81	0.75	0.89
		400	0.17	0.88	0.71	0.58	0.43	0.96
		600	0.19	0.90	0.60	0.40	0.25	0.97
	Av. width	200	0.48	0.80	0.55	0.60	0.64	0.88
		400	0.44	0.73	0.45	0.49	0.53	0.85
		600	0.43	0.71	0.40	0.44	0.48	0.84

Note: RO is the rolling OLS estimator, while cRO is rolling OLS but uses the critical values from Table 1. SR1, SR2 and SR3 are the smooth rolling estimators using bandwidths $c_2T^{-0.3}$ with $c_2 = 0.75$, 1.0 and 1.25, respectively, and using the critical values in Table 1. Rolling OLS with Bonferroni critical values is labeled as Bon. For data generating process (b), $\beta_1(t/T)$ is 1 in the first half of the sample and 1.5 in the second half. For case (c), we add two additional covariates to data generating process (b). Data generating process (d) is a random coefficient model for both parameters, where $v_{it} \sim N(0, 0.01)$. The coverage probabilities are for the rolling average $\beta_1(t/T)$ in the data with a 95% target.

and width of the interval. As expected, the smoothing procedures do not work as well and get worse as the sample size increases.

The next 6 experiments consider data generating processes with autoregressive models and we consider the uniform coverage of the estimated average of the autoregressive parameter. These models have the potential to generate the bias process of Theorem 1, so that the SR procedures should have the potential advantage. To begin, we start with a baseline model from a simple autoregressive model with no parameter changes. The data generating process is given by

$$y_t = 0.80y_{t-1} + \epsilon_t, \tag{M5}$$

where ϵ_t is $N(0, 0.25)$. The results are presented in Panel (a) of Table 3. Given the rolling fraction of $\lambda = 0.20$, the procedures are attempting to estimate the AR parameter based on 40 observations. The SR3 procedure is best for this data generating process, with narrow confidence bands, but the adjusted rolling OLS procedure performs respectably. Again, the naive traditional rolling OLS (point-wise bands) procedure performs poorly, with coverage of 17% even with a sample of 600 observations. The Bonferroni procedure has good coverage, but this is due to having the widest intervals in every case.

Now, we consider a model with changing autoregressive coefficient

$$y_t = \rho(t/T)y_{t-1} + \epsilon_t, \tag{M6}$$

where $\rho(t/T) = 0.8\,[1 - (t/T)]$ and ϵ_t is $N(0, 0.25)$, so that the process becomes less persistent throughout the sample. Each of the procedures attempts to estimate the average slope throughout the rolling interval. So, the average persistence is changing through the sample, with a steady decline. Our results are described in Panel (b) of Table 3. This experiment also highlights the advantages of the smoothed procedures. The adjusted rolling OLS estimator performs well once the sample size reaches 600 (effectively 120 with $\lambda = 0.2$), while the smoothed rolled estimators with larger bandwidths (SR2 and SR3) perform well for all sample sizes and have narrower confidence bands. The Bonferroni procedure provides adequate coverage at the largest sample size, but is almost 50% wider than the intervals from SR3.

For the next experiment, the autoregressive coefficients do not change, but there is an omitted trend variable. This type of data generating process may cause bias in the autoregressive parameter if one uses standard OLS, which is illustrated in

Table 3
Autoregressive models from model (M5) to model (M10).

Model		T	RO	cRO	SR1	SR2	SR3	Bon
(a) M5	Coverage	200	0.06	0.61	0.76	0.92	0.97	0.75
		400	0.13	0.80	0.79	0.92	0.97	0.92
		600	0.17	0.86	0.81	0.92	0.97	0.96
	Av. width	200	0.40	0.67	0.56	0.56	0.56	0.74
		400	0.28	0.46	0.38	0.38	0.38	0.53
		600	0.22	0.37	0.30	0.30	0.30	0.44
(b) M6	Coverage	200	0.15	0.77	0.89	0.96	0.99	0.86
		400	0.20	0.87	0.91	0.97	0.99	0.96
		600	0.22	0.90	0.92	0.96	0.99	0.96
	Av. width	200	0.53	0.89	0.70	0.72	0.74	0.98
		400	0.38	0.64	0.50	0.51	0.52	0.74
		600	0.31	0.52	0.41	0.41	0.42	0.62
(c) M7	Coverage	200	0.01	0.33	0.96	0.99	0.99	0.48
		400	0.00	0.15	0.95	0.98	0.99	0.37
		600	0.00	0.06	0.95	0.98	0.99	0.21
	Av. width	200	0.56	0.94	0.76	0.78	0.81	1.04
		400	0.41	0.69	0.55	0.56	0.57	0.80
		600	0.34	0.57	0.45	0.46	0.47	0.68
(d) M8 mean break	Coverage	200	0.12	0.73	0.96	0.99	0.99	0.83
		400	0.09	0.69	0.96	0.98	0.99	0.84
		600	0.06	0.59	0.95	0.97	0.98	0.79
	Av. width	200	0.60	0.96	0.76	0.79	0.81	1.05
		400	0.42	0.70	0.55	0.56	0.57	0.81
		600	0.35	0.58	0.45	0.46	0.47	0.69
(e) M9 persistence break	Coverage	200	0.09	0.69	0.80	0.91	0.95	0.80
		400	0.13	0.80	0.83	0.91	0.93	0.92
		600	0.14	0.81	0.83	0.90	0.91	0.94
	Av. width	200	0.47	0.79	0.64	0.65	0.66	0.87
		400	0.33	0.55	0.45	0.46	0.46	0.64
		600	0.27	0.45	0.36	0.36	0.37	0.54
(f) M10 mean & persistence break	Coverage	200	0.06	0.61	0.82	0.93	0.96	0.73
		400	0.06	0.55	0.85	0.93	0.94	0.71
		600	0.04	0.42	0.85	0.91	0.92	0.60
	Av. width	200	0.46	0.77	0.64	0.64	0.65	0.85
		400	0.32	0.54	0.44	0.44	0.45	0.63
		600	0.27	0.44	0.36	0.36	0.36	0.52

