The Distribution of Rolling Regression Estimators *†

Zongwu Cai, Ted Juhl
University of Kansas

January 4, 2021

Abstract

We find 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 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.

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

1 Introduction

Rolling regression is often employed in many applied fields as a method to characterize changing 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 methods of examining the stability of statistical relationships over time. A cursory search will reveal that there are rolling regression routines written

*We thank Graham Elliott, Bruce Hansen, and Ulrich Müller for comments on earlier versions of this paper.

†Contact information: Zongwu Cai, Department of Economics, University of Kansas, Lawrence, KS, 66045 (caiz@ku.edu); Ted Juhl, School of Business, University of Kansas, Lawrence, KS 66045 (juhl@ku.edu)

in statistical computing languages or packages such as R, SAS, 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 some type of ocular inference about whether there are changes in 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 Swanson and Williams (2014), Linnainmaa and Roberts (2017), Adrian et al. (2015), Blanchard (2018), Georgiev et. al. (2018), Jiménez et. al. (2017), and López-Salido et al. (2017), among others.[1]

In this paper, we characterize the population parameters that rolling regression is attempting to estimate and we 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 we want 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, we 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 proce-

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

dure 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. Finally, all technical proofs are gathered in Section 7.

2 Model and Assumptions

It is natural to ask what we are attempting to estimate by employing rolling regression. Consider a standard regression model given by

$$y_t = x_t^\top \beta + \epsilon_t, \quad 1 \le t \le T,$$

where x_t is a p-dimensional regressor. 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 we index each of the periods with r so that we have

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

Clearly, 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 will be inefficient relative to the full sample regression estimator. Consider the model

$$y_t = x_t^\top \beta \left(\frac{t}{T} \right) + \epsilon_t$$

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); see, for example, Cai (2007) for details. 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 non-stationarity. In particular, given that we are attempting to estimate time-varying parameters, it is natural to allow for non-stationary processes. Changing values of $\beta(t/T)$ necessarily induces non-stationarity in y_t. Moreover, if the model is dynamic, then x_t is also non-stationary if $\beta(t/T)$ changes. To this end, consider processes depending on deterministic functions coupled with the iid process v_t. Let $\mathcal{F}_t = \{\ldots, v_{t-1}, v_t\}$. The processes for the covariates and errors are given by

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

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

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

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

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

Statisticians characterize dependence in data in several ways; linear processes, α-mixing, β-mixing, etc. Recent papers by Chen and Hong (2012) and Cai (2007) both make use of β-mixing and α-mixing assumptions, but the data are assumed to be stationary in both papers. The assumption in Inoue et al. (2017) allows the data to be near-epoch dependent (NED). As argued in Inoue et al. (2017), the NED assumption is more general than the α-mixing assumption, allows for heterogeneity over time which is necessary for time-varying parameter framework, and overcomes several undesirable features of the α-mixing assumption as addressed by Lu and Linton (2007). In this paper, we follow Zhou and Wu (2008) and allow for non-stationary 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 $\mathcal{F}_j^* = \{\ldots, v_{-1}, v_0', v_1, \ldots, v_{j-1}, v_j\}$. 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

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

for some $q \geq 1$, 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.

The dependence in these processes over time is a separate issue than whether the process is stationary. One class of non-stationary 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 non-stationary in a way that is separate from the dependence over time. To this end, we denote a process G to be stochastically Lipshitz continuous if

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

for finite $c_1 > 0$.

Assumption 1 *The true data generating process is given by*

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

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 Lipshitz 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 martingale differences for the error process. In particular, autoregressive models with time varying coefficients are considered in Zhang and Wu (2012). They show that, given (standard) conditions on the time varying roots of the characteristic function, the process is locally stationary and they characterize 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 Section 7.

Theorem 2.1 *Suppose that Assumptions 1 to 5 hold and that $r \in [\lambda, 1]$. Then,*

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

where

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

serves as the asymptotic bias term, $\Rightarrow$ indicates weak convergence, and $Q(r)$ denotes a p-dimensional Gaussian process with covariance $E[Q(r_1)Q(r_2)^\top] = \int_0^{\min(r_1,r_2)} \Omega(s)$.

The result 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 the Monte Carlo section, using standard normal critical values for confidence bands for the estimate of the average parameter vector, $\bar{\beta}_\lambda(r)$, will 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.

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 will 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 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 non-zero mean. The intuition for the bias process is that regression over the rolling interval will weight 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 the second moments of the regressors.

