Rolling Window Selection for Out-of-Sample Forecasting with Time-Varying Parameters

Atsushi Inoue* Lu Jin† Barbara Rossi‡§

Vanderbilt University StataCorp ICREA-Universitat Pompeu Fabra Barcelona GSE CREI and CEPR

December 26, 2015

Abstract: While forecasting is a common practice in academia, government and business alike, practitioners are often left wondering how to choose the sample for estimating forecasting models. When we forecast inflation in 2018, for example, should we use the last 30 years of data or the last 10 years of data? There is strong evidence of structural changes in economic time series, and the forecasting performance is often quite sensitive to the choice of such window size. In this paper, we develop a novel method for selecting the estimation window size for forecasting. Specifically, we propose to choose the optimal window size that minimizes the forecaster's quadratic loss function, and we prove the asymptotic validity of our approach. Our Monte Carlo experiments show that our method performs quite well under various types of structural changes. When applied to forecasting US real output growth and inflation, the proposed method tends to improve upon conventional method, especially for output growth.

Keywords: Macroeconomic forecasting; parameter instability; nonparametric estimation; bandwidth selection.

*Department of Economics, Vanderbilt University, Nashville, TN 37235-1819. Email: atsushi.inoue@vanderbilt.edu.

†StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845-4512. Email: jl817cn@gmail.com.

‡ICREA-Universitat Pompeu Fabra, Barcelona GSE and CREI, Carrer Ramon Trias Fargas 25–27, 08005 Barcelona, Spain. Email: barbara.rossi@upf.es.

§Atsushi Inoue and Barbara Rossi gratefully acknowledge NSF support via grants ♯1022125 and ♯1022159, respectively.

1 Introduction

Parameter instability is widely recognized as a crucial issue in forecasting (Stock and Watson, 1996; Rossi, 2013; Giacomini and Rossi, 2009; Paye and Timmermann, 2006; Koop and Potter, 2004; Goyal and Welch, 2003; Clements and Hendry, 1998). The empirical evidence of parameter instability is widespread in financial forecasting (Goyal and Welch, 2003), exchange rate prediction (Schinasi and Swamy, 1989, and Wolff, 1987), and macroeconomic forecasting (Stock and Watson, 1996, 2003, 2007), to name a few. To handle such instability, it is quite common to use only the most recent observations to estimate parameters of forecasting models rather than all available observations (the so-called "rolling estimation" method). Examples of rolling estimation include: in finance, Goyal and Welch (2003) to evaluate the power of dividend ratios in predicting stock market returns and the equity premium; in macroeconomics, Swanson (1998) to investigate the extent to which fluctuations in the money stock predict fluctuations in real income; in exchange rate forecasting, Molodtsova and Papell (2009) to investigate the predictability of models that incorporate Taylor rule fundamentals for exchange rate.

In rolling out-of-sample forecasting, one produces a sequence of pseudo out-of-sample forecasts using a fixed number of the most recent data at each point of time. One practical issue with rolling out-of-sample forecasting is how many recent observations should be used in the estimation. The number of the recent observations used in estimation is referred to as the window size. Conventionally, the window size is arbitrarily determined by forecasters or based on past experience. For instance, Molodtsova and Papell (2009) use a 10-year window of monthly data to predict exchange rates; and Stock and Watson (2007) forecast inflation with a 10-year window of quarterly observations. However, we often find that the forecasting performance of the rolling scheme is sensitive to the choice of the window size (see Inoue and Rossi, 2012).

While the problem of selecting the estimation window size is similar to the problem of bandwidth selection in nonparametric estimation, methods to select the window size in rolling out-of-sample forecasting have received little attention. Among recent papers focusing on how to determine the optimal window size: Pesaran and Timmermann (2007) propose five methods to select the window size when the forecasting model is subject to one or multiple discrete breaks; Pesaran, Pick and Pranovich (2013) derive optimal weights under continuous and discrete breaks; and Giraitis, Kapetanios and Price (2013) develop a cross-validation-based method to select a tuning parameter to downweight older data in the presence of structural change.

In this paper, we develop a new approach for selecting the size of the rolling estimation window in forecasting models with potential breaks. More specifically, parameters are specified as smooth functions of time and the functional forms are unknown. This setting, in which structural changes may occur in every point in time and are small, is consistent with empirical findings of small instability in some forecasting areas, such as forecasting inflation (Stock and Watson, 1999). This setup is also adopted in the nonparametric literature, for example, Robinson (1989), Cai (2007) and Chen and Hong (2012).

Our approach has three advantages over existing methods. First, the error term and regressors can be weakly dependent, and the regressors can include both exogenous and lagged dependent variables, while existing methods rely on more stringent assumptions. The five window selection methods developed in Pesaran and Timmermann (2007) require serially uncorrelated errors and strictly exogenous regressors. Pesaran, Pick and Pranovich's (2013) approach needs independent errors and exogenous regressors. Giraitis, Kapetanios and Price (2013) focus on models without regressors. Thus, our approach can be used for a wider range of forecasting models than existing methods. Second, our approach allows multiple-step-ahead forecasting, while existing methods only consider one-step-ahead. Third, our procedure for selecting the optimal window size is feasible in practice and we also prove its asymptotic validity. Pesaran and Timmermann's (2007) methods are designed for discrete breaks, rather than the continuous breaks assumed in the present paper.

Our new approach proposes to choose the optimal window size that minimizes the conditional mean square forecast error (MSFE), which is commonly used as the forecasters' loss function at the end of the sample. Since the conditional MSFE is infeasible, we construct an approximate conditional MSFE by replacing the unknown parameters in the conditional MSFE with estimates from local linear regressions, and then choose the window size that minimizes this approximate conditional MSFE. We show that choosing the optimal window size based on our approximate criterion is asymptotically equivalent to choosing the window size based on the infeasible one. Choosing the window size for the conditional MSFE as opposed to the integrated MSFE and establishing their asymptotic justification under the aforementioned general framework are our new contributions to the literature. Our Monte Carlo simulations suggest that using the window size selected by our procedure can improve the forecasting performance vis-à-vis an ad-hoc choice of the window size.

Moreover, we empirically assess the practical value of our procedure in forecasting real output growth and inflation. As shown in Stock and Watson (2003, 2007), the predictive ability of standard forecasting models suffers from instability; that is, finding a predictor useful in one period does not guarantee that it will predict well in later periods. In our empirical analysis, we examine whether we can improve forecasts by using our proposed window selection procedure. Our results suggest that asset prices, unemployment and monetary measures have useful predictive content for forecasting output growth at short horizons. When forecasting inflation, measures of unemployment are useful at long horizons, confirming the usefulness of the unemployment-based Phillips curve for inflation forecasting in the presence of parameter instability. In general, the forecast improvements generated by the optimal window size are more substantial when forecasting output growth than inflation, since, as we show, parameters are more likely to vary in the former than in the latter.

When the optimal window sizes are used, the number of building permits has useful predictive content for long-term output growth forecasts, and measures of unemployment are useful for inflation forecasts. One possible economic interpretation is that building constructions typically take a long time to complete, so investment in the construction sector has a long-term effect on output growth. The unemployment-based Phillips curve is useful in predicting inflation, possibly because the non-accelerating inflation rate of unemployment (NAIRU) is unstable and the optimal window size captures time variation.