Note: RO is the rolling OLS estimator, while cRO is rolling OLS but uses the critical values from Table 1. SR1, SR2 and SR3 are the smooth rolling estimators using bandwidths $c_2T^{-0.3}$ with $c_2 = 0.75$, 1.0 and 1.25, respectively, and using the critical values in Table 1. Rolling OLS with Bonferroni critical values is labeled as Bon. Specification (d) has $\alpha(t/T)$ set to 1 in the first half of sample and 1.5 in the second half, while $\rho(t/T)$ is zero, so that it has a mean break. Data generating process (e) sets $\rho(t/T)$ equal to 0.80 in the first half of the sample and 0.40 in the second half with $\alpha(t/T) = 0$, so that it has a persistence break. Data generating process (f) modifies (e) by adding a change of $\alpha(t/T)$ from 1/10 in the first half to 5/6 in the second half of the sample, so that it has both mean and persistence breaks. The coverage probabilities are for the rolling average $\rho(t/T)$ in the data with a 95% target.

Fig. 1. Hence, we expect poor coverage for RO and cRO and the smoothed versions of the rolling estimators are designed for this type of data generating process. Then, we have the following

$$y_t = 4 \times (t/T) + \rho y_{t-1} + \epsilon_t, \tag{M7}$$

where ϵ_t is $N(0, 0.25)$ and the true value of ρ is set to zero, but the estimated model is $y_t = \alpha + \phi y_{t-1} + \epsilon_t$. The rolling estimator of ρ is the basis for comparison of the techniques. Also, Fig. 1 highlights the potential bias in such a data generating process and this is where we expect to see that the bias process $(B_T(r))$ has effects on the techniques. We report the coverage for the uniform bands for the average autoregressive coefficient in Panel (c) of Table 3. The biases in standard OLS are apparent in both the rolled OLS and adjusted rolled OLS based confidence bands with poor coverage for all sample sizes. The bias affects the coverage in adjusted rolling OLS, rolling OLS and Bonferroni procedures. This illustrates the same phenomenon as in Perron (1989) that an incorrectly specified deterministic component in a dynamic model is misconstrued as a more persistent time series model. The SR procedures are excellent for this data generating process, with excellent coverage and the narrowest bands.

Another data generating process with a structural change is

$$y_t = \alpha(t/T) + \rho(t/T)y_{t-1} + \epsilon_t, \tag{M8}$$

where ϵ_t is $N(0, 0.25)$, $\rho(t/T) = 0$ and $\alpha(t/T) = 1$ if $t \le T/2$ and it is 3/2 if $t > T/2$. We estimate the average slope parameter to see the effects of the bias process on the autoregressive parameter. The simulation results are displaced in Panel (d) of Table 3 with a label of "mean break". The rolling OLS and adjusted rolling procedure are both susceptible to the effects of a break in the intercept term. The Bonferroni procedure also gets worse as the sample size increases as the bias process dominates. Although the intercept change is not smooth, the SR procedures do an effective job of removing the bias process, so that their coverage is quite good for each of the bandwidth parameters. Coverage improves and the average width of the coverage intervals declines as the sample size increases. Both RO and cRO have poor coverage and get worse as the sample size increases.

The next process we simulate has a sudden break in the persistence, so that

$$y_t = \rho(t/T)y_{t-1} + \epsilon_t, \tag{M9}$$

where ϵ_t is $N(0, 0.25)$ and $\rho(t/T)$ is 0.80 in the first half of the sample and 0.40 in the second half of the sample. The experiment results are summarized in Panel (e) of Table 3 with a label of "persistence break". The SR methods provide good coverage, but the RO, cRO and Bonferroni procedures have poor performance. Moreover, the width of the intervals for SR2 and SR3 get much smaller as the sample size increases.

