Analyzing Cross-Validation for Forecasting with Structural Instability

Keisuke Hirano[*] Jonathan H. Wright[†]

DRAFT: October 24, 2018

Abstract

When forecasting with economic time series data, researchers often use a restricted window of observations or downweight past observations in order to mitigate the potential effects of parameter instability. In this paper, we study the problem of selecting a window for point forecasts made at the end of the sample. We develop asymptotic approximations where there is local parameter instability of various sorts. We examine the risk properties of cross-validation as a scheme for selecting the window, and compare it to some alternatives. We also propose a quasi-Bayesian form of cross-validation that we find to have good risk properties.

[*]Department of Economics, Pennsylvania State University, 602 Kern Graduate Building, University Park PA 16802. Email: hirano@psu.edu

[†]Department of Economics, Johns Hopkins University, 3400 North Charles St., Baltimore MD 21218. Email: wrightj@jhu.edu

1 Introduction

Parameter instability is widely viewed as an obstacle to time series forecasting, and as a leading reason why econometric relationships that appear to be strong in sample can perform poorly for out-of-sample prediction. Motivated by concerns of parameter instability, researchers often want to make predictions using relatively recent data. They may do this by using a window of recent data, where it is hoped that the effects of parameter instability is negligible within the window (e.g. Pesaran and Timmermann (2007)).

For this approach, the researcher is faced with a choice of what window to use. If the true model has a discrete structural break, the optimal estimation window might include some pre-break data because the bias that this induces might be more than offset by the variance reduction, as noted by Pesaran and Timmermann (2007). The approach that we focus on in this paper is cross-validation, in the form suggested by Pesaran and Timmermann (2007). Cross-validation reserves a fraction of the data at the end of the sample for out-of-sample evaluation, and selects the start date for the estimation window as the one that minimizes the mean square forecast error over this out-of-sample period. The cross-validation method has also been analyzed by Giratis, Kapetanios, and Price (2013), and is part of the procedure proposed by Inoue, Jin, and Rossi (2017).

We analyze cross-validation and other approaches to selecting the estimation window, when the underlying model has structural instability of various forms. We model the degree of structural instability as being small enough that no test for a structural break will reject with probability approaching one. We derive the properties of cross-validation with this kind of local instability, and compare it to some other approaches such as imposing a possibly misspecified parametric model. Not surprisingly, we find that cross validation does less well in terms of forecast accuracy than estimating a parametric model if that model is correctly specified. But cross-validation can be a competitive alternative if the model is misspecified.

While cross-validation brings a degree of robustness to model misspecification, the cross-validation criterion function can be noisy, and the estimated loss can be strongly asymmetric about the minimizing value. Motivated by these observations, we also consider a variant of cross-validation that uses the entire cross-validation risk function as a pseudo-likelihood, where the "parameter" is the window length. Rather than taking the optimizer of the pseudo-likelihood, our proposed "Laplace" cross-validation scheme calculates a quasi-Bayesian posterior mean based on the cross-validation pseudo-likelihood. We find that this has attractive risk properties.

Another approach to dealing with structural instability is to downweight past observations, as in exponential smoothing, an idea that goes back to Holt (1957). This is motivated by a model of slowly changing parameters. We also analyze the properties of cross-validation and other approaches to selecting the smoothing coefficient under various forms of instability.

The plan for the remainder of this paper is as follows. We set up the basic model and forecasting methods, and obtain local asymptotic approximation results in Section 2. Numerical work based on the local asymptotics is in Section 3. Section 4 contains Monte Carlo simulations and Section 5 gives some illustrative empirical applications. Section 6 concludes.

2 Analysis

We consider a linear regression model with structural instability. The model specifies that y_{t+1} is a dependent variable, x_t is a $K \times 1$ vector of regressors, and ε_t is a sequence with mean zero and variance σ^2 such that:

$$y_{t+1} = x_t' \beta_t + \varepsilon_{t+1}, \quad t = 1, \ldots, T.$$

We will take limits as $T \to \infty$, rescaling time to the interval $[0,1]$. We make the following assumptions:

A1 $T^{-1} \sum_{t=1}^{[Tr]} x_t x_t' \to_p rM$ for $r \in [0,1]$, where M is a positive definite bounded matrix.

A2 $T^{-1/2}\sum_{t=1}^{[Tr]} x_t \varepsilon_{t+1} \to \Omega^{1/2} V(r)$, where Ω is a symmetric $K \times K$ matrix, with symmetric square root $\Omega^{1/2}$, and $V(r)$ is a standard K-dimensional Brownian motion for $r \in [0,1]$.

These assumptions essentially require that x_t is stationary but allows the errors to have serial correlation and conditional heteroskedasticity. If the errors are homoskedastic and serially uncorrelated, then $\Omega = \sigma^2 M$. Although we denote the dependent variable with a $t+1$ subscript, our assumptions allow for k-step ahead forecasting for arbitrary fixed k.

The parameter β_t can change over time. In order to develop large-sample approximations, we suppose that β_t, appropriately scaled, converges to a process in the following sense:

$$T^{1/2}\beta_{[Tr]} \to M^{-1}\Omega^{1/2} H(r),$$

where $H(r)$ is either a deterministic or stochastic function of $r \in [0,1]$, which represents the path of the time-varying parameter in a rotated space. To develop concrete results we consider three specific models for the time-varying parameter β_t:

M1 A one-time structural break at date $[cT]$. Before the break, $\beta_t = \beta_{PRE} = T^{-1/2}\Omega^{1/2} M^{-1}\mu_0$ while after the break $\beta_t = \beta_{POST} = T^{-1/2}\Omega^{1/2} M^{-1}\mu_1$. The local asymptotic sequence makes the degree of structural instability small. In this model,

$$T^{1/2}\beta_{[Tr]} \to M^{-1}\Omega^{1/2}\mu_1 + M^{-1}\Omega^{1/2}\mu 1(r < c),$$

where $\mu = \mu_0 - \mu_1$. Here $H(r) = \mu_1 + \mu 1(r < c)$.

M2 A random walk parameters model:

$$\beta_t = \beta_{t-1} + \delta_T v_t,$$
$$\beta_0 = 0$$

where $v_t \sim (0, M^{-1}\Omega M^{-1})$, and $\delta_T = \mu/T$, for a scalar μ.[1] In this model,

$$T^{1/2}\beta_{[Tr]} \to \mu M^{-1}\Omega^{1/2}W_v(r),$$

where $H(r) = \mu W_v(r)$, and $W_v(r)$ is a standard K-dimensional Brownian motion.

M3 A Poisson breaks model of the sort considered by Koop and Potter (2007). In this model, there are breaks with intensity λ/T, and the break dates are $t_1, t_2, \ldots$. Let β_{t_i-1,t_i} denote the constant value of the parameter vector between t_{i-1} and t_i. We assume that $\beta_{t_{i-1},t_i} = \beta_{t_{i-2},t_{i-1}} + T^{-1/2}\mu\Omega^{1/2}M^{-1}\xi_i$ where μ is a scalar and ξ_i is iid standard normal. Here we have $T^{1/2}\beta_{[Tr]} \to M^{-1}\Omega^{1/2}H(r)$ where $H(r) = \mu\sum_{j=1}^{C(r)}\xi_j$, where $C(r)$ is the number of Poisson jumps by time r.

Define a rolling window estimator using data from date h up to date t_1 as:

$$\hat{\beta}_{(h,t_1)} = \left(\sum_{t=h}^{t_1} x_t x_t'\right)^{-1} \sum_{t=h}^{t_1} x_t y_t.$$

More generally, we could consider weighted least squares estimators with weights $w = \{w_t\}$, which we denote

$$\hat{\beta}_{(w,t_1)} = \left(\sum_{t=1}^{t_1} w_t x_t x_t'\right)^{-1} \sum_{t=1}^{t_1} w_t x_t y_t.$$

The rolling window estimator corresponds to $w_t = 1(t \geq h)$, but we can consider more general weighting schemes. We require that the weights converge in the following sense:

$$w_{[Tr]} \to \omega(r),$$

for some function $\omega(r)$. For example, if we define the rolling window estimator as $w_t = 1(t \geq$

[1] If the errors are homoskedastic and not serially correlated, this is the same parametrisation chosen by Nyblom (1989). This ensures that the parameter innovations are uncorrelated if the regressors are rotated to be uncorrelated.

$[\eta T]$), we have

$$w_{[Tr]} \to 1(r \geq \eta).$$

Another weighting scheme is

$$w_t = \left(1 - \frac{\eta}{T}\right)^{t_1 - t},$$

which amounts to exponential smoothing. If $t_1 = [r_1 T]$, this satisfies the convergence requirement:

$$w_{[Tr]} \to \exp(-\eta(r_1 - r)).$$

In model M1, M2 or M3:

$$T^{1/2}(\hat{\beta}_{(w,[Tr]-1)} - \beta_{[Tr]}) \to_d M^{-1}\Omega^{1/2}\left[\frac{\int_0^r \omega(s)H(s)ds}{\int_0^r \omega(s)ds} - H(r)\right] + M^{-1}\Omega^{1/2}\frac{\int_0^r \omega(s)dV(s)}{\int_0^r \omega(s)ds} \equiv M^{-1}\Omega^{1/2}G(r,\eta).$$

The limiting function $G(r,\eta)$ depends on the model and weighting scheme. With the rolling weighting scheme:

$$G(r,\eta) = \frac{1}{r-\eta}\int_\eta^r H(s)ds - H(r) + \frac{1}{r-\eta}(V(r) - V(\eta)). \tag{2.1}$$

With the one-time break, this further simplifies as

$$\frac{1}{r-\eta}\int_\eta^r H(s)ds - H(r) = \frac{c-\eta}{r-\eta}\mu 1(\eta < c < r).$$