Theorem 2.1 allows for both x_t and ϵ_t to exhibit a form of non-stationary behavior, which complicates the limiting distribution of the rolling estimator. The following corollary provides a simplification. If x_t and ϵ_t are both stationary, the result simplifies. First, the bias process $B_T(r)$ disappears. 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

$$y_t = \alpha(\frac{t}{T}) + \phi(\frac{t}{T})y_{t-1} + \epsilon_t \quad \text{where} \begin{pmatrix} \alpha(\frac{t}{T}) \\ \phi(\frac{t}{T}) \end{pmatrix} = \begin{pmatrix} k_1 \times \dfrac{t}{T} \\ \phi \end{pmatrix},$$

so that the autoregressive coefficient $\phi(t/T)$ is constant but the intercept term is changing

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

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

$$k_1^2(1-\phi)(1-\phi^2)/\left[k_1^2(1-\phi^2)+12\sigma^2\right].$$

To simplify further, if $\phi = 0$, bias is

$$\left[1+12\sigma^2/k_1^2\right]^{-1}.$$

This simple case is much like the result in Perron (1989) for 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 $B_T(r)$ will be larger in magnitude as there is more correlation in the second moment of x_t and the parameter vector $\beta(t/T)$.

2.1 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

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

If there is no bias process, it is easy to see that the standardized process will converge 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 non-zero covariance when $r_1 > r_2 - \lambda$.

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

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

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. To illustrate the distribution under this case when $k = 1$, 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 0.85. 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 0.99 quantiles of empirical distribution are provided in Table 1.

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 will be 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, these tables provide 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 will 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

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

Table 1. Critical Values for Rolling Confidence Bands

	Confidence Level				Confidence Level				Confidence Level		
λ	0.90	0.95	0.99	λ	0.90	0.95	0.99	λ	0.90	0.95	0.99
0.050	3.498	3.708	4.148	0.320	2.795	3.059	3.590	0.590	2.423	2.707	3.262
0.060	3.440	3.648	4.092	0.330	2.780	3.048	3.581	0.600	2.404	2.695	3.260
0.070	3.395	3.614	4.074	0.340	2.767	3.028	3.564	0.610	2.389	2.685	3.252
0.080	3.346	3.570	4.016	0.350	2.753	3.016	3.533	0.620	2.380	2.672	3.266
0.090	3.311	3.535	3.992	0.360	2.730	3.000	3.533	0.630	2.380	2.674	3.248
0.100	3.271	3.499	3.956	0.370	2.712	2.983	3.523	0.640	2.357	2.655	3.238
0.110	3.245	3.478	3.945	0.380	2.700	2.970	3.512	0.650	2.343	2.636	3.214
0.120	3.213	3.444	3.912	0.390	2.690	2.962	3.496	0.660	2.332	2.628	3.208
0.130	3.178	3.415	3.895	0.400	2.677	2.942	3.481	0.670	2.322	2.623	3.214
0.140	3.159	3.388	3.866	0.410	2.662	2.935	3.477	0.680	2.309	2.600	3.176
0.150	3.125	3.356	3.843	0.420	2.646	2.920	3.469	0.690	2.288	2.585	3.163
0.160	3.105	3.344	3.823	0.430	2.635	2.917	3.462	0.700	2.283	2.578	3.176
0.170	3.080	3.319	3.808	0.440	2.617	2.896	3.449	0.710	2.264	2.562	3.155
0.180	3.056	3.299	3.780	0.450	2.603	2.887	3.436	0.720	2.254	2.551	3.145
0.190	3.035	3.284	3.784	0.460	2.591	2.870	3.419	0.730	2.239	2.540	3.135
0.200	3.013	3.265	3.755	0.470	2.582	2.854	3.410	0.740	2.221	2.524	3.136
0.210	2.993	3.240	3.729	0.480	2.563	2.846	3.398	0.750	2.212	2.512	3.115
0.220	2.976	3.225	3.701	0.490	2.554	2.834	3.384	0.760	2.202	2.501	3.079
0.230	2.942	3.194	3.704	0.500	2.544	2.825	3.388	0.770	2.190	2.491	3.093
0.240	2.926	3.182	3.708	0.510	2.530	2.818	3.367	0.780	2.167	2.473	3.062
0.250	2.921	3.174	3.691	0.520	2.513	2.796	3.355	0.790	2.155	2.461	3.058
0.260	2.896	3.154	3.656	0.530	2.499	2.779	3.342	0.800	2.154	2.459	3.059
0.270	2.876	3.134	3.660	0.540	2.483	2.770	3.326	0.810	2.134	2.438	3.029
0.280	2.863	3.120	3.646	0.550	2.468	2.758	3.322	0.820	2.114	2.424	3.024
0.290	2.844	3.103	3.621	0.560	2.459	2.749	3.317	0.830	2.112	2.414	3.027
0.300	2.828	3.091	3.619	0.570	2.443	2.733	3.295	0.840	2.095	2.399	2.995
0.310	2.818	3.078	3.600	0.580	2.428	2.718	3.284	0.850	2.072	2.377	2.999

The simulated (standardized) process will 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 assumption that the bias process is zero, as would be the case if the 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 estimator for $\beta(t/T)$ was analyzed in Cai (2007) for stationary α-mixing data, and is given by $\tilde{\beta}(t/T)$, where

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

with

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

$K_{t,s} = K(t-s/Th)$, and $K(\cdot)$ being a kernel function. It is well known that the nonparametric estimator is consistent and point-wise normally distributed. 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.