Finally, we take the same data generating process as in (M9), but add a change in the intercept,

$$y_t = \alpha(t/T) + \rho(t/T)y_{t-1} + \epsilon_t, \tag{M10}$$

where $\alpha(t/T)$ is 1/10 in the first part of the sample and 5/6 in the second half of the sample. Given the AR parameters, this process has a mean shift and a shift in persistence with mean of 0.5 that changes to 1.39. The results are reported in Panel (f) of Table 3 with a label of "mean & persistence break". This experiment again shows the effects of the bias process, so that rolling regression, critical value adjusted rolling regression and the Bonferroni procedure are all negatively affected. The SR2 and SR3 procedures exhibit robustness to the changes, with effective coverage and narrow bands.

In summary, the Monte Carlo experiments conducted above suggest that for static models, using the new critical values from Table 1 provide a simple and effective way to adjust the rolling estimators to obtain accurate confidence bands (the cRO procedure). However, from the theorems in Sections 2 and 3, we know that bias results when second moments of the regressors are related to the parameters in the model. A well-known case of this phenomenon is the time varying autoregressive model, which we also explore in the Monte Carlo experiments. We see that employing rolling regression with the use of the time-varying coefficient regression of Cai (2007) in the first stage removes the bias from rolling regression. Given that dynamic models with time varying parameters are particularly susceptible to the bias process, we suggest smoothed procedures for finding the average coefficients in dynamic models.[10]

5. An empirical illustration

We illustrate our proposed procedures along with rolling regression using intervals (point-wise) such as those that appear in applied work using rolling regression. Rolling regression was employed by O'Reilly and Whelan (2005) to examine the persistence of inflation over time in the Euro area. They find that despite many different policy regimes, the reduced form coefficient estimates appear to be stable in the Euro area, based on evidence from the rolling regression results and other tests. We consider a similar exercise for the U.S. using rolling regression and the smoothed regression procedure proposed in this paper. Suppose that one estimates an AR(4) model for the U.S. inflation given by

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \epsilon_t.$$

A common transformation that is used to evaluate persistence is

$$\Delta y_t = \alpha + \omega y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \gamma_3 \Delta y_{t-3} + \epsilon_t,$$

where $\Delta y_t - y_t - y_{t-1}$. If $\omega = \phi_1 + \phi_2 + \phi_3 + \phi_4 - 1$ is zero, the autoregressive process has a unit root which is an extreme form of persistence. If the process is stationary, ω is negative and the closer to -1, the less persistent is inflation. We construct the inflation series from data for CPI-U,[11] with monthly observations from February of 1947 to May of 2022. As a preliminary analysis, we estimate the AR(4) model and test for serial correlation in the residuals. We fail to reject

10 A recent paper by Inoue et al. (2021) estimates time-varying-parameter models along with GARCH effects. They note that it may be possible to establish uniform inference (over time) which would establish confidence bands. There are Bayesian models of time-varying-parameters that can estimate the function over the entire range as well. Both types of models might be used as the first stage estimates to include in rolling averages, but the distributional results need to be proven.

11 The index consists of all U.S. urban consumers with seasonal adjustment and the inflation series is calculated as the percent change from month to month. For details, please see the website https://www.bls.gov/help/one_screen/cu.htm.

the null of no serial correlation at the 5% level and we also reject the unit root hypothesis at the 1% level.[12] We proceed to analyze rolling estimates of the persistence parameter ω. Given our preliminary findings of no unit root, our rolling analysis is not intended to check for unit roots but to look for changing persistence.

Our analysis in this paper treats rolling regression and the variants we propose as a way to estimate "average" parameters over each of the windows in the rolling procedure. Hence, the model allows for time varying parameters in the underlying model. In particular, we have

$$\Delta y_t = \alpha(t/T) + \omega(t/T)y_{t-1} + \gamma_1(t/T)\Delta y_{t-1} + \gamma_2(t/T)\Delta y_{t-2} + \gamma_3(t/T)\Delta y_{t-3} + \epsilon_t.$$

This autoregressive based model has two main parameters in $\alpha(t/T)$ and $\omega(t/T)$. If ω (and the other parameters) are constant and $\alpha(t/T)$ increases, this increases the mean of the process. As the results in Perron (1989) show, changes in the constant term have the effect of increased *estimated* persistence in the process. This well-known result is also related to the motivation for our smoothed regression procedure that accounts for the bias in the autoregressive parameters. Hence, we anticipate that a change in mean through $\alpha(t/T)$ may affect the estimated persistence for RO, but not for SR. Based on this underlying model, we consider estimating the average parameters over rolling window over the sample of data. Adjusting for endpoints, we are left with 900 observations, so that using 120 observations in the rolling window is approximately $\lambda = 0.133$. Hence, we are using 10 years of data in each of the possible windows, a common choice in empirical studies.