With exponential weighting, the limit is instead:

$$G(r,\eta) = \frac{\eta\int_0^r \exp(-\eta(r-s))H(s)ds}{1-\exp(\eta r)} - H(r) + \frac{\eta\int_0^r \exp(-\eta(r-s))dV(s)}{1-\exp(\eta r)}. \tag{2.2}$$

Suppose that the forecaster observes x_T and wants to predict y_{T+1}. The researcher does this by estimating the model using an estimator $\hat{\beta}_{(w,T)}$. The loss function is squared error loss:

$$L(y_{T+1}, \hat{y}_{T+1}) = \left(y_{T+1} - \hat{y}_{T+1}\right)^2.$$

where $\hat{y}_{T+1} = x_T' \hat{\beta}_{(w,T)}$. Since the forecaster cannot do anything about the future shock ϵ_{T+1}, we subtract σ^2 from the loss to obtain the regret loss. The rescaled regret risk[2] when using the parameter estimator $\hat{\beta}_{(w,T)}$ is:

$$R = T\left[E\left((y_{T+1} - x_T'\hat{\beta}_{(w,T)})^2\right) - \sigma^2\right] = TE\left(((\hat{\beta}_{(w,T)} - \beta)' x_T)^2\right) \tag{2.3}$$

Then, as $T \to \infty$, we have $R \to R^*$, where:

$$R^* = E\left(\left[\frac{\int_0^1 \omega(s)H(s)ds}{\int_0^1 \omega(s)ds} - H(1)\right]' \Lambda \left[\frac{\int_0^1 \omega(s)H(s)ds}{\int_0^1 \omega(s)ds} - H(1)\right]\right) + E\left(\left[\frac{\int_0^1 \omega(s)dV(s)}{\int_0^1 \omega(s)ds}\right]' \Lambda \left[\frac{\int_0^1 \omega(s)dV(s)}{\int_0^1 \omega(s)ds}\right]\right)$$

and $\Lambda = \Omega^{1/2} M^{-1} \Omega^{1/2}$. If $\Omega = \sigma^2 M$, this simplifies to:

$$R^* = \sigma^2 E\left(\left[\frac{\int_0^1 \omega(s)H(s)ds}{\int_0^1 \omega(s)ds} - H(1)\right]' \left[\frac{\int_0^1 \omega(s)H(s)ds}{\int_0^1 \omega(s)ds} - H(1)\right]\right) + \sigma^2 E\left(\left[\frac{\int_0^1 \omega(s)dV(s)}{\int_0^1 \omega(s)ds}\right]' \left[\frac{\int_0^1 \omega(s)dV(s)}{\int_0^1 \omega(s)ds}\right]\right). \tag{2.4}$$

With the rolling weighting scheme and fixed η, it simplifies further to:

$$R^* = \sigma^2 E\left(\left[\frac{1}{1-\eta}\int_\eta^1 H(s)ds - H(1)\right]' \left[\frac{1}{1-\eta}\int_\eta^1 H(s)ds - H(1)\right]\right) + \frac{\sigma^2 K}{1-\eta}. \tag{2.5}$$

[2]We consider the unconditional mean square forecast error, though it would be possible in principle to consider the mean square forecast error conditional on x_T.

In model M1, this reduces to:

$$R^* = \sigma^2 \mu' \mu \max\left(0, \frac{c-\eta}{1-\eta}\right)^2 + \frac{\sigma^2 K}{1-\eta}. \tag{2.6}$$

In model M2, it instead reduces to:

$$R^* = \frac{\sigma^2 \mu^2 K(1-\eta)}{3} + \frac{\sigma^2 K}{1-\eta}, \tag{2.7}$$

With the exponential weighting scheme and fixed η, equation (2.4) instead simplifies to:

$$R^* = \mu^2\sigma^2 E\left(\left[\frac{\eta\int_0^1 \omega(s)\beta(s)ds}{1-\exp(-\eta)} - \beta(1)\right]' \left[\frac{\eta\int_0^1 \omega(s)\beta(s)ds}{1-\exp(-\eta)} - \beta(1)\right]\right) + \frac{\eta K\sigma^2(1-\exp(-2\eta))}{2(1-\exp(-\eta))^2}. \tag{2.8}$$

In model M1, this further reduces to:

$$R^* = \mu'\mu\sigma^2\left[\frac{\exp(-\eta)(\exp(\eta c)-1)}{1-\exp(-\eta)}\right]^2 + \frac{\eta K\sigma^2(1-\exp(-2\eta))}{2(1-\exp(-\eta))^2}. \tag{2.9}$$

In model M2, it instead reduces to:

$$R^* = \mu^2 K\sigma^2\left[\frac{e^{-2\eta}}{(1-\exp(-\eta))^2}\frac{e^{2\eta}-4e^{\eta}+2\eta+3}{2\eta}\right] + \frac{\eta K\sigma^2(1-\exp(-2\eta))}{2(1-\exp(-\eta))^2}. \tag{2.10}$$

It is important to note that these risk functions are asymmetric, and the degree and even direction of asymmetry depends on the magnitude of the structural break, μ. Figure 1 plots the limiting risk function in equations (2.6) and (2.7), in which the asymmetry can clearly be seen.

2.1 Rolling Weighting Scheme

Consider first the case of rolling weighting, where h is the start date of the window. There are a number of schemes that have been proposed in the literature including:

1. We could use the least squares estimator of the break date, which is also the pseudo-Gaussian maximum likelihood estimator assuming homoskedastic and serially uncorrelated errors: $\hat{h}_1 = \arg\min_{h_{\min} \le h \le h_{\max}} S(h)$ where

$$S(h) = \sum_{t=1}^{h-1} (y_{t+1} - x_t'(\sum_{s=1}^{h} x_s x_s')^{-1} \sum_{s=1}^{h} x_s y_s)^2 + \sum_{t=h}^{T} (y_{t+1} - x_t'(\sum_{s=h+1}^{T} x_s x_s')^{-1} \sum_{s=h+1}^{T} x_s y_s)^2$$

   where $h_{\min} = [T\eta_{\min}]$ and $h_{\max} = [T\eta_{\max}]$.

2. As a variant on this, we could take the least squares estimates of μ and c, $\hat{\mu}_{LS}$ and $\hat{c}_{LS} = \hat{h}_1/T$, plug these estimates into the risk function (equation (2.6)) and minimize, giving a window start date of:

$$\hat{h}_2 = \left[T \arg\min_{\eta} \hat{\mu}_{LS}' \hat{\mu}_{LS} \max\left(0, \frac{\hat{c}_{LS} - \eta}{1-\eta}\right)^2 + \frac{K}{1-\eta} \right].$$

   Note that $\hat{h}_2 \le \hat{h}_1$. One might want to use a little data before the estimated break to reduce bias, but the researcher has no motivation not to use all the data after the estimated break. A finite-sample version of this method was proposed by Pesaran and Timmermann (2007) and referred to as the *tradeoff* method, in recognition of the bias-variance trade-off. We henceforth refer to it as the tradeoff method.

3. Pesaran and Timmermann (2007) proposed selecting the window as $\hat{h}_3 = \arg\min_{1 \le h \le h_{\max}} C(h)$ where $C(h)$ is the following cross-validation criterion:

$$C(h) = \sum_{t=r}^{T-1} (y_{t+1} - x_t' \hat{\beta}_{(h,t)})^2,$$

where $w = [\rho T]$, $\hat{\beta}_{(h,t)}$ is the rolling window estimator with start date h, and $h_{\max} = [T(\rho - \zeta)]$ for some fixed $\zeta > 0$.

4. A variant of this cross-validation approach, also proposed by Pesaran and Timmermann (2007) is to select the window as $\hat{h}_4 = \arg\min_{1 \le h \le \min(\hat{h}_1, h_{\max})} C(h)$, in other words only to consider rolling window start dates that come on or before the estimated break date.

The following proposition gives the limiting distributions of these estimators:

Proposition 2.1. *Using least squares (method 1) gives an estimate of the break fraction ($\hat{h}_1/T$) that converges to:*

$$\eta_{LS} = \arg\max_{\eta_{\min} < \eta < \eta_{\max}} \frac{Q(\eta)'\Lambda Q(\eta)}{\eta} + \frac{(Q(1) - Q(\eta))'\Lambda(Q(1) - Q(\eta))}{1-\eta} \tag{2.11}$$

where $Q(\eta) \equiv V(\eta) + \int_0^\eta H(s)ds$. In the model with the one-time break, $\int_0^\eta H(s)ds = \mu\min(c,\eta)$.

The tradeoff (method 2) estimate of the break fraction ($\hat{h}_2/T$) converges to:

$$\eta_{TO} = \arg\min_\eta \mu_{LS}'\mu_{LS} \max\left(0, \frac{\eta_{LS} - \eta}{1-\eta}\right)^2 + \frac{K}{1-\eta}. \tag{2.12}$$

where the least squares estimate of μ converges to $\mu_{LS} = \frac{Q(\eta_{LS})}{\eta_{LS}} - \frac{Q(1)-Q(\eta_{LS})}{1-\eta_{LS}}$.

The cross-validated estimate (method 3) of the break fraction ($\hat{h}_3/T$) converges to:

$$\eta_{CV} = \arg\min_{0 \le \eta \le \rho - \zeta} \int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr - 2\int_\rho^1 G(r,\eta)'\Lambda dV(r). \tag{2.13}$$

with $G(r,\eta)$ given by equation (2.1).