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

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

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 altogether. We list three additional assumptions for the data generating process and the bandwidth.

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

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

Assumption 8 *The kernel function* $K(z)$ *is second order 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 will converge at the usual parametric rate, and Assumption 8 is useful to limit the observations that enter the smoothed estimators. The Epanechnikov kernel is a popular example of such a kernel and we employ this kernel in our Monte Carlo and empirical examples. We state the theorem for the modified procedure below with its proof relegated to Section 7.

Theorem 3.1 *Suppose that Assumptions 1-8 hold with* $r \in (\lambda, 1)$.

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

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

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

The new estimator is consistent for $\hat{\beta}_\lambda(r)$. 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 $\hat{\beta}_\lambda(r)$. 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 semi-parametric models involving averages of nonparametric estimates. 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.

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. Let

$$\tilde{S}(s) = \begin{bmatrix} S_{T,0}(s) & S_{T,1}^{\top}(s) \\ S_{T,1}(s) & S_{T,2}(s) \end{bmatrix}, \quad \tilde{\Omega}_{\ell}(s) = \frac{1}{Th}\sum_{r=1}^{T} x_r x_r^{\top} \tilde{\epsilon}_r^2 \left(\frac{s-r}{Th}\right)^{\ell} K\left(\frac{s-r}{Th}\right),$$

$$\tilde{\Omega}(s) = \begin{bmatrix} \tilde{\Omega}_0(s) & \tilde{\Omega}_1(s) \\ \tilde{\Omega}_1(s) & \tilde{\Omega}_2(s) \end{bmatrix}, \quad \text{and} \quad \tilde{\epsilon}_r = y_r - x_r^{\top}\tilde{\beta}(r).$$

Then, the standard errors are estimated via the variance matrix

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

The standardized estimator will be 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 $r_1 < r_2 - \lambda$).

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 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.

3.1 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 Studies

The theorems from Sections 2 and 3 show 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 3.1 is to provide a bias corrected estimate of the average slope.

We explore the performance of various rolling regression estimators in this section. In particular, the naive rolling regression estimator that uses standard normal critical values is denoted RO for Rolling OLS. A second estimator is considered but uses the new critical values, and we refer to this estimator as adjRO. 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 $-1/4 < \delta < -1/3$. We use $\delta = 0.30$ and set $c_2 = 0.5$, 0.75, and 1. The estimators are denoted SRb1, SRb2, and SRb3.

Before we consider rolling regression, we illustrate the bias that one encounters if an autoregressive process has an omitted trend, which could be considered a case where $\lambda = 1$. To this end, we simulate the process discussed in Section 2 with a missing trend. The bias of the estimated AR parameter 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 $T = 200$ and 1000 replications for each value of k_1. The resulting bias for OLS and the smoothed estimates (Sb1, Sb2, Sb3) appears in Figure 1 (see later). 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 of 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 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 value of $\beta(r)$ over each relevant subset of time. Our experiments report the estimated coverage probability for a 95% uniform confidence band for $\bar{\beta}_\lambda(r)$. We report the mean average deviation over the range. In addition, we list the average width of the uniform confidence band. For example, if two competing procedures both have 95% coverage, we would prefer the method generating narrow bands.

Denote the current naive technology of using rolling OLS and employing critical values from the standard normal distribution as RO (rolled OLS). If we use the adjusted critical values tabulated in Table 1 with rolled OLS, we list the procedure as adjRO. Our smoothed rolling procedure using the adjusted critical values is indexed by the bandwidths as SRb1, SRb2, and SRb3.[3]

Our first experiment uses a simple static regression model given by

$$y_t = 2x_t + \epsilon_t$$

where $\epsilon_t \sim N(0, 0.25)$. The results are summarized in Table 2.

We first examine the rolling OLS (RO) estimator using the (incorrect) standard normal distribution critical values. The coverage for the procedure is below the nominal 95%, and ranges from around 19% to 26% 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 91% when the sample size is 600. The smoothed rolling procedures have much better coverage for all the considered bandwidth parameters. The SRb2 performs best, does not vary as much with the sample size. The mean absolute deviation (MAD) is similar for all procedures, which is to be expected since the coverage issues arise from incorrect critical values. The smoothed procedures has the narrowest uniform bands (average width) while having the best coverage.