The naive approach (point-wise confidence intervals at each point) just estimates the parameter ω with OLS using 120 observations for each window and includes heteroskedasticity robust standard errors. However, the confidence bands incorporate the incorrect critical value of 1.96. Given that $\lambda = 0.133$, using Table 1 and interpolation gives the appropriate critical value that accounts for the uniform nature of the confidence bands, with a value of 3.415. We plot the resulting rolling estimator of ω and the competing confidence bands in Fig. 2, where the estimated average coefficients are labeled as $\hat{\omega}(r)$ and $\hat{\omega}^*(r)$ for rolling and smoothed rolling regression respectively. In particular, we plot the rolling OLS regression estimator with the incorrect confidence bands and the smoothed rolling regression with the corrected bands that account for the rolling multiple periods.[13]

There are several conclusions that we see from the empirical exercise of comparing the existing naive procedure from the new technique. The confidence bands for the existing RO procedure to not contain our adjusted SR estimates from July of 1968 to July of 1973, October of 1973 to June of 1976, and January of 1986 to December of 1989. In addition, the naive RO procedure was shown to have narrow bands so that confidence levels are far below the prescribed target. A by-product of the incorrect narrow bands of the rolling OLS procedure is the illusion of a more volatile inflation persistence. That is, the smoothed rolling regression estimates with more accurate confidence bands and lack of bias indicates a gradual increase in persistence in the early 1970's, a peak in the early 1980's, and a gradual decline until the early 2000's. Both procedures show a resurgence of inflation persistence in the 2022 estimates of the average coefficients.

The use of our new procedure corrects for the poor numerical performance of confidence bands by increasing the width. Moreover, autoregressive processes estimated via rolling OLS are susceptible to upward bias in the persistence estimate arising from changes in the mean of the process through the change in intercept parameters. We mitigate this bias to isolate the persistence from the change in mean of the process. In the present case, increases in the level of inflation are misinterpreted by some researchers, as increases in persistence relying on rolling OLS.

For the U.S. inflation, the evidence of the stability of coefficients governing inflation depends on the procedure considered. If one employs the traditional rolling regression methods with no modifications, the persistence of inflation appears to change dramatically over the sample. However, after accounting for the potential bias process and using the corrected critical values proposed in this paper, our results are in line with those in O'Reilly and Whelan (2005) for the Euro area, where more stability is observed.[14]

6. Conclusion

The results in this paper provide an asymptotic analysis of the ubiquitous rolling regression estimator for a class of potentially nonstationary processes. Our analysis covers processes that allow forms of nonstationary properties which may

[12] Unit root tests were performed using an Augmented Dickey Fuller test with an intercept and trend, with lag length selected via AIC and the modified AIC method of Ng and Perron (2001). In addition, we also tested for unit roots using tests that allow for breaks in the intercept and trend as developed in Perron (1997) with the minimum t-test in an Augmented Dickey Fuller regression. Both innovative and additive outlier models are considered and lag lengths using AIC and modified AIC were employed. We reject the null of a unit root at the 1% level in all cases allowing for breaks and we reject at the 5% level in the standard ADF test using modified AIC but at the 1% level using AIC.

[13] One could also appeal to the Bonferroni procedure to construct conservative confidence bands. Given the 900 time series observations (for 780 possible sub-samples), Bonferroni critical values would be 3.99, which are even wider than our critical values. We kindly thank a referee for the suggested comparison.

[14] Stock and Watson (2007) provide evidence that inflation is better modeled as an integrated moving average process as opposed to a rolling AR process. In particular, when forecasting one period ahead, the rolling integrated moving average process has lower MSE forecast errors than a rolling AR process for much of the sample. We do not have results for estimating moving average models with time varying parameters and constructing confidence bands for the average parameters. Moving average processes can be approximated by autoregressive models, but an effective approximation should require higher order AR models.