The rest of the paper is organized as follows. Section 2 presents a model, motivates our problem and describes our proposed window selection procedure. Section 3 provides theoretical justifications for our window selection procedure. Section 4 reports Monte Carlo simulation results. Section 5 applies our procedure to forecasting output growth and inflation in the United States, and Section 6 concludes. The appendix provides proofs of the theorems.

2 Motivation and Setup

Assume the data generating process (DGP) is:

$$y_{t+h} = \beta'_{h,t} x_t + u_{t+h}, \;\; t = 1, 2, ..., T, \tag{1}$$

where $x_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$ is a $p \times 1$ vector of stochastic regressors, $\beta_{h,t} = (\beta_{h,t,1}, \beta_{h,t,2}, ..., \beta_{h,t,p})'$ is a $p \times 1$ vector of time-varying parameters; u_{t+h} is an unobservable disturbance; h denotes the forecast horizon, where $1 \leq h < \infty$ and $h \in \mathbb{Z}^+$; and T denotes the full sample size. The regressor vector x_t may include exogenous explanatory variables and lagged values of the dependent variable. Our interest is to predict y_{T+h} using information available at time T.

As in Robinson (1989) and Cai (2007), the time variation in the parameters is represented by a smooth function of the current period t. For each i, $1 \leq i \leq p$, $\beta_{h,t,i}$ is defined as $\beta_{h,t,i} = \beta_i\left(\frac{t}{T}\right)$, where the parametric form of $\beta_i(\frac{t}{T})$ is unknown and its dependence on forecast horizon h is omitted for notational simplicity. Thus equation (1) can be rewritten as:

$$y_{t+h} = \beta\left(\frac{t}{T}\right)' x_t + u_{t+h}, \tag{2}$$

where $\beta(\frac{t}{T}) = (\beta_1(\frac{t}{T}),\, \beta_2(\frac{t}{T}), ..., \beta_p(\frac{t}{T}))'$ is a vector of unknown smooth functions of time t. This framework avoids parametric restrictions on $\beta_{t,i}(\cdot)$. Note that $\beta_{t,i}(\cdot)$ is defined on an equally spaced grid over $(0, 1)$, which becomes finer as $T \to \infty$. According to Robinson (1989), this requirement is important for deriving consistent nonparametric estimates, since the amount of local information on which an estimator depends increases suitably as sample size T increases. Although the functional form of $\beta(\cdot)$ is unspecified, we require it is smooth enough.

The rolling OLS estimator is commonly used in forecasting because parameters are often found to be time-varying. While rolling OLS estimators may look like a parametric estimator, it is a local constant estimator and thus is a nonparametric estimator of $\beta_h(\cdot)$ in equation (2), where the estimation window size plays the role of the bandwidth. [1]

We focus on how to determine the size of the estimation window for forecasting in the framework

[1] The rolling window estimator is a local constant estimator with the truncated kernel that assigns 0-1 to the observations. While such weights may not be optimal, we focus on the rolling window estimator because it is widely used in practice. We refer to Pesaran, Pick and Pranovich (2013) for the analysis of optimal weights.

described above. Our new approach chooses the optimal window size that minimizes the conditional MSFE. The conditional MSFE is a commonly used measure of forecast accuracy. Both rolling windows and MSFE are used in Bacchetta, van Wincoop and Beutler (2010), Carriero, Kapetanios and Marcellino (2009), Chen, Rogoff and Rossi (2010), Cheung, Chin and Pascual (2005), Della Corte, Sarno and Sestieri (2012), Faust, Rogers and Wright (2003), Meese and Rogoff (1983a,b), Molodtsova and Papell (2009, 2012), Pesaran and Timmermann (2007), and Welch and Goyal (2007), to name a few. It should be noted that the conditional MSFE is not the only loss function that yields the conditional mean as the optimal forecasts (Patton, 2015).

The population MSFE at the end of the sample is defined by

$$E_T[(y_{T+h} - \beta_h(1)'x_T)^2], \tag{3}$$

where $E_T(\cdot)$ is the conditional expectation operator based on the information set at time T and $\beta_h(1) = \beta_h(T/T)$ is the parameter value at time T Because $\beta_h(\cdot)$ is unknown, we replace (3) by

$$E_T[(y_{T+h} - \widehat{\beta}_R(1)'x_T)^2], \tag{4}$$

where $\widehat{\beta}_R(1)$ is the rolling OLS estimate of $\beta(1)$ based on the last R observations in the sample. We choose the window size R to minimize (4). Since

$$E_T[(y_{T+h} - \widehat{\beta}_R(1)'x_T)^2] \;=\; \sigma_h^2 + (\widehat{\beta}_R - \beta(1))'x_T x_T'(\widehat{\beta}_R - \beta(1)), \tag{5}$$

where σ_h^2 is the variance of u_{t+h}, minimizing (4) is equivalent to minimizing

$$(\widehat{\beta}_R - \beta(1))'x_T x_T'(\widehat{\beta}_R - \beta(1)). \tag{6}$$

However, (6) is not feasible because it depends on the unknown parameter value $\beta_h(1)$. Replacing it with a local linear estimate with initial window size R_0, $\tilde{\beta}(1)$, yields the following feasible criterion:

$$(\widehat{\beta}_R - \tilde{\beta}(1))'x_T x_T'(\widehat{\beta}_R - \tilde{\beta}(1)). \tag{7}$$

Our proposal is to minimize (7) with respect to R to achieve (3).

Specifically, we implement our proposed window selection method in the Monte Carlo simulation and empirical application in this paper as follows:
Step 1: Test whether the parameters are constant using Bai and Perron's (1998, Section 4.1) test.
Step 2: If we fail to reject the null hypothesis of constant parameters in Step 1, we set the optimal window R to the full sample size. Otherwise, we set initial window size R_0 to the one chosen by Pesaran and Timmermann's (2007) cross validation method with unknown break dates. We then select the window R from a set $(\underline{R}, \bar{R})$ to minimize the approximate conditional MSFE, $(\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1))$.

3 Theory

3.1 Assumptions

First we define the notation. For a $p \times 1$ random vector $X \equiv (X_1, \ldots, X_p)'$, $\|X\|_r$ denotes the L_r-norm of X, i.e. $\|X\|_r = (\sum_{i=1}^{p} E(|X_i|^r))^{1/r}$. For a $k \times 1$ real vector $x \equiv (x_1, \ldots, x_k)'$, $\|x\|$ denotes the Euclidean norm of vector x, i.e. $\|x\|^2 = \sum_{i=1}^{k} x_i^2$. For any $m \times n$ matrix $A \equiv (a_1, a_2, ..., a_n)$, where a_j is the $j-$th column of matrix A, and $a_j = (a_{1j}, a_{2j}, ..., a_{mj})'$ for $j = 1, 2, \ldots, n$, $vec(A)$ is an $mn \times 1$ vector, i.e. $vec(A) \equiv (a_1', a_2', ..., a_n')'$. From this point on we write $\beta_h(\cdot)$ as $\beta(\cdot)$ to simplify the notation.