Considering only windows starting on or before $\hat{h}_1$, the cross-validated estimate (method 4) of the break fraction ($\hat{h}_4/T$) converges to:

$$\eta_{CV2} = \arg\min_{0 \le \eta \le \min(\rho-\zeta, \eta_{LS})} \int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr - 2\int_\rho^1 G(r,\eta)'\Lambda dV(r). \tag{2.14}$$

Proof: The proof of equation (2.11) is from Elliott and Müller (2007). Equation (2.12) follows immediately. The cross-validation estimate of the risk function is:

$$C(h) = \sum_{t=r}^{T-1}(y_{t+1} - x_t'\hat{\beta}_{(t)})^2 = \sum_{t=r}^{T-1}(\varepsilon_{t+1} - (\hat{\beta}_{(t)} - \beta_{t+1})'x_{t+1})^2$$
$$= \sum_{t=r}^{T-1}\varepsilon_{t+1}^2 + \sum_{t=r}^{T-1}(\hat{\beta}_{(t)} - \beta_{t+1})'x_{t+1}x_{t+1}'(\hat{\beta}_{(t)} - \beta_{t+1})$$
$$-2\sum_{t=r}^{T-1}(\hat{\beta}_{(t)} - \beta_{t+1})'x_{t+1}\varepsilon_{t+1}.$$

Hence,

$$C(h) - \sum_{t=T-r}^{T-1}\varepsilon_{t+1}^2 \to_d \int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr - 2\int_\rho^1 G(r,\eta)\Lambda dV(r).$$

implying equations (2.13) and (2.14). ■

It is important that method 3 evaluates the risk over a fixed window from r to $T-1$. That's why the term $\sum_{t=r}^{T-1}\varepsilon_{t+1}^2$ drops out of the minimization over h (i.e. $\arg\min_h C(h) = \arg\min_h C(h) - \sum_{t=r}^{T-1}\varepsilon_{t+1}^2$). We can imagine schemes in which the sample for evaluation changes as a function of h. But then $\Sigma\varepsilon_{t+1}^2$ does not drop out and in fact because it is $O_p(T)$ whereas the other terms in $C(h)$ are $O_p(1)$, it is the only term that matters asymptotically. This is effectively shown in Ploberger and Krämer (1990) who show that the CUSUM of squares tests has only trivial local asymptotic power in a $\sqrt{T}$-neighborhood.

2.2 Laplace Cross-validation

In our time-series setting, the cross-validation criterion function $C(h)$ is based on out-of-sample evaluations over a limited time span, and may be fairly noisy.[3] Moreover, as we saw in Figure 1, the underlying risk function being approximated by cross-validation can be asymmetric.

Motivated by these observations, we consider a pseudo-Bayesian alternative to minimization of

[3] An alternative approach to accounting for estimation error in the cross-validation criterion was proposed by Lei (2017) in a cross-sectional setting.

$C(h)$. Consider $C(h)/\hat{\sigma}^2$, the cross-validation function scaled by $\hat{\sigma}^2$ (a consistent estimate of the variance). We treat $L(h) = \exp(-0.5C(h)/\hat{\sigma}^2)$ as a pseudo-likelihood in terms of the parameter h, and then take:

$$\hat{h}_5 = \frac{\int hL(h)dh}{\int L(h)dh}$$

as the estimate of the window start date. This is a pseudo-posterior mean, with a flat prior for the rolling window start date. It is based on the entire cross-validation function, not just the minimizer of $C(h)$. It takes account of the asymmetry of the risk function, which was shown in Figure 1 to be potentially important, whereas simply minimizing $C(h)$ does not. We call this the Laplace cross-validation estimator. From the Continuous Mapping Theorem, $\frac{\hat{h}_5}{T} \to_d \eta_{CVL}$, where

$$\eta_{CVL} = \frac{\int_0^{\rho-\zeta} \eta \exp[-0.5\int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr + \int_\rho^1 G(r,\eta)'\Lambda dV(r)]d\eta}{\int_0^{\rho-\zeta} \exp[-0.5\int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr + \int_\rho^1 G(r,\eta)'\Lambda dV(r)]d\eta}. \quad (2.15)$$

with $G(r,\eta)$ given by equation (2.1).

Laplace-type estimators and inference procedures have been proposed for a number of econometric settings, including Chernozhukov and Hong (2003), Christensen, Chen, and Tamer (2017), and Inoue and Shintani (2018). Our proposal has a similar form, but its properties are quite distinct because the criterion function does not having a limiting log-quadratic form under our local asymptotics. As a consequence the limiting distribution of our window estimator $\hat{h}_5$ is non-Gaussian.

Laplace cross-validation is also related to Bayesian Model Averaging. Geweke and Amisano (2011) use the one-step ahead predictive likelihood as weights in Bayesian Model Averaging where there is uncertainty about which variables to include in the model. This is similar to Laplace cross-validation, except that Laplace cross-validation picks a single window, and so does not involve any kind of model averaging, and we are considering uncertainty about the window length rather than the variables to be included.

2.3 Exponential Weighting Scheme

Consider the case of the exponential weighting scheme. Here we can also select η by cross validation. This gives the estimate $\hat{\eta}_1 = \arg\min_{\eta\geq 0} C(\eta)$ where:

$$C(\eta) = \sum_{t=r}^{T-1} (y_{t+1} - x_t'\hat{\beta}_{(w,t)})^2$$

$r = [\rho T]$ and $\hat{\beta}_{(w,t)}$ is the exponential weighted estimator with parameter η. The cross-validated estimate of η converges to:

$$\eta_{CV} = \arg\min_{\eta\geq 0} \int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr - 2\int_\rho^1 G(r,\eta)'\Lambda dV(r).$$

with $G(r,\eta)$ given by equation (2.2).

Alternatively, we can pick η by Laplace cross-validation. Letting $g(\eta)$ be a prior for the exponential smoothing parameter, the cross-validated Laplace estimator is:

$$\hat{\eta}_2 = \frac{\int_0^\infty \eta C(\eta)g(\eta)d\eta}{\int_0^\infty \eta g(\eta)d\eta}.$$

This converges to:

$$\eta_{CVL}^{EW} = \frac{\int_0^\infty g(\eta)\eta\exp[-0.5\int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr + \int_\rho^1 G(r,\eta)'\Lambda dV(r)]d\eta}{\int_0^\infty g(\eta)\exp[-0.5\int_\rho^1 G(r,\eta)'\Lambda G(r,\eta)dr + \int_\rho^1 G(r,\eta)'\Lambda dV(r)]d\eta}. \tag{2.16}$$

The parameter μ can also be estimated by pseudo-Gaussian maximum likelihood in model M2 and can be used as η in the exponential smoothing scheme. Maximum likelihood estimation of μ can be done by way of the Kalman filter. Model M2 with $x_t = 1$ is equivalent to an MA(1) specification (Shephard, 1993; Stock and Watson, 1998) with an MA unit root of $1 - \frac{\mu}{T} + o(\frac{1}{T})$. Davis and Dunsmuir (1996) and Müller and Wang (2017) give the asymptotic distribution of the maximum likelihood (ML) estimator in this case, but this only applies in the case of the random walk parameter model.

3 Numerical Work

3.1 Rolling Window

Figures 2-4 give the local asymptotic distributions from equations (2.11)-(2.13) and (2.15) in models M1, M2 and M3. The different estimators are labeled LS, TO, CV and CVL, respectively. The settings are $\Omega = \sigma^2 M$, $K = 1$, $\sigma^2 = 1$, $M = 1$, $\eta_{\min} = 0.15$, $\eta_{\max} = 0.85$ and $\omega = 0.9$ in all cases. In model M1, we set $\mu = 10$ and consider $c = 0.25$, 0.5, 0.75. The risk-minimizing fixed η from equation (2.6) is 0.24, 0.49 and 0.74 respectively. In model M2, we set $\mu = 1$, 5, 10. The risk-minimizing η from equation (2.7) is 0, 0.65 and 0.83 respectively. In model M3, we set $\mu = 3$ and $\lambda = 1, 2, 3$. The risk-minimizing η is 0.43, 0.59 and 0.75, respectively.

In Figure 2, the LS estimates are tightly concentrated around the true break fraction, even though not consistent. So are the TO estimates. The CV estimates are more dispersed, and the CVL estimate is shrunk towards the middle of the sample.

In Figures 3 and 4, the single break model is misspecified. The LS estimate is not centered anywhere in particular, which is not surprising. The CV estimator is also very dispersed. It has a distribution with some point mass at 0 and $\eta_{\max}$. The CVL estimator is centered around the middle of the sample, but the higher is λ, the more likely CVL is to pick a larger η. It is also influenced by the asymmetry of the loss function.

3.2 Exponential Weighting

With the exponential weighting scheme, model M2 is correctly specified and models M1 and M3 are misspecified. Figures 5-7 give the local asymptotic distributions of the CV and CVL estimators of η in models M1-M3. The settings are the same as for the rolling window, but the prior $g(\eta)$ specifies that the midpoint of the exponential weighting function[4] is uniform on [0.5,0.95].

In Figure 6, the model is correctly specified. The well-known pileup problem applies to the CV estimate in that the estimate of η has point mass at zero. The CVL estimator does not have any pileup, as expected given that it is a pseudo-posterior mean.

In Figures 5 and 7, the model is misspecified. The mean of the CV estimator is close to the risk-minimizing η. The CV estimate has point mass at zero, but the pileup probability is declining in the break-date parameter c (in Figure 5) and in the jump intensity parameter λ (in Figure 7). The CVL estimator is shrunk towards the prior mean, and is also influenced by the asymmetry of the loss function. Again, the CVL estimator does not exhibit any pileup.

3.3 Local Asymptotic Risk

Table 1 gives numerical calculations of local asymptotic risk. For each draw of $\beta(.)$ and $V(.)$ we compute the chosen η and hence the weight function $\omega(s)$. We then plug these into:

$$\mu^2\left(\left[\frac{\int_0^1 \omega(s)H(s)ds}{\int_0^1 \omega(s)ds} - H(1)\right]^2\right) + \left(\left[\frac{\int_0^1 \omega(s)dV(s)}{\int_0^1 \omega(s)ds}\right]^2\right), \tag{3.1}$$

and average across all the draws. Table 1 reports results for both the rolling window (where model M1 is correctly specified) and exponential weighting (where model M2 is correctly specified).