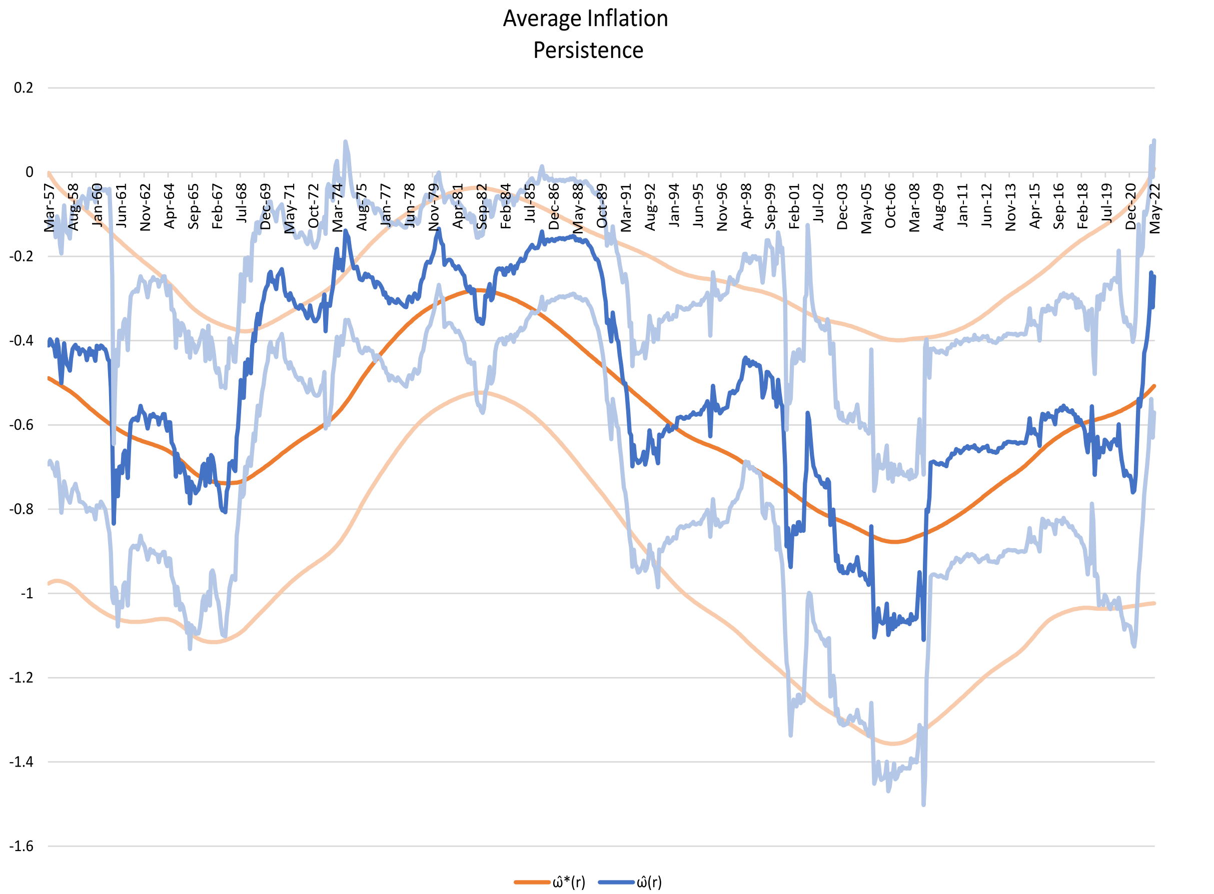

Fig. 2. The potential bias using different methods.
Note: The CPI inflation is calculated as the percentage change from the previous month. The persistence measure is $\omega(t/T)$ calculated from $\Delta y_t = \alpha(t/T) + \omega(t/T)y_{t-1} + \gamma_1(t/T)\Delta y_{t-1} + \gamma_2(t/T)\Delta y_{t-2} + \gamma_3(t/T)\Delta y_{t-3} + \epsilon_t$. We denote $\hat{\omega}(t/T)$ for rolling OLS and $\hat{\omega}^*(t/T)$ represents the smoothed rolling procedure using bandwidth $0.75T^{-0.3}$ (SR1). Confidence bands for RO are calculated using standard normal critical values, while bands for SR1 use critical values based on Table 1.

arise from the changing parameters in the model. In particular, we can cover classes of varying parameter autoregressions. In the simplest cases, the usual procedure for using point-wise confidence bands based on plus and minus 1.96 times the standard error may lead to confidence bands that are much too narrow. The limiting distribution is a functional of Gaussian processes, but is readily tabulated. Our results suggest that one should use our new critical values which depend on the window width to determine uniform confidence bands.

In addition to the new critical values, we show the potential for a bias process arising from a relationship between the distribution of the regressors and the regression parameters. From an empirical standpoint, a dynamic model should be most susceptible for such a process. However, we propose a procedure of averaging smooth coefficient time-varying regression estimators over the relevant window. The resulting distribution is the same as in the case with no bias process and should be used when applying rolling regression in dynamic models. The empirical example covers time-varying persistence in inflation and we show that the new corrected bands suggest that the persistence is historically less variable than previous studies would suggest. However, recent data confirms that inflation has not only recently increased, but has become more persistent as well.