The assumptions imposed on the data generating process are as follows:

Assumption 1 *$\{u_{t+h}\}_{t=1}^{T}$ is a sequence such that $E(u_{t+h}|\Omega_t) = 0$, $\sigma_t^2 = Var(u_{t+h}|\Omega_t)$ is well-defined a.s., where $\Omega_t = \sigma(x_t', x_{t-1}', \ldots, y_t, y_{t-1}, \ldots)$ is the information observed at time t, and all eigenvalues of $E(u_{t+h}^2 x_t x_t')$ are finite and bounded away from zero uniformly in t.*

Assumption 1 imposes that the forecast error is a martingale difference sequence when $h = 1$. However, Assumption 1 rules out unit roots.

Assumption 2 *Let $\{Z_t\} \equiv \{(u_t, x_{t-h}')'\}$, $t = h+1, \ldots, T+h$. For $r > 2$, the sequence $\{Z_t\}$ is (i) $L_{4r/(r-1)}$-NED of size -2, with positive constants $d_t = O(\|Z_t - EZ_t\|_{4r/(r-1)})$, on a sequence $\{V_t\}_{-\infty}^{\infty}$, where $\{V_t\}$ is α-mixing of size $-2r/(r-2)$; (ii) $\|vec(Z_t Z_t')\|_r \leq M$, for some constant M, $0 < M < \infty$, uniformly in t.*

While it is common to assume that data are α-mixing (Cai, 2007; Clark and McCracken, 2001; and West, 1996), Assumption 2 allows the data to be near-epoch dependent (NED). The NED assumption is more general than the α-mixing assumption, allows for heterogeneity over time which is necessary for our time-varying parameter framework, and overcomes several undesirable features of the α-mixing assumption (Lu and Linton, 2007). Since we allow for parameter instability in eq. (1), the mean of the dependent variable varies over time. Therefore we do not impose stationary conditions on Z_t, that may include lagged dependent variables; this differentiates our approach from Cai (2007), who assumes Z_t to be strictly stationary.

Assumption 2(ii) requires $vec(Z_t Z_t')$ to be L_r-bounded uniformly in t, so is its subcomponent $vec(x_t x_t')$. As $r > 2$, this assumption also ensures the existence of the fourth and second moments of Z_t uniformly in t.

Assumption 3 *All eigenvalues of $E(x_t x_t')$ are bounded away from zero uniformly in t, $1 \leq t \leq T$, $T \geq 1$.*

Assumption 3 requires the matrix $E(x_t x_t')$ be positive definite and non-singular uniformly in t, which is necessary for nonparametric estimation of $\beta(t/T)$.

Assumption 4 *(i) $\beta(\cdot)$ is twice continuously differentiable over the real line $\mathbb{R}$; (ii) $\|\beta^{(i)}(\frac{t}{T})\|$ is bounded uniformly in t, for $i = 1, 2$, where $\beta^{(i)}(\cdot)$ denotes the $i-th$ derivative of $\beta(\cdot)$.*

Assumption 4 imposes a smoothness condition on $\beta(\cdot)$. This condition is necessary because $\beta_j^{(1)}(\cdot)$ and $\beta_j^{(2)}(\cdot)$ appear in the Taylor expansion of $\beta_j(\cdot)$ and the bias of the rolling OLS estimate of $\beta_j(\cdot)$. Assumption 4(ii) is used to derive the rate of the optimal window size. It is important to note that we do not specify the parametric form of $\beta(\cdot)$.

Assumption 5 *$R_0, R \to \infty$, $R/R_0 = o(1)$ and $R_0^2 R/T = o(1)$ as $T \to \infty$.*

Here R denotes the number of the most recent observations used to predict y_{T+h}, R_0 is the number of the most recent observations used to construct local linear estimates, and T is the full sample size. Assumption 5 requires that the window sizes R and R_0 go to infinity as the sample size T goes to infinity, but the divergent rates of R and R_0 are slower than T.

Assumption 6 *R belongs to a set $\Theta_R \subseteq \mathbb{Z}^+$ and $\Theta_R \subset [\underline{R}, \bar{R}]$, where $\underline{R}$ and $\bar{R}$ satisfy the assumption for R in Assumption 5. Also the cardinality of Θ_R, denoted by $\#\Theta_R$, satisfies $\#\Theta_R = \underline{R}^\rho$, for some ρ, $0 < \rho < 1$.*

Assumption 6 implies that the number of elements in Θ_R grows at the rate of T^τ for some $0 < \tau < 1$. Thus the cardinality of the set Θ_R is cT^τ, for some $c > 0$. This assumption is useful to derive results uniform in R, as in Marron (1985), Marron and Härdle (1986) and Härdle and Marron (1985).

3.2 Infeasible Conditional MSFE

We choose the most recent R observations to estimate the forecasting model, then use the estimated coefficients to produce the forecast. The h-step ahead forecast $\hat{y}_{T+h} \equiv \hat{\beta}_R(1)' x_T$ is based on OLS estimation using the most recent R observations:[2]

$$\hat{\beta}_R(1) \equiv \hat{\beta}_R\left(\frac{T}{T}\right) = \left(\sum_{t=T-R+1}^{T-h} x_t x_t'\right)^{-1} \left(\sum_{t=T-R+1}^{T-h} x_t y_{t+h}\right). \tag{8}$$

The accuracy of the forecast $\hat{y}_{T+h}$ depends on the choice of R. Including too distant information reduces the forecast variance but increase its bias; on the other hand, if R is too small, the forecast variance increases although the bias decreases. So the optimal estimation window resolves the trade-off between forecast variance and bias.

The optimal window size minimizes the conditional MSFE, $E((y_{T+h} - \hat{y}_{T+h})^2 | \Omega_T)$. Expanding the conditional MSFE gives

$$\begin{aligned} E\left((y_{T+h} - \hat{y}_{T+h})^2 \middle| \Omega_T\right) &= E\left((\beta(1)' x_T + u_{T+h} - \hat{\beta}_R(1)' x_T)^2 \middle| \Omega_T\right) \\ =& E\left(u_{T+h}^2 | \Omega_T\right) - 2E\left((\hat{\beta}_R(1) - \beta(1))' x_T u_{T+h} \middle| \Omega_T\right) \\ &+ E\left((\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1)) \middle| \Omega_T\right) \end{aligned} \tag{9}$$

Because $\hat{\beta}_R(1)$ and x_T are deterministic given the information set Ω_T, we have

$$E\left((\hat{\beta}_R(1) - \beta(1))' x_T u_{T+h} \middle| \Omega_T\right) = ((\hat{\beta}_R(1) - \beta(1))' x_T) E\left(u_{T+h} \middle| \Omega_T\right) = 0,$$

[2] The subscript R means that the estimate is computed using the most recent R data.

where the last inequality is also implied by Assumption 1. Thus the second term is zero. Since the first two terms in (9) are independent of R, minimizing the conditional MSFE with respect to R is equivalent to minimizing $(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$ with respect to R.