[3]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.

Table 2. Static Regression

		SRb1	SRb2	SRb3	adjRO	RO
	$T = 200$	0.8924	0.9575	0.9839	0.8357	0.1940
Coverage	$T = 400$	0.9052	0.9586	0.9819	0.8991	0.2386
	$T = 600$	0.9081	0.9530	0.9773	0.9152	0.2623
	$T = 200$	0.0577	0.0520	0.0478	0.0650	0.0650
MAD	$T = 400$	0.0414	0.0382	0.0355	0.0455	0.0455
	$T = 600$	0.0336	0.0314	0.0294	0.0367	0.0367
	$T = 200$	0.3918	0.4001	0.4098	0.4978	0.2988
Av. Width	$T = 400$	0.2800	0.2837	0.2883	0.3585	0.2152
	$T = 600$	0.2293	0.2318	0.2348	0.2944	0.1767

Next, we generate data from a simple autoregressive model with no parameter changes. The data generating process is given by

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

where ϵ_t is N(0,0.25). The results are given in Table 3.

Table 3. Constant AR, $\rho = 0.80$

		SRb1	SRb2	SRb3	adjRO	RO
	$T = 200$	0.3502	0.7647	0.9210	0.6154	0.0552
Coverage	$T = 400$	0.4406	0.7947	0.9231	0.8041	0.1266
	$T = 600$	0.4865	0.8076	0.9224	0.8579	0.1734
	$T = 200$	0.1975	0.1350	0.1059	0.1124	0.1124
MAD	$T = 400$	0.1200	0.0832	0.0666	0.0672	0.0672
	$T = 600$	0.0893	0.0630	0.0513	0.0511	0.0511
	$T = 200$	0.5782	0.5623	0.5592	0.6724	0.4036
Av. Width	$T = 400$	0.3871	0.3773	0.3753	0.4480	0.2689
	$T = 600$	0.3075	0.3008	0.2995	0.3686	0.2213

Given the rolling fraction of $\lambda = 0.20$, the procedures are attempting to estimate the AR parameter based on 40 observations. The SRb3 procedure is best for this data generating process, with small MAD and narrow confidence bands, but the adjusted rolling OLS procedure performs respectably. Again, the naive traditional rolling OLS procedure performs poorly, with coverage of 17% even with a sample of 600 observations.

The next process we consider has a changing autoregressive coefficient

$$y_t = \rho(t/T)y_{t-1} + \epsilon_t \quad \text{with} \quad \rho(t/T) = 0.8\,[1 - (t/T)]\,,$$

and our results appear in Table 4.

Table 4. AR, $\rho = 0.80[1 - (t/T)]$

		SRb1	SRb2	SRb3	adjRO	RO
	$T = 200$	0.6907	0.8937	0.9631	0.7701	0.1451
Coverage	$T = 400$	0.7559	0.9109	0.9651	0.8702	0.1950
	$T = 600$	0.7815	0.9156	0.9642	0.8989	0.2236
	$T = 200$	0.1417	0.1131	0.0983	0.1235	0.1235
MAD	$T = 400$	0.0948	0.0784	0.0696	0.0837	0.0837
	$T = 600$	0.0748	0.0632	0.0569	0.0672	0.0672
	$T = 200$	0.6856	0.7028	0.7195	0.8850	0.5313
Av. Width	$T = 400$	0.4916	0.5000	0.5080	0.6380	0.3830
	$T = 600$	0.4040	0.4094	0.4146	0.5248	0.3150

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 perform well for all sample sizes and have narrower confidence bands.

For the next experiment, the autoregressive coefficients do not change, but there is an omitted trend variable. This type of data generating process will cause bias in the autoregressive parameter if one uses standard OLS, which are illustrated in Figure 1. The process is given by

$$y_t = 2 \times (t/T) + \rho y_{t-1} + \epsilon_t.$$

We report the coverage for the uniform bands for the autoregressive coefficient in Table 5.

The biases in standard OLS are apparent in both the rolled OLS and adjusted rolled OLS based confidence bands with coverage of 0% for all sample sizes. The bias appears in