If regression parameters are changing over time, we would ideally prefer to obtain estimates of the parameters at each point in time and estimate confidence bands for the entire function. There is a large literature that suggests procedures to accomplish this task and that convergence rate for the estimated functions and bands is even slower than the rate from using nonparametric confidence intervals at a given point. Rolling regression originally developed as a way to combine observations close to a particular date as a type of local regression. Rolling regression is a natural procedure, which may be employed in many recent empirical studies. The choice of window width for the typical rolling procedure is often based on having "enough" observations in order to estimate parameters, or based on some relevant time frame for the question at hand. Our results show that one can retain the original idea of rolling regression, as well as the usual parametric convergence rate, but the statistical distribution is adjusted to obtain proper coverage for the rolling average of the parameters over the entire range. Moreover, the adjusted rolling regression procedures are simple to implement, with narrower confidence bands relative to fully nonparametric time-varying coefficient models with uniform confidence bands.

Appendix. Proofs

Proof of Theorem 1. A simple algebra leads to

$$\begin{aligned}\sqrt{T}\left(\hat{\beta}_\lambda(r)-\bar{\beta}_\lambda(r)\right)&=\sqrt{T}\left[\left(\sum_{s=[rT-T\lambda+1]}^{[rT]}x_sx_s^\top\right)^{-1}\sum_{s=[rT-T\lambda+1]}^{[rT]}x_sy_s-\bar{\beta}_\lambda(r)\right]\\&=\sqrt{T}\left[\left(\sum_{s=[rT-T\lambda+1]}^{[rT]}x_sx_s^\top\right)^{-1}\sum_{s=[rT-T\lambda+1]}^{[rT]}x_s\left(x_s^\top\beta\left(\frac{s}{T}\right)+\epsilon_s\right)-\bar{\beta}_\lambda(r)\right]\\&=\sqrt{T}\left[\left(\sum_{s=[rT-T\lambda+1]}^{[rT]}x_sx_s^\top\right)^{-1}\sum_{s=[rT-T\lambda+1]}^{[rT]}x_s\epsilon_s\right]+\sqrt{T}B_T(r)\\&\quad+\sqrt{T}\left[\frac{1}{T\lambda}\sum_{s=[rT-T\lambda+1]}^{[rT]}\beta\left(s/T\right)-\bar{\beta}_\lambda(r)\right],\end{aligned}$$

where $B_T(r)$ is the bias process defined in the statement of the theorem. The final term is $o(1)$ by Riemann integrability of $\beta(s/T)$. Then, applying Corollary 2 of Wu and Zhou (2011) for locally stationary processes, we have

$$\frac{1}{\sqrt{T}}\sum_{s=[rT-T\lambda+1]}^{[rT]}x_s\epsilon_s\Rightarrow Q(r)-Q(r-\lambda).$$

Now, consider $x_sx_s^\top$. The process $x_sx_s^\top-M(s/T)$ is mean zero and following the proof of Lemma 6 of Zhou and Wu (2010), we apply Doob's inequality to

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{t}\left(x_sx_s^\top-M(s/T)\right),$$

so that

$$\frac{1}{T}\sum_{s=[rT-T\lambda+1]}^{[rT]}x_sx_s^\top\overset{p}{\to}\int_{r-\lambda}^{r}M(s)ds$$

uniformly in r. Therefore, this completes the proof of Theorem 1.

Proof of Theorem 2. For the local linear estimator, define the term

$$R_{T,0}(s/T)=\frac{1}{T}\sum_{t=[Th]}^{[T(1-h)]}\frac{1}{h}K\left(\frac{t-s}{Th}\right)x_t\epsilon_t.$$

Given our bandwidth choice, we have

$$\sup_{h\le s/T\le 1-h}\left\|M(G,s/T)\{\tilde{\beta}(s/T)-\beta(s/T)\}-R_{T,0}(s/T)\right\|=O_p(\kappa_T\xi_T)$$

where

$$\kappa_T=(Th)^{-1}\left[T^{1/4}+(Th\log T)^{1/2}\right]+h^2\quad\text{and}\quad\xi_T=T^{-1/2}h^{-1}+h$$

from equations (26) and (27) of Zhou and Wu (2010) and (A.3) of Zhang and Wu (2012). Then, we have

$$\begin{aligned}\sqrt{T}\left[\hat{\beta}^*_\lambda(r)-\bar{\beta}_\lambda(r)\right]&=\frac{1}{\sqrt{T}\lambda}\sum_{s=[rT-T\lambda+1]}^{[rT]}\left[\tilde{\beta}(s/T)-\beta(s/T)\right]\\&=\frac{1}{\sqrt{T}\lambda}\sum_{s=[rT-T\lambda+1]}^{[rT]}M(G,s/T)^{-1}R_{T,0}(s/T)+O_p(T^{1/2}\kappa_T\xi_T),\end{aligned}$$