We derive the rate of the optimal window size in the following theorem:

Theorem 1 *In addition to Assumption 1–6, assume that $\beta'(\cdot)$ is bounded away from zero uniformly. Then the optimal window size R that minimizes $(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$ is of order $T^{2/3}$ in probability.*

The proof of this theorem is in the appendix. Theorem 1 shows that, in the presence of smoothly time-varying parameters, the optimal window, which minimizes the conditional MSFE, should equal $cT^{2/3}$ with probability going to unity, for some constant c, $0 < c < \infty$. However, this rate is not useful in practice, because the constant c is still unknown to practitioners. If one attempts to search all the possible values of R to minimize $(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$, he/she will soon find it is still infeasible, because the true parameter $\beta(1)$ is unknown.

3.3 Approximate MSFE using Local Linear Regressions

In this section, we replace the unknown $\beta(1)$ in the infeasible criterion $(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$ by a local linear regression estimate. The local linear regression method is considered to be a superior method in theory and applications among non-parametric regressions, see Fan and Gijbels (1996) and Cai (2007). One desirable feature of the local linear regression estimator is that it has the same asymptotic behavior at the interior points and the boundaries, whereas the Nadaraya-Watson estimator regression has a larger bias at the boundaries. Also the bias of the Nadaraya-Watson estimator at the boundaries is larger than the bias of the local linear regression at the boundaries. Here $\hat{\beta}_R(1)$ is actually a special Nadaraya-Watson estimate, which uses the uniform kernel and is evaluated at the end of the sample. Using the fact that the bias of a local linear estimate is smaller than the Nadaraya-Watson estimate at the end of the sample, the error introduced by the approximation of local linear estimates is negligible. The local linear regression proceeds as follows.

Provided that the parameter function $\beta(\cdot)$ is twice continuously differentiable over the real line in Assumption 4, for any $t = 1, \dots, T$, we can approximate $\beta(\frac{t}{T})$ by:

$$\beta\left(\frac{t}{T}\right) = \beta(1) + \beta^{(1)}(1)\left(\frac{t-T}{T}\right) + \frac{\beta^{(2)}(c)}{2!}\left(\frac{t-T}{T}\right)^2, \tag{10}$$

where $c = \lambda\frac{t}{T} + (1-\lambda)\frac{T}{T}$, for $\lambda \in (0,1)$. $\beta^{(i)}(\cdot)$ denotes the ith derivative of $\beta(\cdot)$. By substituting eq. (10) into eq. (2), we obtain

$$\begin{aligned} y_{t+h} =& \beta(1)' x_t + \beta^{(1)}(1)' x_t \left(\frac{t-T}{T}\right) + \frac{\beta^{(2)}(c)'}{2} x_t \left(\frac{t-T}{T}\right)^2 + u_{t+h} \\ =& \beta(1)' x_t + \beta^{(1)}(1)' x_t \left(\frac{t-T}{T}\right) + \epsilon_{t+h}, \end{aligned} \tag{11}$$

where ϵ_{t+h} is a composite error term of u_{t+h} and the second order term in eq. (10).

Let $\tilde{\beta}(1)$ and $\tilde{\beta}^{(1)}(1)$ be the estimates for $\beta(1)$ and $\beta^{(1)}(1)$ in eq. (11); then the OLS estimator is given by

$$\begin{bmatrix} \tilde{\beta}(1) \\ \tilde{\beta}^{(1)}(1) \end{bmatrix} = \begin{bmatrix} \sum x_t x_t' & \sum x_t x_t' \left(\frac{t-T}{T}\right) \\ \sum x_t x_t' \left(\frac{t-T}{T}\right) & \sum x_t x_t' \left(\frac{t-T}{T}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t y_{t+h} \\ \sum x_t y_{t+h} \left(\frac{t-T}{T}\right) \end{bmatrix} \tag{12}$$

where the summation $\sum$ represents $\sum_{t=T-R_0+1}^{T-h}$. $\tilde{\beta}(1)$ and $\tilde{\beta}^{(1)}(1)$ are estimated using the most recent R_0 data, where $R_0 = 2p, \dots T$, is a given pilot window size for the local linear regression.

Next, replacing the unknown parameter $\beta(1)$ with the local linear estimate $\tilde{\beta}(1)$ leads to a feasible window selection criterion: the optimal window size $\hat{R}$ satisfies

$$\hat{R} = \arg\min_{R \in \Theta_R} (\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1)) \tag{13}$$

where $\tilde{\beta}(1)$ is computed using R_0 observations and the estimate $\hat{\beta}_{\hat{R}}(1)$ means that it is estimated using $\hat{R}$ observations. Here R_0 is treated as a given value. Theorem 2 shows that this approximate MSFE rule is asymptotically optimal relative to the infeasible conditional MSFE. In other words, the error introduced by replacing $\beta(1)$ with $\tilde{\beta}(1)$ is asymptotically negligible.

Theorem 2 *Under Assumptions 1–6, choosing R to minimize $(\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1))$ is asymptotically optimal in the sense that*

$$\frac{(\hat{\beta}_{\hat{R}}(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_{\hat{R}}(1) - \tilde{\beta}(1))}{\inf_{R \in \Theta_R} (\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))} \xrightarrow{p} 1, \tag{14}$$

where $\hat{R} = \arg\min_{R \in \Theta_R} (\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1))$ *and* $\tilde{\beta}(1)$ *is the estimates from the local linear regression in eq.* (12) *using the* R_0 *most recent observations.*

Theorem 2 provides a formal justification for using the approximate rule in eq. (13) as a proxy for the infeasible MSFE. The asymptotic optimality suggests that $\hat{R}$ chosen from the approximate rule can yield the same forecasts as the true optimal window size chosen by the infeasible MSFE with probability approaching one. The proof is shown in the appendix. The emphasis of the proof is to show that the left-hand side of eq. (14) converges to one in probability uniformly in $R \in \Theta_R$. Conditioning on R_0 with different orders of magnitude, the asymptotic optimality holds uniformly for R over its corresponding range. The growing cardinality of Θ_R given by Assumption 6 plays an important role in the proof. Similar techniques are used in Marron (1985), Marron and Härdle (1986) and Härdle and Marron (1985).

4 Monte Carlo Experiments

We now turn to a Monte Carlo analysis of the performance of the window selection procedure described above. The purpose of this section is to replicate existing methods for selecting estimation window size (Pesaran and Timmermann, 2007; Cai, 2007; Anatolyev and Kitov, 2007; Pesaran, Pick and Pranovich, 2013), and compare their performance with our window selection procedure based on the approximate MSFE.

4.1 DGPs

The DGPs are based on:

$$\begin{bmatrix} y_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} \beta_t \\ 0 \end{bmatrix} + \begin{bmatrix} a_t & b_t \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} + \begin{bmatrix} u_{y,t+1} \\ u_{x,t+1} \end{bmatrix}, \tag{15}$$

where the error terms satisfy

$$\begin{bmatrix} u_{y,t+1} \\ u_{x,t+1} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right) \tag{16}$$

Detailed setups for these two types of DGPs are listed in Tables 1. DGPs 1 to 7 are based on univariate version with $a_t = b_t = 0$ for all t, while DGPs 8 to 20 are based on the bivariate VAR(1) model used in Pesaran and Timmermann (2007).