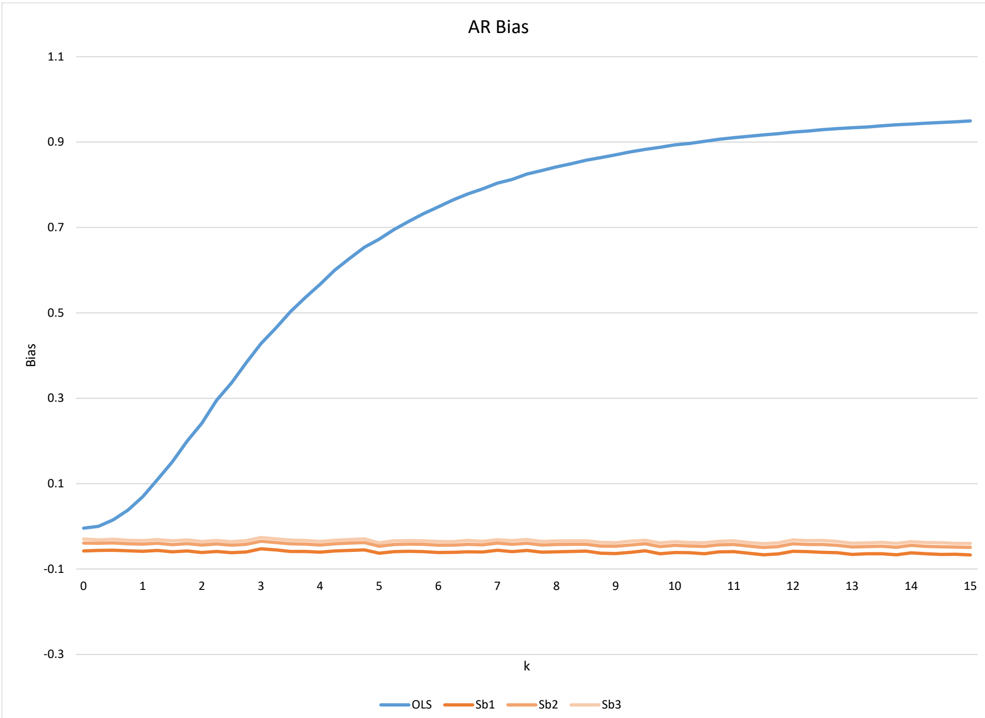

Figure 1: Bias for OLS and the smoothed estimates (Sb1, Sb2, Sb3).

Table 5. Omitted trend, $2 \times (t/T)$

		SRb1	SRb2	SRb3	adjRO	RO
Coverage	$T = 200$	0.8601	0.9480	0.9830	0.0000	0.0000
	$T = 400$	0.8859	0.9490	0.9783	0.0000	0.0000
	$T = 600$	0.8917	0.9497	0.9774	0.0000	0.0000
MAD	$T = 200$	0.1151	0.1031	0.0942	0.5620	0.5620
	$T = 400$	0.0820	0.0753	0.0699	0.5663	0.5663
	$T = 600$	0.0674	0.0626	0.0586	0.5678	0.5678
Av. Width	$T = 200$	0.7278	0.7570	0.7781	0.7692	0.4618
	$T = 400$	0.5323	0.5471	0.5576	0.5545	0.3347
	$T = 600$	0.4410	0.4508	0.4577	0.4559	0.2737

the MAD rows of Table 5 where both adjusted rolling OLS and rolling OLS are over five times larger than the smoothed rolling estimators.

The next data generating process is given by

$$y_t = \rho(t/T)y_{t-1} + \epsilon_t \quad \text{with} \quad \rho(t/T) = 0.8\sin[2\pi \times (t/T)],$$

so that the autoregressive coefficient starts at zero, and oscillates between 0.80 and zero. The results are displayed in Table 6.

Table 6. AR, $\rho = 0.80\sin[2\pi \times (t/T)]$

		SRb1	SRb2	SRb3	adjRO	RO
	$T = 200$	0.6669	0.8111	0.8484	0.6594	0.0808
Coverage	$T = 400$	0.7212	0.8347	0.8535	0.6928	0.0766
	$T = 600$	0.7445	0.8408	0.8477	0.6683	0.0621
	$T = 200$	0.1334	0.1214	0.1256	0.1202	0.1202
MAD	$T = 400$	0.0895	0.0833	0.0876	0.0848	0.0848
	$T = 600$	0.0710	0.0668	0.0708	0.0710	0.0710
	$T = 200$	0.6222	0.6384	0.6581	0.8040	0.4826
Av. Width	$T = 400$	0.4439	0.4528	0.4637	0.5789	0.3475
	$T = 600$	0.3635	0.3695	0.3773	0.4763	0.2859

The performance of rolling OLS is again poor, and the adjusted rolling OLS is also below nominal confidence levels. The smoothed rolling procedures improve as the sample size increases, up to coverage of 84%.

Finally, we treat the case where the intercept term follows increases and then decreases with

$$y_t = \sin[2\pi \times (t/T)] + \rho y_{t-1} + \epsilon_t.$$

The results are presented in Table 7.

The estimators based on OLS perform poorly and get worse as the sample size increases. Our smoothed rolling estimators perform well and improve as the sample size increases. The MAD gets smaller and the width of the confidence bands declines with increases in sample size.