The tradeoff rolling window gives smaller risk than the LS estimate of the break date, in all the cases considered here regardless of whether the model is correctly specified or not. Ordinary cross-validation gives larger risk than the correctly specified LS estimate, in model M1. But

[4]This is $\bar{r} : \frac{\int_{\bar{r}}^1 \exp(-\eta(1-r))dr}{\int_0^1 \exp(-\eta(1-r))dr} = 0.5$.

cross-validation gives lower asymptotic risk than misspecified least squares estimation in models M2 and M3. Moreover, the Laplace form of cross-validation gives lower asymptotic risk than ordinary cross-validation in most cases considered here. Overall, these results indicate that the Laplace form of cross-validation has attractive asymptotic risk properties.

In this numerical work, the LS procedure always estimates a break. Another variant of course is to do a sup-F pretest, and use the whole sample if no break is found. This doesn't change the main conclusions about the local asymptotic performance of the different procedures.

Within ordinary or Laplace cross-validation, it would also be possible to compute the choices of ρ and ζ that minimize average risk in model M1, averaging over possible choices of the break date, c. This will, of course, depend on the magnitude of the break, μ.

4 Monte Carlo Simulations

We report some Monte Carlo evidence of the performance of the different procedures in a model with instability in the conditional mean.

The design is:

$$y_t = b_t + u_t$$

where u_t is an AR(1) process with coefficient ϕ and iid normal innovations with mean zero and variance scaled such that 2π times the spectral density of u_t at frequency zero is 1, $t = 1, ...T$. We have 10 different specifications for b_t:

1. $b_t = 0$.
2. $b_t = 10T^{-1/2}1(\frac{t}{T} > 0.25)$
3. $b_t = 10T^{-1/2}1(\frac{t}{T} > 0.5)$
4. $b_t = 10T^{-1/2}1(\frac{t}{T} > 0.75)$

5. $b_0 = 0$, $b_t = b_{t-1} + T^{-1}\xi_t$ for $t \geq 1$ where ξ_t is iid standard normal.

6. $b_0 = 0$, $b_t = b_{t-1} + 5T^{-1}\xi_t$ for $t \geq 1$ where ξ_t is iid standard normal.

7. $b_0 = 0$, $b_t = b_{t-1} + 10T^{-1}\xi_t$ for $t \geq 1$ where ξ_t is iid standard normal.

8. $b_0 = 0$ and there are Poisson jumps with an intensity $\frac{1}{T}$ at each of which the parameter increases by $3T^{-1/2}\xi$ where ξ is standard normal.

9. As in (8) except that the Poisson intensity is $\frac{2}{T}$.

10. As in (9) except that the Poisson intensity is $\frac{3}{T}$.

The first design has no structural break, designs 2-4 have a discrete break, designs 5-7 have a random parameter and designs 8-10 feature Poisson jumps. The models are chosen to be directly comparable to the local asymptotics that we considered earlier, to let us assess whether the local asymptotics provide a useful guide to finite sample performance. In these models, we then forecast y_{T+1} and consider various rolling windows:

1. A window starting at the least squares estimate of a single break date.

2. A rolling window selected by the break date estimation method of Bai and Perron (1998) with a number of breaks between 0 and 5 determined by the BIC. If no break is found, the whole sample is used for forecasting. If one or more breaks are found, then the window starting on the last estimated break date is used for forecasting.

3. The tradeoff rolling window based on the least squares estimate in (1).

4. Cross-validation using the last 10 percent of the sample for evaluation considering all possible window start dates ($\hat{h}_3$).

5. Cross-validation using the last 10 percent of the sample for evaluation considering only start dates before the last estimated break date ($\hat{h}_4$).

6. The Laplace cross-validation scheme, as proposed in subsection 2.2.

7. The Laplace cross-validation scheme, considering only start dates before the last estimated break date.

8. The method for window selection of Inoue, Jin, and Rossi (2017). This uses cross-validation without an estimated break date to select an initial window, uses local linear regression to estimate the parameter at the end of the sample, and then uses the window that gets closest to this local linear estimate.[5]

9. Exponential weighting using the maximum likelihood estimate of a random walk plus noise model.

10. The corresponding tradeoff exponential weighting, selecting η to minimize equation (2.10) where the parameters μ and σ^2 are repliaced by their maximum likelihood estimates.

11. Exponential weighting using cross-validation.

12. Exponential weighting with Laplace cross-validation.

With each of these methods for selecting the window start date, we computed the mean square error of the forecast as a predictor of b_{T+1}. Let $\hat{y}_{i,T+1}$ denote the forecast of the $T+1$th observation using a rolling window in the ith replication and let $b_{i,T+1}$ denote the draw of b_{T+1} in this replication. The mean square forecast error is:

$$MSFE = T\sum_{i=1}^{R}\left(\hat{y}_{i,T+1}-b_{i,T+1}\right)^2 \tag{4.1}$$

where R is the number of Monte Carlo replications, which we set to 5,000. We scale the mean square forecast error by the sample size to make it comparable to the rescaled risk function in equation (2.3). Note also that we are viewing the forecasts as predictors of b_{T+1}rather than y_{T+1}.

[5] Concretely, we are using method OptR1 in the notation of Inoue, Jin, and Rossi (2017).

This is so as to remove the effect of noise that is going to be the same for all methods and just serves to obscure the difference between them. The performance of these methods has been evaluated in Monte Carlo simulations in existing work, including Pesaran, Pick, and Pranovich (2013), except that the assessment of the Laplace cross-validation and exponential weighting cross-validation schemes are new.

The results are reported in Table 2 with $\phi = 0$ (meaning that the errors are iid) and sample sizes $T = 100$ and $T = 200$. Results with $\phi = 0.7$ and the same sample sizes are shown in Table 3.

In the models with a single discrete break, the rolling window methods that estimate this break date (or use the tradeoff method) give smallest MSFE, although the efficiency losses from using cross-validation to select the window can be small. In the models with slow-moving parameters, exponential weighting with maximum likelihood estimation does best, but the cross-validation rolling window does quite well too. The Laplace forms of cross-validation gives smaller loss than ordinary cross-validation in most cases. In the models with Poisson jumps, one of the two versions of Laplace cross-validation gives the smallest MSFE with both sample sizes, and with both iid and AR(1) errors. In general, the conclusions are consistent with the local asymptotic risk calculations of the previous section, and are quite consistent across sample size and error persistence. However, persistence of the errors substantially degrades the performance of the maximum-likelihood estimator of the random walk plus noise model and the associated tradeoff estimator.

5 Empirical Applications

5.1 Inflation Forecasting

We consider an illustrative application to the use of the Phillips curve in inflation forecasting. A tradeoff between unemployment and inflation is central to new Keynesian macroeconomic models. However, there are widely thought to have been changes to inflation dynamics over the last few decades, with the Phillips curve having flattened and inflation having become less persistent (e.g. Stock and Watson (2010)).

We define annualized inflation in quarter t as π_t and define k-period inflation as $\pi_t^{(k)} = k^{-1}\Sigma_{i=i}^{k}\pi_{t+i-1}$. Following Stock and Watson (2009), we consider forecasting $k-$period inflation as of time t from the regression:

$$\pi_{t+k}^{(k)} - \pi_t = \beta_0 + \beta_1(\pi_t - \pi_{t-1}) + \beta_2(u_t - \bar{u}_t) + \varepsilon_t \tag{5.1}$$

where u_t is the average civilian unemployment rate in quarter t and $\bar{u}_t$ is the CBO measure of NAIRU in that quarter. In this regression, we select the rolling window using LS estimation of break dates with a single break, LS estimation with the number of breaks set by BIC, the tradeoff method, cross-validation, Laplace cross-validation, and the method of Inoue, Jin, and Rossi (2017).

Our data are quarterly from 1959:Q1 to 2018:Q2, and we measure inflation by either the total or core PCE price index. We assess the methods on the basis of out-of-sample forecast accuracy, with the first forecast made using rolling windows from the first 80 quarters of data. For total PCE inflation, we use the real-time data from the Federal Reserve Bank of Philadelphia. Unfortunately, core PCE inflation was not released at all before 1996, and so we have to use *ex-post* revised data for that.

The results are shown in Table 4. No window selection method does best in all cases. Cross-validation is however competitive, and in nearly all cases, the Laplace form of cross-validation gives a smaller root mean square error than ordinary cross-validation. Laplace cross-validation seems to be especially beneficial at longer forecasting horizons. For both inflation measures, Laplace cross-validation gives lower root mean square prediction error at both the 3 and 4 quarter horizons. Inflation forecasting is hard, and simple univariate benchmarks generally do better than Phillips curve forecasts in terms of out-of-sample forecast accuracy (e.g. Faust and Wright (2013)). However, within Phillips curve forecasts, the Laplace cross-validation method for choosing the rolling window appears to work relatively well.

5.2 Fitting Autoregressions to a Large Dataset

As another illustration, we took the quarterly version of the database of McCracken and Ng (2016) and considered all 210 series for which data are available from 1959Q1-2017Q4. We applied the transformation to induce stationarity as given in that paper to each series, and then forecasted each series by an AR(1), selecting the rolling window by the same 8 different methods as in the previous subsection. For each series, we computed the one-quarter-ahead out-of-sample root mean square prediction error using the different methods for selecting rolling windows, relative to that from simply using the full sample. The exercise is meant to illustrate the relative performance of different methods for choosing windows in a generic environment.

Table 5 reports the averages and 25th, 50th, and 75th percentiles of the relative root mean square error across all 210 series, for each of 8 methods. Table 5 also reports the proportion of series for which the best forecast, in an out-of-sample root mean square error sense, is given by each of the 8 methods, and it reports the proportion of series for which the chosen window is the entire sample, meaning that the relative root mean square error is exactly equal to 1.