Table (1) includes the following functional forms for the parameters: (1) constant parameters; (2) parameters with one-time break; (3) smooth time-varying parameters; and (4) Nyblom's (1989) random walk parameter model. First, constant parameters are used in DGPs 1 and 8. Second, DGPs from 2 to 4 and 9 to 16 consider parameters with a one-time break: the break date is $0.25T$ for DGPs 2, 9 and 13, $0.5T$ for DGPs 3, 10 and 14, $0.75T$ for DGP 4, 11 and 15, $0.95T$ for DGPs 12 and 16, respectively. Third, DGPs 5 and 6 and DGPs from 17 to 20 use smoothly time-varying parameters: in DGPs 5, 17 and 18, parameters are linear functions of t; in DGPs 6, 19 and 20, parameter functions are quadratic in t. Fourth, Nyblom's (1989) random walk parameter is used in DGPs 7, 21 and 22.

In DGPs 2, 3 and 4, the variance of the error term is chosen by equalizing the variance of $\{\beta_t\}$ and the variance of $\{u_{t+1}\}$ for each break date to ensure that the signal to noise ratio is the same across these DGPs. In DGP 7, the variance of the error term of the random coefficient β_t is controlled so that the variance of the random coefficients equals the variance of the error term of the model. In DGPs 7, 21 and 22, the variances of the error term in the random coefficients are assumed to be relatively small ($T/2$) so that the parameters change smoothly. The standard deviation of the error term in the process of the random coefficient a_t is set at a small value, $0.1/\sqrt{T}$. The purpose of this setting is to prevent a_t from exceeding unity, in which case the process y_t would explode and become unstable. Simulations with a_t greater than one are are discarded.

4.2 Window Selection Methods

Tables 2–5 report results using the following window selection methods: (i) the five methods used in Pesaran and Timmermann (2007, "PT" thereafter); (ii) the weighted least squares method of Anatolyev and Kitov (2007) labeled "WLS"; (iii) Cai's (2007) AIC bandwidth selection rule ("Cai1" and "Cai2"); (iv) Pesaran, Pick and Pranovich's (2013, equation 48 on page 144, labeled "PPP") robust optimal weighting method; (v) our proposed method based on the approximate MSFE; and (vi) our proposed method based on the infeasible MSFE.

More in detail, we include the following methods used in PT: (1) the post-break method (labeled

"Postbk" the tables); (2) cross validation ("CV"); (3) weighted average of forecasts ("WA"); (4) pooled forecast combination ("Pooled"); and (5) the trade-off method ("Troff"). These methods are designed to select the rolling window size when parameters are subject to discrete breaks. The results with estimated break dates are reported under the label "Estimated break date ($\hat{T}_1$)".[3] In addition to these two variants, cross validation, weighted average of forecasts and pooled forecast combination can be implemented without taking a break into account; the results without estimating a break date are reported under the label "Unknown break date".

Second, Cai's (2007) method is implemented for local constant regression estimators with the uniform kernel ("Cai1") and the Epanechnikov kernel ("Cai2") and for local linear regression estimators with the uniform kernel ("LL1") and the Epanechnikov kernel ("LL2"). The bandwidth, τ, is chosen to minimize

$$\text{AIC}(\tau) = \log(\hat{\sigma}^2) + 2(n_\tau + 1)/(n - n_\tau - 2), \tag{17}$$

where n is the sample size, $\hat{\sigma}^2 = (1/n) \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$ and n_τ is the trace of the matrix H_τ. Here H_τ satisfies $\hat{Y} = H_\tau Y$, where $Y = (y_1, \cdots, y_n)'$. The product of the sample size n and the bandwidth τ, $n\tau$, equals our window size R, and thus R is selected from the set, $\{0.1T, 0.125T, 0.150T, ..., 0.675T, 0.7T\}$, to minimize the AIC criterion (17). Note that Cai's bandwidth selection method is not designed to produce the best forecasts at time T because the AIC criterion in eq. (17) is based on the sum of the squared residuals from time 1 through time T.

Third, the window selection method developed in this paper is implemented as follows:

Step (1): Test whether the parameters are constant using Bai and Perron's (1998, Section 4.1) test. Critical values are set at the 5% significance level. The trimming range for the possible break dates is $[0.15T, 0.85T]$. We will perform robustness checks with respect to the significance level and trimming ranges.

Step (2): If we fail to reject the null hypothesis of constant parameters in Step (1), we set the optimal window R to the full sample size. Otherwise, we set R_0 to the one chosen by PT's CV with unknown break dates, $\underline{R} = \max(1.5T^{2/3}, 20)$, $\bar{R} = \min(4T^{2/3}, T - h)$ ("OptR1") $\bar{R} = \min(5T^{2/3}, T - h)$

[3] We have also implemented these methods imposing the true break date in DGPs 2–4 and 9–16. Because the results are qualitatively similar to those with estimated break dates and because, in practice, imposing the true break date is infeasible, we report only the results with estimated break dates to save space. The results with the true break dates imposed are available upon request from the authors.

("OptR2") and $\bar{R} = \min(6T^{2/3}, T-h)$ ("OptR3"). We then select the window R from $(\underline{R}, \bar{R})$ to minimize the approximate conditional MSFE $(\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1))$.

Fourth, the infeasible window selection criterion $(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$ is also considered (labeled "True"). Here R is chosen in the range $[0.1T, 0.9T]$. This infeasible version, which uses the true value of $\beta(1)$ instead of the estimated value $\tilde{\beta}(1)$, should always perform better than our approximate MSFE criterion.

4.3 Simulation Results

We evaluate the performance of the out-of-sample prediction of y_{T+h} over 5,000 Monte Carlo simulations for $T = 100, 200$ and $h = 1, 2$. Tables 2–5 report the ratios of the RMSFEs (square root MSFEs) produced by the optimal window size relative to the RMSFEs produced in the full sample:

$$\sqrt{\frac{\sum_{m=1}^{5{,}000} (y_{T+1}^{(m)} - \hat{y}_{T+1}^{(m)})^2}{\sum_{m=1}^{5{,}000} (y_{T+1}^{(m)} - \tilde{y}_{T+1}^{(m)})^2}}, \tag{18}$$

where $\hat{y}_{T+1}^{(m)}$ is the forecast computed using the optimal window size obtained using the window selection methods for the $m-$th replication and $\tilde{y}_{T+1}^{(m)}$ is the forecast computed using the full sample for the $m-$th replication. The benchmark forecast, based on the full sample, should perform the best for models with constant parameters. If the relative RMSFE given in equation (18) is less than one, the forecast estimated using the window size chosen by the window selection methods performs better than the benchmark forecast. The smallest number is in bold face excluding the infeasible MSFE criterion (labeled "True").

We summarize the results as follows:

1. The infeasible MSFE criterion almost always produces the smallest relative RMSFEs in all DGPs.[4] While our MSFE criterion is designed for smoothly time-varying parameters, it also works very well for discrete breaks and random-walk parameters.

2. When the parameters are constant (DGPs 1 and 8), most of the relative RMSFEs are close to one. This is because the bias of the estimates is zero when parameters are constant, thus using the full sample size yields the smallest variance of the estimates, which also minimizes the MSFE.