where the last term is $o_p(1)$ given our bandwidth choice. Thus,

$$\begin{aligned}\sqrt{T}\left[\hat{\beta}^*_\lambda(r)-\bar{\beta}_\lambda(r)\right] &= \frac{1}{\sqrt{T}\lambda}\sum_{s=[rT-T\lambda+1]}^{[rT]} M(G,s/T)^{-1}\frac{1}{Th}\sum_{t=[Th]}^{[T(1-h)]} K\left(\frac{t-s}{Th}\right)x_t\epsilon_t + o_p(1)\\ &= \frac{1}{\sqrt{T}\lambda}\sum_{t=[Th]}^{[T(1-h)]}\frac{1}{Th}\sum_{s=[rT-T\lambda+1]}^{[rT]} M(G,s/T)^{-1}K\left(\frac{t-s}{Th}\right)x_t\epsilon_t + o_p(1)\\ &= \frac{1}{\sqrt{T}\lambda}\sum_{s=[rT-T\lambda+1]}^{[rT]} M(G,s/T)^{-1}K\left(\frac{t-s}{Th}\right)\frac{1}{Th}\sum_{t=[Th]}^{[T(1-h)]} x_t\epsilon_t + o_p(1).\end{aligned}$$

Consider

$$D(r)=\frac{1}{T}\sum_{s=[Th]}^{[rT]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}\lambda}\sum_{t=[Th]}^{[T(1-h)]} x_t\epsilon_t.$$

By writing $D(r)=D^*(r)-D_1(r)+D_2(r)$, we have

$$D^*(r)=\frac{1}{T\lambda}\sum_{s=[Th]}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Th]}^{[Tr]} x_t\epsilon_t,$$

$$D_1(r)=\frac{1}{T\lambda}\sum_{s=[Tr]+1}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Th]}^{[Tr]} x_t\epsilon_t,$$

and

$$D_2(r)=\frac{1}{T\lambda}\sum_{s=[Th]}^{[Tr]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Tr]+1}^{[T(1-h)]} x_t\epsilon_t.$$

We show that $D_1(r)$ and $D_2(r)$ converge to zero uniformly in r. To this end, it is easy to see that

$$\|D_1(r)\|\le\left\|\frac{1}{\sqrt{T}\lambda}\sum_{t=[Th]}^{[Tr]} x_t\epsilon_t\right\|\left\|\frac{1}{T}\sum_{s=[Tr]+1}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\right\|.$$

Consider the second term on the right hand side. Denote the eigenvalue decomposition as $M(G,s/T)=\Gamma(s/T)\Theta(s/T)\Gamma(s/T)^{-1}$ so that

$$\begin{aligned}&\left\|\frac{1}{T}\sum_{s=[Tr]+1}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\right\|\le\frac{1}{Th}\sum_{s=[Tr]+1}^{[T(1-h)]}\left\|M(G,s/T)^{-1}K\left(\frac{t-s}{Th}\right)\right\|\\ &=\frac{1}{Th}\sum_{s=[Tr]+1}^{[T(1-h)]} K\left(\frac{t-s}{Th}\right)\sqrt{tr\left[\Gamma(s/T)\Theta(s/T)^{-2}\Gamma(s/T)^{-1}\right]}\\ &\le\frac{1}{Th}\sum_{s=[Tr]+1}^{[T(1-h)]} K\left(\frac{t-s}{Th}\right)c_M^{-2k},\end{aligned}$$

where tr denotes the trace of a matrix and c_M is the lower bound of eigenvalues of $M(s/T)$. Note that $t\le[Tr]$, so that $t<s$. Then

$$\frac{1}{Th}\sum_{s=[Tr]+1}^{[T(1-h)]} K\left(\frac{t-s}{Th}\right)\approx\int_r^{1-h}\frac{1}{h}K\left(\frac{u-v}{h}\right)dv=\int_{\frac{r-u}{h}}^{\frac{1-h-u}{h}} K(z)dz,$$

where $z=(v-u)/h$, $u<1-h$ and $r<u$. Then, as $h\to 0$, this integral is $O(h)$, since $K(z)=0$ for $|z|\ge 1$. Hence, $D_1(r)$ converges in probability to 0 uniformly in r. The argument is similar to that for $D_2(r)$. For $D^*(r)$, we note that

$$\begin{aligned}&\frac{1}{Th}\sum_{s=[Th]}^{[T(1-h)]} M(G,s/T)^{-1}K\left(\frac{t-s}{Th}\right)\approx\int_h^{(1-h)}\frac{1}{h}M(G,v)^{-1}K\left(\frac{u-v}{h}\right)dv\\ &=\int_{\frac{u-r}{h}}^{\frac{u-r+\lambda}{h}} M(G,u-zh)^{-1}K(z)dz=M(G,u)^{-1}+O(h).\end{aligned}$$

Then,

$$D^*(r) = \frac{1}{\sqrt{T}\lambda} \sum_{t=[Th]}^{[rT]} M(G, t/T)^{-1} x_t \epsilon_t + o_p(1)$$

uniformly in r. Combining these results, we have

$$\sqrt{T}\left[\hat{\beta}_\lambda^*(r) - \bar{\beta}_\lambda(r)\right] = D^*(r) - D^*(r-\lambda) + o_p(1).$$