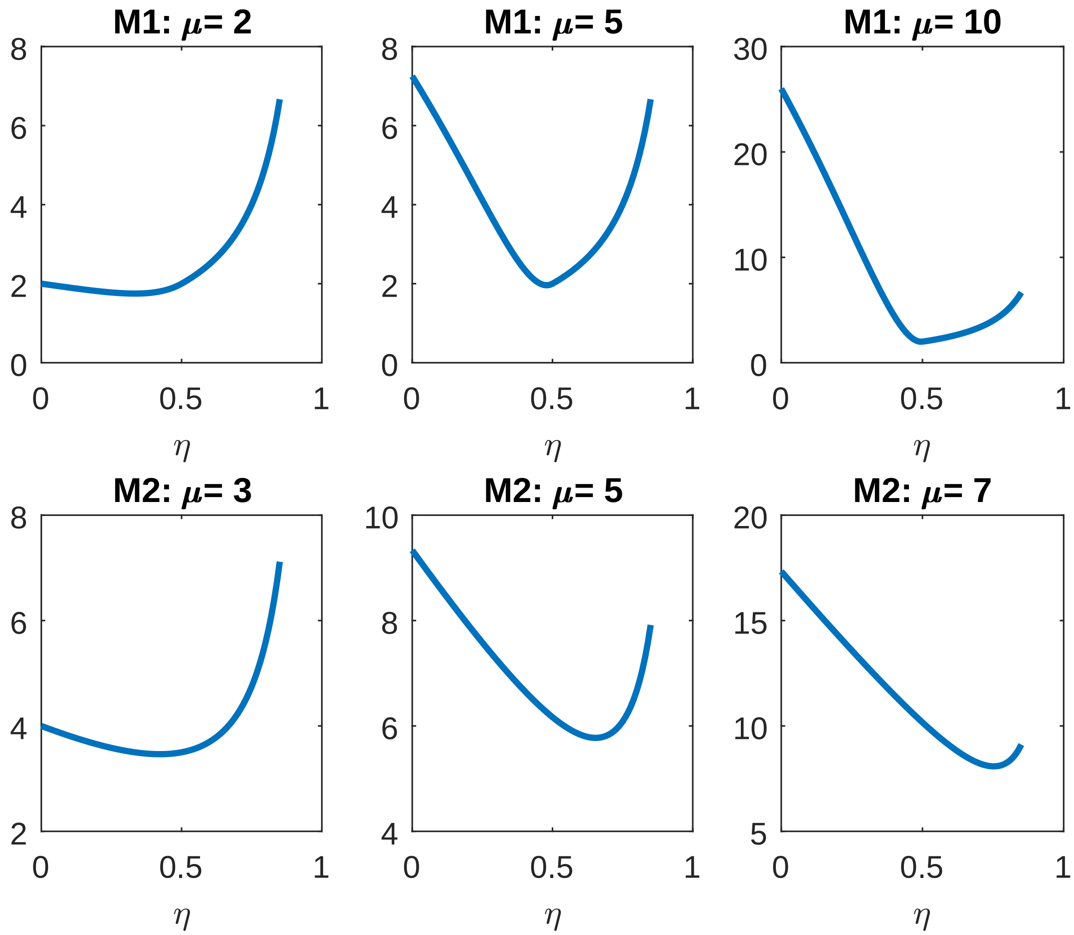
LS estimation with the BIC number of breaks allows for the information criterion to pick anything from 0 to 5 breaks. For 68 percent of the series, it picks no breaks at all, meaning that the relative root mean square prediction error is exactly 1. The cross-validation methods allow for the possibility of using the whole sample as the estimation window, but this is not actually chosen for any of the 210 series. Therefore only LS estimation with the BIC number of breaks gives a positive probability of the relative root mean square error being exactly 1.

For all 8 methods, the average relative root mean square error across all series is around 1, indicating that picking rolling windows does not typically help much, and might well actually worsen forecasting performance. But the two Laplace cross-validation methods are the only two approaches that give average relative root mean square error that are very slightly below one. For nearly half of the series, the best forecasts are given by LS estimation with the BIC number of dates. But cross-validation methods are also competitive and one of the two forms of Laplace cross-validation are optimal for about a quarter of the series.

6 Conclusions

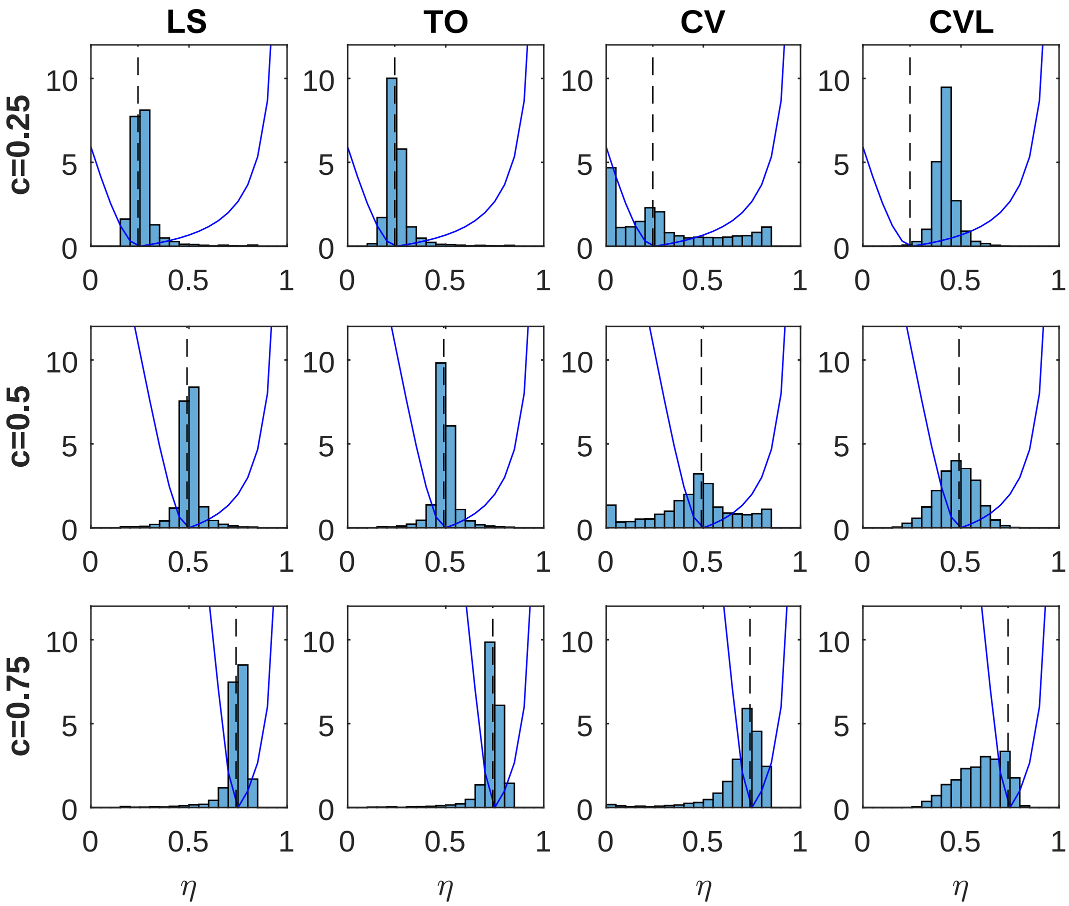
In this paper we have considered the risk properties of various methods for selecting the estimation window to be used for point forecasting in the linear regression model in the presence of local parameter instability of various forms. Cross-validation, and especially a quasi-Bayesian form of cross-validation are found to have good risk properties. These predictions are confirmed in Monte-Carlo simulations and are also borne out in out-of-sample accuracy in some illustrative empirical applications.

Figure 1: Asymptotic Loss as a Function of η in Models M1 and M2



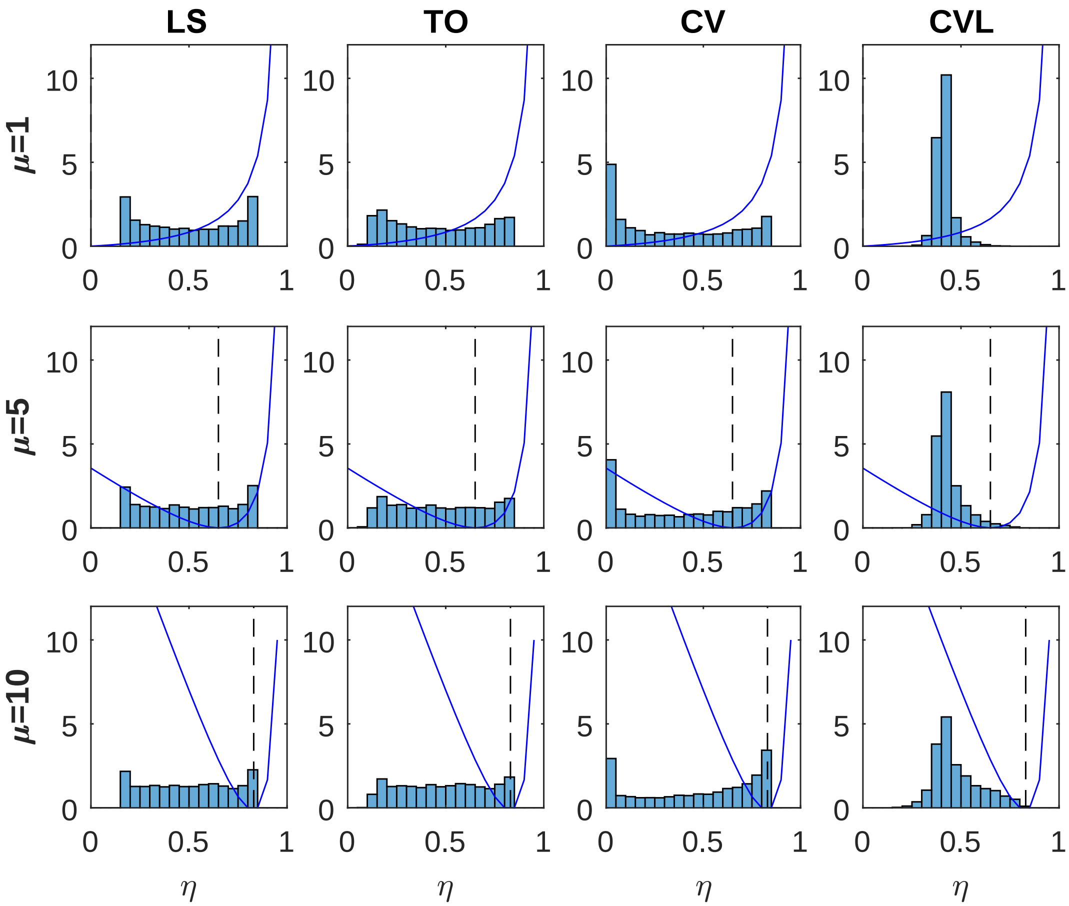
Notes: The figures show the risk functions in model M1 (Equation (2.6)) and in model M2 (Equation (2.7)), as functions of η for $K = 1$, $\sigma^2 = 1$ and selected choices of μ. In model M1, we set $c = 0.5$.