[4] There are three exceptions in which the local linear estimator outperforms the local constant estimator based on the infeasible MSFE criterion. The latter is designed to produce optimal forecasts based on local constant estimators and is not guaranteed to yield better forecasts than local linear estimators.

However, the relative RMSFEs of the approximate MSFE criterion are not necessarily the smallest because we falsely reject the null hypothesis of parameter constancy 5% of the times, in which case we do not use the full sample size.

3. While our method is not designed to handle discrete time breaks, its performance tends to be close to, if not better than, that of the best performing existing methods even when there is a one-time break (DGPs 2, 3, 4, 9–16). Our method tends to work well even when the break is near the end of the sample (DGPs 12 and 16).

4. When the parameters are smoothly time-varying (DGPs 5, 6 and 17–20), the proposed approximate MSFE criteria (OptR1, OptR2 and OptE3) perform very well. The approximate MSFE criteria do not perform as well as the infeasible MSFE criterion ("True") because replacing $\beta(1)$ with the local linear estimates $\tilde{\beta}(1)$ introduces additional noise into the MSFE criterion. The improvements over the existing methods tend to be greater for larger sample sizes.

5. When the parameter follows a random walk process (DGPs 7, 21 and 22), the proposed approximate MSFE criterion tends to perform well. Even when the random coefficient does not evolve as smoothly as in DGPs 5 and 6, the approximate MSFE criterion still works provided that the variance of the error in the random coefficient function is small.

6. When $h = 1$, the local linear estimator based on Cai's (2007) method works well overall. While it can outperform our proposed method, the performance of the local linear estimator is quite sensitive to the DGP and forecast horizons. For example, the local linear estimator tends to perform poorly when $h = 2$. Overall, the rolling OLS estimator based on our method tends to perform similarly or better than the local linear estimators. The equal or better performance of the local constant estimators suggests that the choice of the window size plays a more important role than the choice of the order of polynomials in out-of-sample forecasting performance.

7. Pesaran et al's (2013) method performs well in the case of one regressor (DGPs 1 to 7).[5] Regarding this and the other methods proposed in the literature, overall our method performs well relative to them, and often improves upon them.

To summarize, when the underlying models have smooth time-varying parameters, the improve-

[5] We only report results for PPP for the univariate model case (DGPs 1 to 7), which is the case for which Pesaran et al. (2013) derive their formula. Results for DGPs 8-22 are available upon request. Also, we implement PPP only for $h = 1$, to satisfy their assumptions.

ment in forecasting obtained by choosing the window size by the approximate MSFE criterion is remarkable. Even when the parameters are not smoothly time-varying, the performance of the proposed approximate MSFE criterion is competitive relative to existing methods.

The above results for our method are based on a 5% significance level and trimming rate of 0.15 when implementing Bai and Perron's (1998) test. In Tables 6–9, we report results at the 10% significance level and trimming rate equal to 0.05. The tables show that the results are not very sensitive to these choices.

5 Empirical Analysis

This section examines the practical value of the approximate conditional MSFE criterion developed in Section 2 in forecasting output growth and inflation – see Stock and Watson (1999, 2003, 2007). The latter find strong evidence of instability in predictive relations, which means that finding a predictor useful in one period does not guarantee that it will predict well in later periods.[6] The main results reported in Stock and Watson (1999, 2003, 2007) are based on recursive out-of-sample forecasting, which uses all the data available up to the time the forecast is made, although they also experiment with rolling out-of-sample forecasting using a fixed window size. The purpose of this section is to check whether we can improve forecasts of output growth and inflation using the window size chosen by our new approach.

5.1 Data

We use quarterly data to forecast output growth and inflation for the United States. Quarterly values of monthly series are computed by averaging monthly values over the three months in the quarter. We use the growth rate of real GDP to measure output growth and the GDP deflator to measure inflation. The series of exogenous predictors, described in Table 10, are publicly available from the Federal Reserve Economic Data of the St. Louis Fed. Most of these predictors appear in Stock and Watson (2003, 2007). The exogenous predictors mainly consist of asset prices, measures of real economic activity, price indices and monetary measures. We interpret asset prices as including

[6] For instance, they find that forecasts of output growth based on the term spread, (that is, the long-term government bond rate minus the federal funds rate), improve upon a simple AR model from 1971 through 1984, but are worse than the AR forecasts for the post 1984 period.

interest rates, the interest rate spread and the value of financial assets such as the S&P 500 stock index.

Table 10 lists the data transformation we used in the regressions as well as the mnemonics for the predictors. The full sample starts in 1960:Q1 and ends in 2014:Q3; however, for series with starting date later than 1960:Q1, we use the series from their starting dates through 2014:Q3 as the full sample. The out-of-sample forecast period is 1984:Q1-2014:Q3.

5.2 Forecasting Models

The h-step ahead linear forecasting model for output growth is:

$$y^h_{t+h} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+h}, \tag{19}$$

where the dependent variable is $y^h_{t+h} = (400/h)\ln(Q_{t+h}/Q_t)$, x_t denotes the exogenous predictor, $y_t = 400\ln(Q_t/Q_{t-1})$, and Q_t denotes quarterly real GDP in levels.

The h-step ahead linear forecasting model for inflation is:

$$\pi^h_{t+h} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+h}, \tag{20}$$

where $\pi_t = 400\ln(P_t/P_{t-1})$, $\Delta\pi_t = \pi_t - \pi_{t-1}$, $\pi^h_{t+h} = h^{-1}\sum_{i=1}^{h}\pi_{t+i}$, P_t is the quarterly GDP deflator in levels, and x_t is the exogenous predictor. Furthermore, $\alpha_t(L)x_t$ denotes the lag polynomial, $\alpha_t(L)x_t = \alpha_{1t}x_t + \alpha_{2t}x_{t-1} + \ldots + \alpha_{qt}x_{t-q+1}$, where q is the number of lags. We refer to x_t as a lagged value because it is lagged relative to the dependent variable to be forecast. The same definition applies to $\beta_t(L)y_t$ and $\beta_t(L)\Delta\pi_t$.

5.3 Empirical Results

The results of forecasting one-step-ahead output growth and inflation are summarized in Tables 11–14. The first panel in these tables, labeled "Univariate Models", considers two type of models: autoregressive (AR) models and autoregressive distributed lag (ADL) models, eqs. (19) and (20), respectively. In the AR model, only a constant and lagged values of the variable to be forecast appear as regressors. In the ADL model, regressors include an intercept term, the exogenous variable x_t and the lagged dependent variable (y_t for forecasting output growth or $\Delta\pi_t$ for forecasting inflation). We use ADL models to evaluate the predictive ability of the exogenous predictor in the presence of the lagged dependent variables because, when the series to be forecast is serially correlated, its own past values may be themselves useful predictors.