Given this representation, we apply Corollary 2 of Wu and Zhou (2011) to obtain

$$\frac{1}{\sqrt{T}\lambda} \sum_{t=[rT-T\lambda+1]}^{[rT]} M(G, t/T)^{-1} x_t \epsilon_t \Rightarrow Q_2(r) - Q_2(r-\lambda),$$

where $Q_2(r)$ is a p dimensional Gaussian process with covariance $E\left[Q_2(r_1)Q_2(r_2)^\top\right] = \int_0^{\min(r_1,r_2)} \Lambda(s)$ and

$$\Lambda(s) = \frac{1}{\lambda^2} M(s)^{-1} \Omega(s) M(s)^{-1}.$$

Therefore, the proof of Theorem 2 is complete.

References

Adrian, T., Crump, R.K., Moench, E., 2015. Regression-based estimation of dynamic asset pricing models. J. Financ. Econ. 118, 211–244.
Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.
Belloni, A., Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. J. Econometrics 186, 345–366.
Blanchard, O., 2018. Should we reject the natural rate hypothesis? J. Econ. Perspect. 32, 97–120.
Blanchard, O., Gali, J., 2009. The macroeconomic effects of oil shocks: Why are the 2000s so different from the 1970s? In: Gali, J., Gertler, M. (Eds.), International Dimensions of Monetary Policy. University of Chicago Press, pp. 373–428.
Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. J. Econometrics 136, 163–188.
Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. J. Econometrics 167, 413–425.
Campello, M., Galvao, A., Juhl, T., 2019. Testing for slope heterogeneity bias in panel data models. J. Bus. Econom. Statist. 37, 749–760.
Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80, 1157–1183.
Clark, T.E., McCracken, M.W., 2009. Improving forecast accuracy by combining recursive and rolling forecasts. Internat. Econom. Rev. 50, 363–395.
Eubank, R.L., Speckman, P.L., 1993. Confidence bands in nonparametric regression. J. Amer. Statist. Assoc. 88, 1287–1301.
Georgiev, I., Harvey, D.I., Leybourne, S.J., Taylor, A.R., 2018. Testing for parameter instability in predictive regression models. J. Econometrics 204, 101–118.
Giacomini, R., Rossi, B., 2010. Forecast comparisons in unstable environments. J. Appl. Econometrics 25, 595–620.
Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Inoue, A., Jin, L., Pelletier, D., 2021. Local linear estimation of time-varying-parameter GARCH models and associated risk measures. J. Financ. Econom. 19, 202–234.
Jimeńez, G., Ongena, S., Peydró, J.-L., Saurina, J., 2017. Macro-prudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the Spanish dynamic provisioning experiments. J. Polit. Econ. 125, 2126–2177.
Johnston, G., 1982. Probabilities of maximal deviations for nonparametric regression function estimates. J. Multivariate Anal. 12, 402–414.
Li, J., Liao, Z., 2020. Uniform nonparametric inference for time series. J. Econometrics 219, 38–51.
Linnainmaa, J.T., Roberts, M.R., 2018. The history of the cross section of stock returns. Rev. Financ. Stud. 31, 2606–2649.
Løpez-Salido, D., Stein, J.C., Zkrajšek, E., 2017. Credit-market sentiment and the business cycle. Q. J. Econ. 132, 1373–1426.
Müller, U.K., Petalas, P.E., 2010. Efficient estimation of the parameter path in unstable time series models. Rev. Econom. Stud. 77, 1508–1539.
Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.
O'Reilly, G., Whelan, K., 2005. Has Euro-area inflation persistence changed over time. Rev. Econ. Stat. 87, 709–720.
Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.
Perron, P., 1997. Further evidence on breaking trend functions in macroeconomic variables. J. Econometrics 80, 355–385.
Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer Verlag, New York, pp. 253–264.
Stock, J., Watson, M., 2007. Why has U.S. inflation become harder to forecast? J. Money Credit Bank. 39, 3–33.
Swanson, E.T., Williams, J.C., 2014. Measuring the effect of the zero lower bound on medium-and longer-term interest rates. Amer. Econ. Rev. 104, 3154–3185.
Wang, J., Yang, L., 2009. Polynomial spline confidence bands for regression curves. Statist. Sinica 19, 325–342.
Wu, W.B., Zhou, Z., 2011. Gaussian approximations for non-stationary multiple time series. Statist. Sinica 21, 1397–1413.
Zhang, T., Wu, W.B., 2012. Inference of time-varying regression models. Ann. Statist. 40, 1376–1402.
Zhou, Z., Wu, W.B., 2010. Simultaneous inference of linear models with time varying coefficients. J. R. Stat. Soc. Ser. B Stat. Methodol. 72, 513–531.