The Monte Carlo experiments 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

Table 7. AR, Omitted trend, $\sin[2\pi \times (t/T)]$

		SRb1	SRb2	SRb3	adjRO	RO
	$T = 200$	0.8860	0.9676	0.9836	0.2161	0.0025
Coverage	$T = 400$	0.8956	0.9622	0.9830	0.0458	0.0001
	$T = 600$	0.9024	0.9591	0.9825	0.0086	0.0000
	$T = 200$	0.1010	0.0962	0.0916	0.1909	0.1909
MAD	$T = 400$	0.0807	0.0725	0.0674	0.1736	0.1736
	$T = 600$	0.0663	0.0610	0.0569	0.1682	0.1682
	$T = 200$	0.7273	0.7561	0.7743	0.9395	0.5640
Av. Width	$T = 400$	0.5319	0.5465	0.5562	0.6846	0.4110
	$T = 600$	0.4407	0.4505	0.4572	0.5651	0.3392

accurate confidence bands. However, from the theorems in earlier sections, we know that bias results when the 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.

5 An Empirical Application

Rolling regression was employed by O'Reilly and Whelan (2005) to examine the persistence of inflation over time in the Euro area. Suppose that one estimates an AR(3) model of 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

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

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 will be negative, and the closer to -1 the less persistent is inflation. We use data on inflation from the CPI, with monthly observations from February of 1947 to May of 2020. As a preliminary analysis, we estimate the AR(4) model and test for serial correlation in the residuals. We fail to reject the null of no serial correlation at the 5% level, and we also reject the unit root hypothesis at the 1% level. We proceed to analyze rolling estimates of the persistence parameter ω_1. 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. To this end, we consider a rolling window over the sample of data. Adjusting for endpoints, we are left with 880 observations, so that using $\lambda = 0.20$ results in using 176 observations in the rolling window.

The naive approach just estimates the parameter ω with OLS using 176 observations for each window and includes heteroskedasticity robust standard errors. However, the confidence bands incorporate the incorrect critical value of 1.96. Given that $\lambda = 0.20$, using Table 1 gives the appropriate critical value that accounts for the uniform nature of the confidence bands, with a value of 3.265, just as used in the Monte Carlo experiments. We plot the resulting rolling estimator of ω and the competing confidence bands in Figure 2. That is, we plot the rolling OLS (RO) regression estimator with the incorrect confidence bands, and the smoothed rolling regression (SR) with the corrected bands that account for the rolling multiple periods.

There are several conclusions that we see from the empirical exercise of comparing the existing naive procedure (RO) from the new technique (SR). The first point is that the naive function is not contained in the corrected bands, suggesting that the bias process is nonzero. For example, in May of 1969 through August of 1980, we see that the naive rolling OLS estimate is outside our bands, and the same is true later in the sample in November of 1991. 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 incorrectly narrow bands of the naive 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 mid 1980's, and a gradual decline until the early 2000's.

The examples with an autoregressive process show that the result of a change in the trend or mean of the process is upward bias in the estimated persistence. In light of this effect, we