In Tables 11 and 12, the numbers in column labeled "Fixed" are the RMSFE based on the fixed window of 40 observations, the same window size used by Stock and Watson (2003). In the other columns, the first number is the ratio of the RMSFE based on the window size chosen by our method over the RMSFE based on the fixed window size, and the second number is the p-value of the DM test against the model based on the fixed window size. If the number in the first row is less than one, it means that the optimal window size improves the forecast performance relative to the fixed window size. In Tables 13 and 14, the numbers in the rows labeled "Fixed" are the RMSFE based on the fixed window size and those in the rows labeled "OptR1" are the ratio of the RMSFE based on the window selected by our proposed method (OptR1) over the RMSFE based on the fixed window size.

We investigated choosing the lag length via AIC, BIC or a fixed lag choice (equal to one). The maximum number of lags in the AIC/BIC selection is 2. The AIC/BIC is computed based on the most recent 40 observations. We present the results for the ADL model with BIC lags in Tables 11 and 12, in the panel labeled "ADL(BIC) Models".[7]

Table 11 shows that forecasts of output growth based on the optimal window size often perform better than those based on the fixed window size, for both AR and ADL models. The improvement ranges from 1.3 to 7.7 percent. Table 13 shows that the forecasting improvements based on the optimal window size appear at all horizons.

In ADL models, output growth forecasts obtained by using the federal funds rate, the term spread and the S&P500 as predictors improve when using the optimal window size procedure. These financial variables are useful in predicting output growth in part because they reveal expectations about the future state of the economy. Stock and Watson (2003) found that the term spread is useful for forecasting output growth, and suspected parameter instabilities in the predictive relations. Our results support this conclusion, as our optimal window size procedure allows the model to select

[7]The results for the ADL models based on AIC and the fixed lag are qualitatively similar to those based on BIC in Tables 11 and 12 and thus are omitted to save space. They are available upon request from the authors.

the best amount of past information to forecast at each point in time, and adapting it as time goes by.

When forecasting output growth with measures of real economic activity, the optimal window size seems to improve forecasts for most models and forecast horizons. Improvements appear in forecasts with real disposable personal income ("rdpi") and industrial production ("ip"), among other series. In addition, the optimal window size can improve forecasts based on unemployment measures and price indices, such as PPI, and monetary measures, such as M2.

When forecasting inflation, Table 12 shows that the optimal window sizes hardly beat the fixed window size, especially at shorter forecast horizons. For forecasting models based on measures of economic activities, such as employment, Table 14 shows that long-term inflation forecasts perform better when the optimal window size is used. The empirical evidence on the usefulness of the optimal window size is mixed for inflation forecasts based on monetary measures.

The scatter plots of p-values of the QLR test for parameter constancy (Andrews, 1993) in Figures 1–4 shed light on the difference in the performance of the optimal window size relative to that of the fixed window size. Figures 1 to 4 plot the p-values of the QLR tests (x-axis) recursively implemented at each point in time vis-à-vis the differences in the squared forecast errors at that time; each panel in the figure corresponds to a model, indicated in the title. Figures 1 and 3 show results for forecasting output growth, while Figures 2 and 4 show results for forecasting inflation. The figures suggest that the parameters are unstable in output growth forecast models while they are stable in inflation forecasts. Because our nonparametric approach is both more appropriate and more advantageous when parameters are time-varying, the lack of time-varying parameters may explain why the optimal window size does not perform better than the fixed window size at shorter forecast horizons.

Finally, Tables 15 and 16 report the forecasting performance during the great recession (2007:Q4–2009:Q2). Our method does not perform as well during the latter period. To shed light on the issue, we plot the squared forecast errors in Figures 5 and 6 for a representative predictor. The figures show that there is an outlier which may explain the difference in the performance: while our method is designed to handle smoothly time-varying parameters, it is not designed to handle outliers.

6 Concluding Remarks

We propose a new approach to select the size of the rolling estimation window allowing for smoothly time-varying parameters. Our optimal window size minimizes the conditional MSFEs. Because the true parameter value is unknown, we propose an approximate conditional MSFE criterion in which the unknown value is replaced by a local linear regression estimate. We show that minimizing the approximate conditional MSFE is asymptotically equivalent to minimizing the infeasible conditional MSFE.

Monte Carlo simulations show that using the window size chosen by our approximate conditional MSFE criterion improves the forecasting performance relative to using the full sample when the underlying model is generated by smoothly time-varying as well as random walk parameters. For processes generated by parameters with discrete breaks, the performance of the approximate conditional MSFE criterion is comparable to that of existing window selection methods designed for discrete breaks.

The empirical analysis shows that our new window selection method can improve forecasts of output growth and inflation. In particular, asset prices, housing starts, building permits and monetary measures have useful predictive content for forecasting output growth at short horizons. In general, the improvement at short forecast horizons is more significant for forecasting output growth than inflation, since parameters are more likely to vary in the former than in the latter.

Some caveats are as follows. First, the new window selection method chooses a uniform window size for all the time-varying parameters. When parameters have different patterns of time variation, one window size may not be optimal for all the parameters. Thus, for models with many predictors, the performance of the window selection method would deteriorate. Second, in practice, it is hard for forecasters to know whether the underlying model is subject to discrete breaks or smooth time variation. A careful forecaster should first compare the window selection methods for discrete breaks and the approximate conditional MSFE criterion developed in this paper, and then choose the best window size using the approach developed in this paper.

We focus on the rolling window OLS estimator and the MSFE because they are most commonly used in macroeconomic forecasting. Since our expansions are specific to the choice of models and loss functions, one has to take a stand on the loss function and we chose the MSFE. However one could derive results for asymmetric loss functions, such as those considered in Laurent, Rombouts and Violante (2012), and other estimators, such as local linear estimators. We leave these extensions for future research.

A Appendix: Detailed Proofs

A.1 Notations

For a $p \times 1$ random vector $X \equiv (X_1, \ldots, X_p)'$, $\|X\|_r$ denotes the L_r-norm of X, i.e. $\|X\|_r = (\sum_{i=1}^{p} E(|X_i|^r))^{1/r}$. For a $k \times 1$ real vector $x \equiv (x_1, \ldots, x_k)'$, $\|x\|$ denotes the Euclidean norm of vector x, i.e. $\|x\|^2 = \sum_{i=1}^{k} x_i^2$ and $|x|$ is the max norm of vector x, i.e. $|x| = \max_i |x_i|$. For any $m \times n$ matrix $A \equiv (a_1, a_2, ..., a_n)$, where a_j is the jth column of matrix A, and $a_j = (a_{1j}, a_{2j}, ..., a_{mj})'$ for $j = 1, 2, \ldots, n$, $vec(A)$ is an $mn \times 1$ vector, i.e. $vec(A) \equiv (a_1', a_2', ..., a_n')'$. Let $|A| = \max_{i,j} |a_{ij}|$ where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. For the jth column of matrix A, a_j, $j = 1, 2, \ldots, n$, let $|a_j| = \max_i |a_{ij}|$. If $\{x_n\}_{n=1}^{\infty}$ is any real sequence, $\{a_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers, then $x_n \ll a_n$ denotes $x_n = o(a_n)$, and say that x_n is of smaller order of magnitude than a_n. Conversely $x_n \gg a_n$ denotes $a_n = o(x_n)$. The notation $x_n \simeq a_n$ indicates that there exist $N \geqslant 0$ and finite constants $A > 0$ and $B \geqslant A$, such that $\inf_{n \geqslant N}(x_n/a_n) \geqslant A$ and $\sup_{n \geqslant N}(x_n/a_n) \leqslant B$. This says that $\{x_n\}$ and $\{a_n\}$ ultimately grow at the same rate. Throughout the proofs, let C denote a generic constant that is positive and finite, i.e., $0 < C < \infty$.