Figure 2: Simulated Local Asymptotic Distributions of Estimates of η in Rolling Window with Model M1



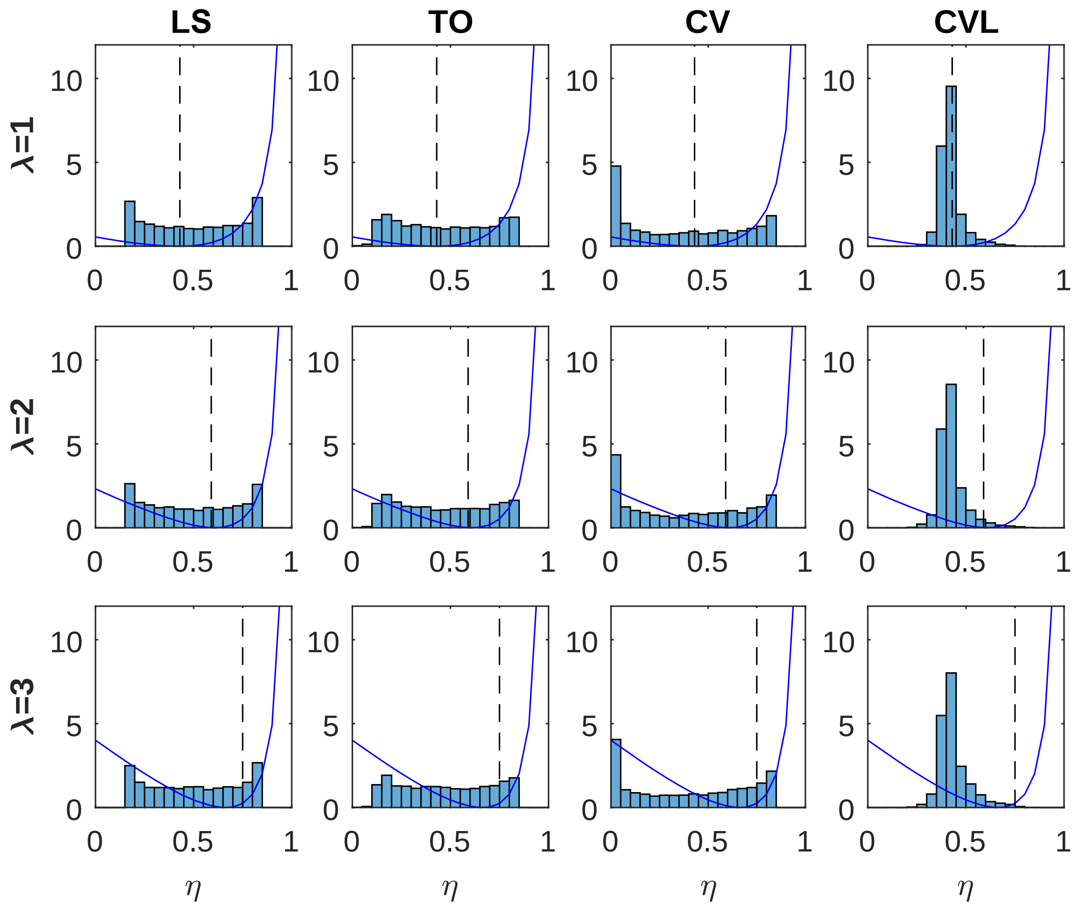
Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from Equation (2.6), and the vertical dashed lines mark the risk-minimizing fixed η.

Figure 3: Simulated Local Asymptotic Distributions of Estimates of η in Rolling Window Model M2



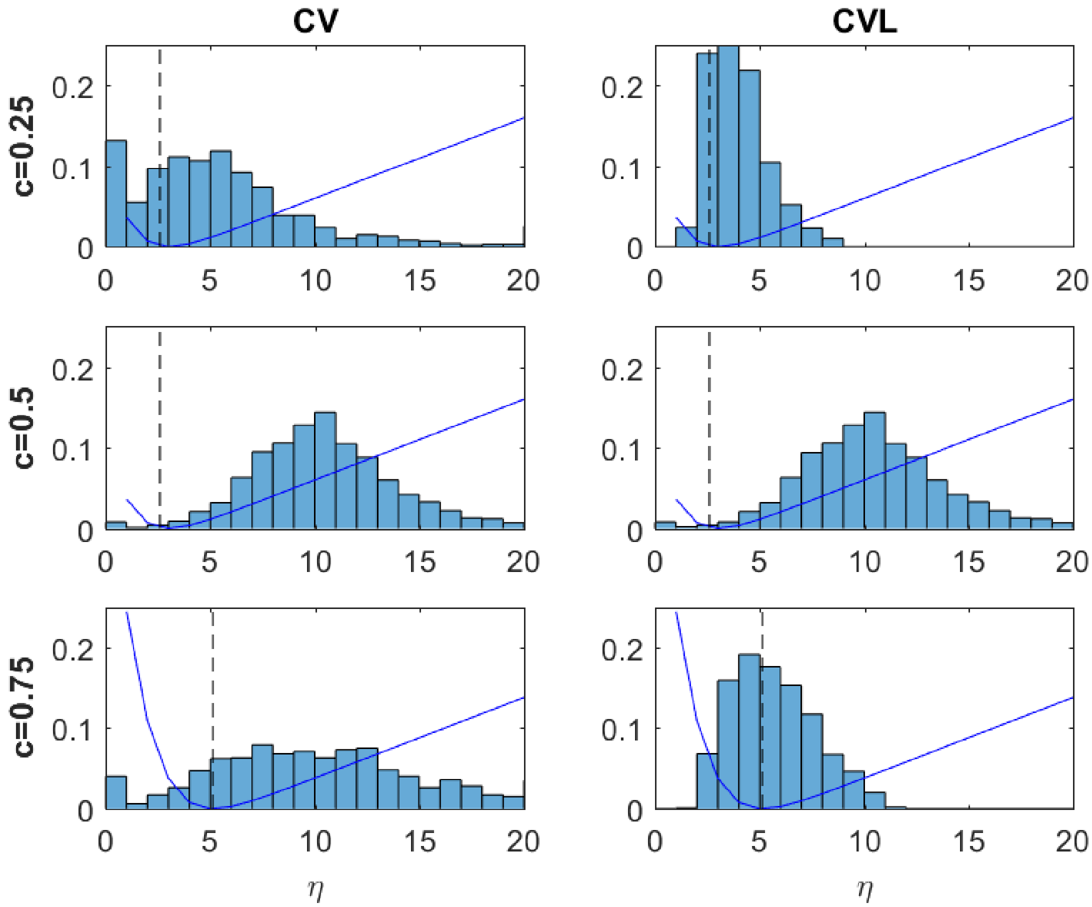
Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from Equation (2.7), and the vertical dashed lines mark the risk-minimizing fixed η.

Figure 4: Simulated Local Asymptotic Distributions of Estimates of η in Rolling Window Model M3