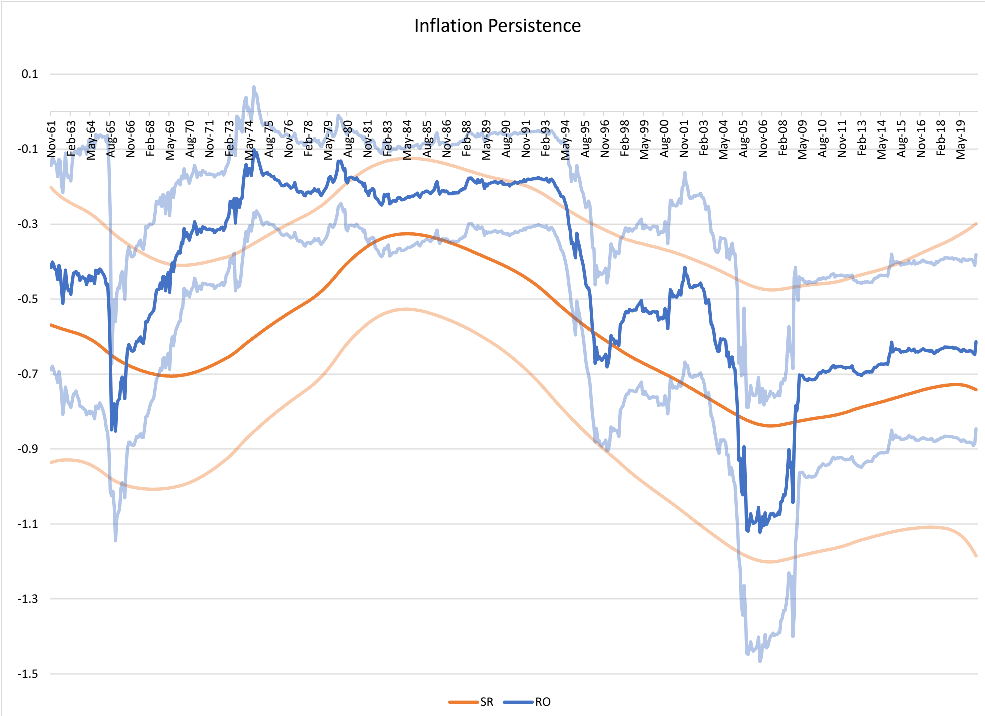

Figure 2: Confidence bands for SR and RO

estimate the α parameter which also influences the mean of the autoregressive process. The smoothed rolling estimate is shown in Figure 3. We anticipate the more that the estimate of α changes, the more upward bias we will see in the estimate of persistence. From Figure 3 coupled with Figure 2, this is indeed the case from May of 1964 until May of 1979, where the naive rolling OLS estimate is much higher than our smoothed estimate, SR.

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 researchers relying on rolling OLS as increases in persistence.

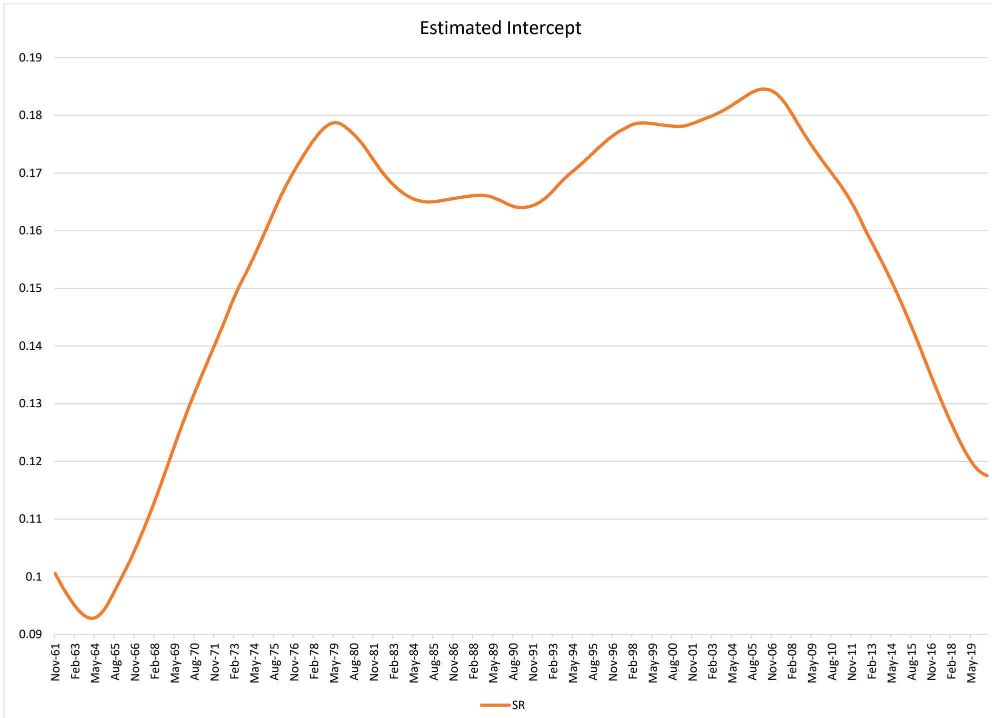

Figure 3: The smoothed rolling estimate.

6 Conclusion

The results in this paper provide an asymptotic analysis of the ubiquitous rolling regression estimator for a class of potentially non-stationary processes. Our analysis covers processes that allow forms of non-stationary properties which may arise from the changing parameters in the model. In particular, we can cover classes of varying parameter autoregressions, for example. In the simplest cases, the usual procedure for using point-wise confidence bands based on plus and minus 1.96 times the standard error will 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 showed 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 will 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 less variable than previous studies would suggest.

Rolling regression is a natural procedure, and one 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 rates, but the statistical distribution is adjusted to obtain proper coverage. 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.

7 Proofs

Proof of Theorem 2.1:

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

where $B_T(r)$ is the bias process defined in the statement of the theorem. The final term is $o(1)$ by Reimann 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 $x_s x_s^\top$. The process $x_s x_s^\top - M(s/T)$ is mean zero, and following the proof of Lemma 6 of Zhou and Wu (2010), we apply Doob's inequality to

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

so that

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

uniformly in r.

Proof of Theorem 3.1 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

$$\kappa_T = (Th)^{-1} T^{1/\iota} + (Th \log T)^{1/2} + h^2$$
$$\xi_T = T^{-1/2} h^{-1} + h$$

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

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

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

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

Consider

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

We write

$$\begin{aligned}D(r) &= D^*(r)-D_1(r)+D_2(r)\\ D^*(r) &= \frac{1}{T\lambda}\sum_{s=[Th]}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Th]}^{[Tr]} x_t\epsilon_t\\ D_1(r) &= \frac{1}{T\lambda}\sum_{s=[Tr]+1}^{[T(1-h)]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Th]}^{[Tr]} x_t\epsilon_t\\ D_2(r) &= \frac{1}{T\lambda}\sum_{s=[Th]}^{[Tr]} M(G,s/T)^{-1}\frac{1}{h}K\left(\frac{t-s}{Th}\right)\frac{1}{\sqrt{T}}\sum_{t=[Tr]+1}^{[T(1-h)]} x_t\epsilon_t\end{aligned}$$

We show that $D_1(r)$ and $D_2(r)$ converge to zero uniformly in r.

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

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 c_M is the lower bound of eigenvalues of $M(s/T)$. Note that $t \leq [Tr]$, so that $t < s$. Then

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

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 $|z| \geq 1$. Hence, $D_1(r)$ converges in probability to 0 uniformly in r. The argument is similar for $D_2(r)$.

For $D^*(r)$ we note that

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

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), so that

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

where $Q_2(r)$ is 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

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

References

Adrian, T., R.K. Crump and E. Moench (2015). Regression-Based Estimation of Dynamic Asset Pricing Models. *Journal of Financial Economics*, **118**, 211-244.

Andrews, D.W.K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, **61**, 821-856.

Blanchard, O. (2018). Should We Reject the Natural Rate Hypothesis?, *Journal of Economic Perspectives*, **32**, 97-120.

Cai, Z. (2007). Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors. *Journal of Econometrics*, **136**, 163-188.

Campello, M., A. Galvao and T. Juhl (2019). Testing for Slope Heterogeneity Bias in Panel Data Models. *Journal of Business and Economic Statistics*, **37**, 749-760.

Chen, B. and Y. Hong (2012). Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression. *Econometrica*, **80**, 1157-1183.

Clark, T.E. and M.W. McCracken (2009). Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts. *International Economic Review*, **50**, 363-395.

Georgiev, I., D.I. Harvey, S.J. Leybourne and A.R. Taylor (2018). Testing for Parameter Instability in Predictive Regression Models. *Journal of Econometrics*, **204**, 101-118.

Hansen, B.E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, **64**, 413-430.

Inoue, A., L. Jin and B. Rossi (2017). Rolling Window Selection for Out-of-Sample Forecasting With Time-Varying Parameters. *Journal of Econometrics*, **196**, 55-67.

Jimeńez, G., S. Ongena, J.-L. Peydró and J. Saurina (2017). Macroprudential Policy, Countercyclical Bank Capital Buffers, and Credit Supply: Evidence from the Spanish Dynamic Provisioning Experiments. *Journal of Political Economy*, **125**, 2126-2177.

Johnston, G. (1982). Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates. *Journal of Multivariate Analysis*, **12**, 402-414.

Linnainmaa, J.T. and M.R. Roberts (2018). The History of the Cross Section of Stock Returns. *Review of Financial Studies*, **31**, 2606-2649.

Løpez-Salido, D., J.C. Stein and E. Zkrajšek (2017). Credit-Market Sentiment and the Business Cycle. *Quarterly Journal of Economics*, **132**, 1373-1426.

Lu, Z. and O. Linton (2007). Local Linear Fitting Under Near Epoch Dependence. *Econometric Theory*, **23**, 37-70.

O'Reilly, G. and K. Whelan (2005). Has Euro-Area Inflation Persistence Change Over Time. *Review of Economics and Statistics*, **87**, 709-720.

Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica*, **57**, 1361-1401.

Robinson, P.M. (1989). Nonparametric Estimation of Time-Varying Parameters. in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, Springer Verlag: New York.

Swanson, E.T. and J.C. Williams (2014). Measuring the Effect of the Zero Lower Bound on Medium-and Longer-Term Interest Rates. *American Economic Review*, **104**, 3154-3185.

Wu, W.B. and Z. Zhou (2011). Gaussian Approximations for Non-Stationary Multiple Time Series. *Statistica Sinica*, **21**, 1397-1413.

Zhang, T. and W.B. Wu (2012). Inference of Time-Varying Regression Models. *Annals of Statistics*, **40**, 1376-1402.

Zhao, Z. and W.B. Wu (2008). Confidence Bands in Nonparametric Time Series Regression. *Annals of Statistics*, **36**, 1854-1878.

Zhou, Z. and W.B. Wu (2010). Simultaneous Inference of Linear Models with Time Varying Coefficients. *Journal of the Royal Statistical Society, Series B*, **72**, 513-531.