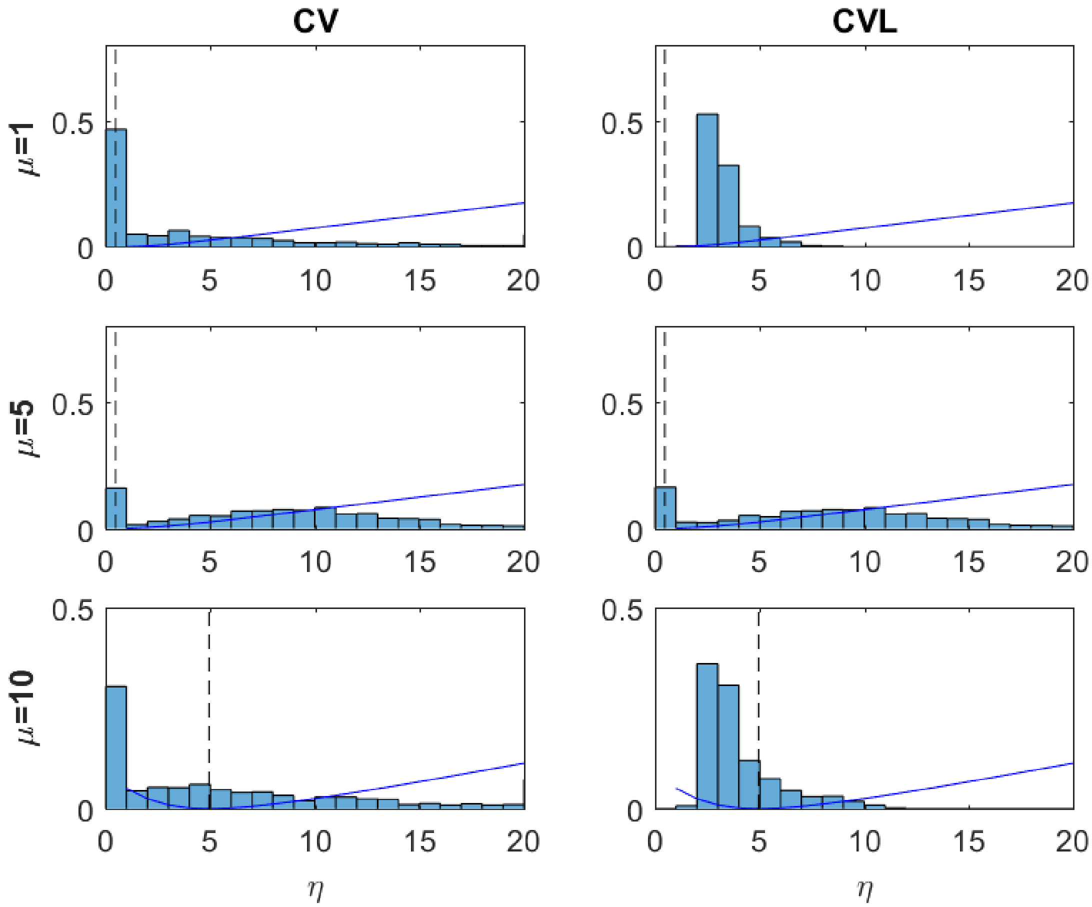
Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from substituting the Poisson breaks process into equation (2.5), and the vertical dashed lines mark the risk-minimizing fixed η.

Figure 5: Simulated Local Asymptotic Distributions of Estimates of η in Exponential Window with Model M1



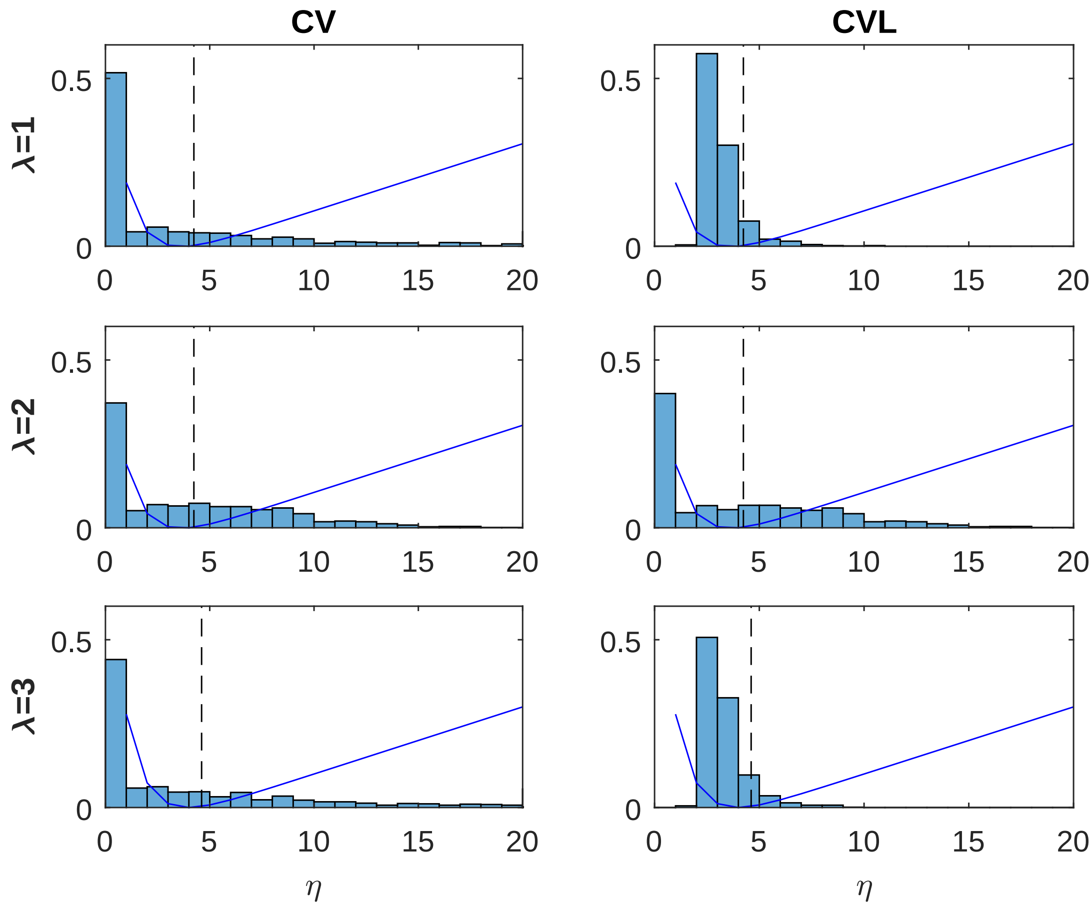
Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from Equation (2.9), and the vertical dashed lines mark the risk-minimizing fixed η.

Figure 6: Simulated Local Asymptotic Distributions of Estimates of η in Exponential Window with Model M2



Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from Equation (2.10), and the vertical dashed lines mark the risk-minimizing fixed η.

Figure 7: Simulated Local Asymptotic Distributions of Estimates of η in Exponential Window with Model M3



Notes: The figures show the simulated densities. The blue lines show the rescaled risk as a function of η, from substituting the Poisson breaks process into equation (2.8), and the vertical dashed lines mark the risk-minimizing fixed η.

Table 1: Local Asymptotic Risk

	Rolling Window				Exp Window	
Rolling Window	LS	TO	CV	CVL	CV	CVL
Model M1						
$c = 0.25$	1.9	1.8	5.8	1.9	7.3	2.4
$c = 0.5$	2.9	2.8	8.0	3.8	9.4	4.8
$c = 0.75$	6.8	6.5	10.3	16.9	11.6	11.4
Model M2						
$\mu = 1$	7.8	4.7	5.3	2.1	6.5	2.2
$\mu = 5$	9.4	7.7	8.5	6.3	9.7	5.5
$\mu = 10$	17.1	16.2	14.0	15.4	13.7	12.5
Model M3						
$\lambda = 1$	7.9	5.3	6.0	3.2	7.4	2.8
$\lambda = 2$	8.5	6.3	7.2	4.8	8.1	4.2
$\lambda = 5$	9.0	7.2	7.9	6.2	9.6	5.9

Notes: This table reports the average of Equation (3.1) averaged across all the draws.

Table 2: Simulated Finite Sample Loss: iid Errors

	Rolling Window								Exp Window			
DGP	LS 1	LS BIC	TO	CV All	CV Pre	CVL All	CVL Pre	IJR	ML	TO	CV	CVL
T=100												
1	5.35	1.42	3.47	4.37	3.40	1.74	1.27	3.10	2.15	2.12	6.84	7.71
2	1.50	2.13	1.54	5.00	3.09	1.74	3.08	3.17	5.88	5.88	7.80	7.74
3	2.23	2.75	2.41	6.54	5.25	3.70	9.80	4.24	6.16	6.16	9.94	7.83
4	4.67	7.09	5.42	7.56	7.26	16.7	22.7	9.06	8.24	8.28	12.2	8.79
5	5.39	1.84	3.59	4.52	3.56	1.97	1.55	3.23	2.44	2.41	7.10	7.91
6	7.26	7.43	6.53	6.78	6.34	6.14	6.92	6.00	6.20	6.23	9.88	9.08
7	15.4	15.6	15.3	11.8	13.8	15.5	19.7	14.3	12.0	12.3	14.4	12.2
8	5.87	3.74	4.48	5.29	4.40	3.39	3.31	4.12	3.80	3.79	7.88	8.41
9	6.42	5.56	5.42	6.05	5.31	4.86	5.16	5.18	5.10	5.11	8.63	8.59
10	7.14	6.84	6.30	6.84	6.26	6.19	6.89	6.18	6.07	6.08	9.78	9.21
T=200												
1	6.16	1.41	3.92	4.79	3.85	1.80	1.30	3.46	2.22	2.18	6.95	7.92
2	1.58	2.32	1.57	5.37	3.31	1.80	3.11	3.73	6.18	6.18	7.83	7.90
3	2.25	2.92	2.31	7.05	5.64	3.93	9.76	4.75	6.48	6.48	10.1	7.93
4	4.84	8.56	5.20	8.22	7.82	17.3	22.8	7.82	8.46	8.49	12.4	8.90
5	6.02	1.67	3.92	4.67	3.76	1.93	1.50	3.53	2.38	2.35	6.76	7.64
6	7.42	7.59	6.54	6.83	6.27	6.07	6.73	6.16	6.12	6.15	9.47	8.79
7	15.2	15.9	14.9	11.5	13.2	15.0	18.9	13.3	11.6	11.7	13.9	11.8
8	6.54	3.72	4.74	5.54	4.69	3.35	3.29	4.42	3.78	3.77	7.48	8.07
9	6.76	5.41	5.42	6.18	5.37	4.71	4.99	5.21	5.02	5.03	8.55	8.50
10	7.21	6.98	6.17	6.83	6.12	6.00	6.64	6.16	5.93	5.95	9.41	8.88

Notes: This table reports the simulated MSFE in Equation (4.1) for the 10 DGPs described in the text and with different ways of selecting the window start date. The methods are (i) Least squares with a single break, (ii) least squares estimation of the break dates following Bai and Perron (1998) in which the number of breaks is from 0 to 5, selected by BIC, (iii) the tradeoff method using least squares with a single break, (iv) cross-validation using the estimate $\hat{h}_3$, (v) cross-validation using the estimate $\hat{h}_4$, (vi) the Laplace cross-validation counterpart of $\hat{h}_3$, (vii) the Laplace cross-validation counterpart of $\hat{h}_4$, (viii) the Opt-R1 method of Inoue, Jin, and Rossi (2017), (ix) exponential weighting using the maximum likelihood estimate of model M2, (x) the tradeoff method for exponential weighting, (xi) exponential weighting with cross validation and (xii) exponential weighting with Laplace cross validation.

Table 3: Simulated Finite Sample Loss: AR(1) Errors

	Rolling Window									Exp Window		
DGP	LS 1	LS BIC	TO	CV All	CV Pre	CVL All	CVL Pre	IJR	ML	TO	CV	CVL
T=100												
1	4.65	4.56	4.35	4.37	3.36	2.64	1.85	2.40	15.17	15.17	8.13	6.96
2	1.39	4.54	1.39	4.73	2.72	2.00	2.76	2.59	15.55	15.55	8.17	6.48
3	2.00	4.58	2.00	5.46	4.20	3.61	6.51	4.08	15.51	15.51	8.58	6.32
4	4.01	3.81	4.02	5.79	5.51	7.17	9.30	8.69	15.52	15.52	8.86	6.69
5	4.68	4.60	4.40	4.46	3.50	2.79	2.05	2.55	15.35	15.35	8.30	7.11
6	6.70	6.39	6.60	6.23	5.90	5.69	5.96	5.29	15.69	15.69	9.12	7.82
7	14.90	11.49	14.87	10.69	13.15	12.19	15.46	12.97	16.71	16.71	11.09	10.05
8	5.20	5.27	4.98	5.00	4.19	3.72	3.25	3.48	15.89	15.89	8.61	7.35
9	5.84	5.84	5.69	5.68	5.02	4.78	4.66	4.56	15.39	15.39	8.73	7.50
10	6.53	6.45	6.41	6.26	5.78	5.60	5.80	5.42	16.09	16.09	9.14	7.82
T=200												
1	5.69	5.49	5.21	5.07	3.96	3.01	2.13	3.07	30.92	30.92	10.68	8.97
2	1.52	5.48	1.52	5.44	3.09	2.43	2.87	3.14	31.10	31.10	10.69	8.45
3	2.14	5.47	2.15	6.47	4.96	4.52	6.54	3.99	31.09	31.09	11.17	8.13
4	4.42	4.41	4.45	6.98	6.62	8.31	9.93	7.05	31.05	31.05	11.57	8.66
5	5.58	5.43	5.15	4.92	3.87	3.02	2.20	3.09	30.29	30.29	10.29	8.61
6	7.15	7.08	6.98	6.61	6.09	5.92	5.97	5.44	30.47	30.47	11.12	9.32
7	14.98	11.97	14.92	10.88	12.87	11.89	14.53	11.70	31.01	31.01	12.99	11.40
8	6.08	6.16	5.71	5.65	4.68	4.04	3.50	3.96	30.79	30.79	10.51	8.79
9	6.34	6.48	6.06	6.13	5.24	4.90	4.60	4.75	31.19	31.19	10.90	9.15
10	6.82	7.00	6.63	6.65	5.89	5.81	5.76	5.40	31.35	31.35	11.15	9.33

Notes: As for Table 2, except that the errors follow an AR(1) with coefficient 0.7.

Table 4: Inflation Root Mean Square Errors

	LS 1B	LS BIC	TO	CV All	CV Pre	CVL All	CVL Pre	IJR
Total PCE								
$h = 1$	1.882	1.767	1.834	1.808	1.787	1.784	1.777	1.824
$h = 2$	1.685	1.570	1.670	1.666	1.610	1.597	1.589	1.642
$h = 3$	1.679	1.490	1.650	1.590	1.586	1.475	1.489	1.513
$h = 4$	1.612	1.572	1.567	1.539	1.521	1.411	1.425	1.428
Core PCE								
$h = 1$	0.806	0.814	0.800	0.809	0.812	0.810	0.807	0.804
$h = 2$	0.706	0.728	0.710	0.718	0.705	0.714	0.707	0.706
$h = 3$	0.726	0.727	0.743	0.773	0.708	0.698	0.685	0.733
$h = 4$	0.770	0.878	0.769	0.843	0.736	0.712	0.732	0.763

Notes: This table reports the out-of-sample root mean square error of the inflation prediction regression in equation (5.1). There are two inflation measures: total PCE (real-time) and core PCE (revised) over the sample period 1959Q1-2018Q2, with out-of-sample forecasting starting after 80 quarters. Units are 100 times annualized log differences. We use rolling windows with 8 methods for selecting the window (i) LS(1B), least squares with a single break, (ii) LS(BIC), least squares estimation of the break dates following Bai and Perron (1998) in which the number of breaks is from 0 to 5 selected by BIC, (iii) TO, the tradeoff method using least squares with a single break, (iv) CV (All) cross-validation using the estimate $\hat{h}_3$, (v) CV (Pre) cross-validation using the estimate $\hat{h}_4$, (vi) CVL (All) and CVL (Pre) the Laplace cross-validation counterparts of these, and (vii) IJR the Opt-R1 method of Inoue, Jin, and Rossi (2017).

Table 5: Root Mean Square Error of AR(1) Forecasts with Alternative Windows

	LS-1	LS-BIC	TO	CV All	CV Pre	CVL All	CVL	Pre IJR
Average RRMSE	1.0368	1.0003	1.026	1.0169	1.0085	0.9951	0.9989	1.0032
25th Pctile RRMSE	1.0065	1.000	1.0041	1.0060	1.0019	0.9919	0.9964	0.9954
50th Pctile RRMSE	1.0322	1.000	1.0189	1.0209	1.0153	1.0036	1.0030	1.0107
75th Pctile RRMSE	1.0569	1.000	1.0415	1.0406	1.0261	1.0088	1.0070	1.0198
P(Min RRMSE)	0.0286	0.4762	0.019	0.0667	0.0476	0.1333	0.1238	0.1048
P(RRMSE=1)	0	0.681	0	0	0	0	0	0

Notes: This table considers out-of-sample root mean square error of prediction of each of the 210 series in the quarterly dataset of McCracken and Ng (2016) that are available from 1959Q1-2017Q4 using an AR(1). All of the window selection methods considered in Table 5 are used. Root mean square prediction errors are computed relative to the benchmark of an AR(1) estimated on the whole sample. The table reports the average and quartiles of the relative root mean square error (RRMSE) for each method across the 210 series. It also reports the proportion of the series that are best predicted by each method. The last line of the table gives the proportion of the series for which the method has RRMSE equal to one, which occurs when the method selects the full window.

References

BAI, J., AND P. PERRON (1998): “Estimating and testing linear models with multiple structural changes,” *Econometrica*, 66, 47–78.

CHERNOZHUKOV, V., AND H. HONG (2003): “An MCMC approach to classical estimation,” *Journal of Econometrics*, pp. 293–346.

CHRISTENSEN, T., X. CHEN, AND E. TAMER (2017): “Monte Carlo Confidence Sets for Identified Sets,” Working Paper.

DAVIS, R. A., AND W. T. DUNSMUIR (1996): “Maximum Likelihood Estimation for MA(1) Processes with a Root on or near the Unit Circle,” *Econometric Theory*, 12, 1–29.

ELLIOTT, G., AND U. K. MÜLLER (2007): “Confidence Sets for the Date of a Single Break in Linear Time Series Regressions,” *Journal of Econometrics*, 141, 1196–1218.

FAUST, J., AND J. H. WRIGHT (2013): “Forecasting inflation,” in *Handbook of Economic Forecasting, Volume 2*, ed. by G. Elliott, and A. Timmermann. Elsevier.

GEWEKE, J., AND G. AMISANO (2011): “Optimal Prediction Pools,” *Journal of Econometrics*, 164, 130–141.

GIRATIS, L., G. KAPETANIOS, AND S. PRICE (2013): “Adaptive forecasting in the presence of recent and ongoing structural change,” *Journal of Econometrics*, 177, 153–170.

HOLT, C. (1957): “Forecasting trends and seasonals by exponential weighted averages,” ONR Memorandum 52/1957. Carnegie Mellon University.

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.

INOUE, A., AND M. SHINTANI (2018): “Quasi-Bayesian Model Selection,” Working Paper.

KOOP, G., AND S. M. POTTER (2007): “Estimation and Forecasting in Models with Multiple Breaks,” *Review of Economic Studies*, 74, 763–789.

LEI, J. (2017): “Cross-Validation with Confidence,” Working Paper.

MCCRACKEN, M. W., AND S. NG (2016): “FRED-MD: A Monthly Database for Macroeconomic Research,” *Journal of Business & Economic Statistics*, 34, 574–589.

MÜLLER, U. K., AND Y. WANG (2017): “Nearly Weighted Risk Minimal Unbiased Estimation,” Working Paper, Princeton University.

NYBLOM, J. (1989): “Testing for the Constancy of Parameters over Time,” *Journal of the American Statistical Association*, 84, 223–230.

PESARAN, M. H., A. PICK, AND M. PRANOVICH (2013): “Optimal forecasts in the presence of structural breaks,” *Journal of Econometrics*, 177, 134–152.

PESARAN, M. H., AND A. TIMMERMANN (2007): "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137, 134–161.

PLOBERGER, W., AND W. KRÄMER (1990): "The Local Power of the Cusum and Cusum of Squares Tests," *Econometric Theory*, 6, 335–347.

SHEPHARD, N. (1993): "Distribution of the ML estimator of a MA(1) and a local level model," *Econometric Theory*, 9, 377–401.

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

——— (2009): "Phillips Curve Inflation Forecasts," in *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, ed. by J. Fuhrer, Y. K. Kodrzycki, J. S. Little, and G. P. Olivei. MIT Press.

——— (2010): "Modeling Inflation After the Crisis," Kansas City Fed symposium, Jackson Hole.