A.2 Lemmas

We use the following lemmas in the proof of the theorems.

Lemma 1 $\hat{\beta}_R(1) - \beta(1) \simeq 1/\sqrt{R} + R/T$ *in probability.*

Lemma 2 *Define the infeasible and approximate loss functions by* $L(R) \equiv (\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1))$ *and* $A(R) \equiv (\hat{\beta}_R(1) - \tilde{\beta}(1))' x_T x_T' (\hat{\beta}_R(1) - \tilde{\beta}(1))$*, respectively, where* $\hat{\beta}_R(1)$ *is the estimate of* $\beta(1)$ *based on the most recent* R *observations and* $\tilde{\beta}(1)$ *is the local linear estimate of* $\beta(1)$ *based on the most* R_0 *observations. Then* $\sup_{R \in \Theta_R} |L(R) - A(R)| / L(R) \xrightarrow{p} 0$.

Lemma 3 *For* $k = 0, 1, 2, 3$, $\|R^{-k-1} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t')(T-t)^k / T^k)\| = C/T^k$ *for some* C.

Lemma 4 *For* $k = 0, 1, 2$ *or* 3*, let* $U_t \equiv x_t u_{t+h}$ *and* $S_U \equiv \sum_{t=T-R+1}^{T-h} U_t \, (t-T)^k / T^k$*. Then,*

(a) $Var(S_U) \simeq R^{2k+1}/T^{2k}$.

(b) $R^{-1} S_U \simeq (R^{(2k-1)/2}/T^k)$ *in probability.*

Lemma 5 *For* $k = 0, 1, 2$ *or* 3*, let* $v_t \equiv x_t x_t' - E(x_t x_t')$ *and* $C_t \equiv vec(v_t)$*. Define* $S_C \equiv \sum_{t=T-R+1}^{T-h} C_t \, (t-T)^k \, / T^k$*. Then,*

(a) $\|vec(Var(S_C))\| = O(R^{2k+1}/T^{2k})$*.*

(b) $\|R^{-1} S_C\| = O_p(R^{k-\frac{1}{2}}/T^k)$*.*

Lemma 6 *Let* $B_R \equiv \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'\right)^{-1}$*,* $B_R^* \equiv \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} E(x_t x_t')\right)^{-1}$*. Then*

(a) for a constant δ*,* $0 < \delta < 1/2$*,* $\sup_{R \in \Theta_R} (R^{\frac{1}{2}-\delta}) \, \|vec(B_R) - vec(B_R^*)\| = O_p(1)$*.*

(b) $\|vec(B_R - B_R^*)\| = O_p(1/\sqrt{R})$*.*

Lemma 7 $R^{-2} \sum_{t=T-R+1}^{T-h} \|E(x_t x_t') \left(\beta(\frac{t}{T}) - \beta(\frac{T}{T})\right)\| = C/T$ *for some* C*.*

Lemma 8 *Let* $G_t \equiv (x_t x_t' - E(x_t x_t'))(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))$*, where* x_t *is a* $p \times 1$ *random vector and* $\beta(\cdot)$ *is* $p \times 1$ *and satisfies Assumption 4. Let* $S_G \equiv \sum_{t=T-R+1}^{T-h} G_t$*. Then,*

(a) $\|vec(Var(S_G))\| = O(R^3/T^2)$*.*

(b) $\|R^{-1} S_G\| = O_p(\sqrt{R}/T)$*.*

(c) *For some constant* δ*,* $0 < \delta < 1/2$*,* $\sup_{R \in \Theta_R} \frac{T}{R^{\frac{1}{2}+\delta}} \left\|R^{-1} S_G\right\| = O_p(1)$*.*

Lemma 9 $\sup_{R \in \Theta_R} \min(R^{\frac{1}{2}-\delta}, T/R)(\hat{\beta}_R(1) - \beta(1))' x_T = O_p(1)$*.*

Lemma 10 $\sup_{R \in \Theta_R} \|R^{-1} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t'))\| = O(1)$*.*

Lemma 11 $\sup_{R \in \Theta_R} \|R^{-\frac{1}{2}-\delta} \sum_{t=T-R+1}^{T-h} x_t u_{t+h}\| = O_p(1)$*, where* $0 < \delta < 1/2$*.*

Lemma 12 $\sup_{R \in \Theta_R} \|T R^{-2} \sum_{t=T-R+1}^{T-h} E(x_t x_t') \left(\beta(\frac{t}{T}) - \beta(\frac{T}{T})\right)\| = O(1)$*.*

Lemma 13 *Let* $\tilde{\beta}(1)$ *and* $\tilde{\beta}^{(1)}(1)$ *be local linear estimates defined in eq. (12).* $\tilde{\beta}(1)$ *and* $\tilde{\beta}^{(1)}(1)$ *are estimated using the most recent* R_0 *observations, then* $\tilde{\beta}(1) - \beta(1) = O_p(1/\sqrt{R_0}) + O_p(R_0^2/T^2)$ *and* $\tilde{\beta}^{(1)}(1) - \beta^{(1)}(1) = O_p(T/R_0^{3/2}) + O_p(R_0/T)$*.*

Lemma 14

$$\frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R})}{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2})} \;=\; o(1), \tag{21}$$

where $0 < \delta < \frac{1}{2}$*.*

Lemma 15 *Suppose $\{U_t\}_{t=1}^T$ is a zero-mean stochastic process, where U_t is $p \times 1$. If some $r > 2$, $\{U_t\}$ satisfies: (i) $\{U_t\}$ is an $L_{r/(r-1)}$-NED process of size -2 on $\{V_t\}$ with constants $\{d_t\}$, where $\{V_t\}$ is α-mixing of size $-2r/(r-2)$ and $\{d_t\}_{-\infty}^{+\infty}$ is a sequence of positive constants; (ii) $\{U_t\}$ is L_r-bounded uniformly in t, i.e. $\sup_t \|U_t\|_r < C$ for some C. Then $\{U_t, \mathcal{F}_{-\infty}^t\}$ is an $L_{r/(r-1)}$-mixingale of size -2 with constants $c_t = O(max\{\|U_t\|_r, d_t\})$.*

Lemma 16 *Define $B_k = \sum_{j=h}^{R} j^k/R^{k+1}$ for $k > -1$, where $0 < h < \infty$, $R \geq h$, $h \in \mathbb{Z}^+$ and $R \in \mathbb{Z}^+$. Then as $R \to \infty$, $B_k \simeq C$ for some C.*

Lemma 17 *For $k > -1$, $\sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} m^{-2}(t-T)^k(t+m-T)^k/T^{2k} = O(R^{2k+1}/T^{2k})$, where $0 < h < \infty$, $R \geq h$, $h \in \mathbb{Z}^+$ and $R \in \mathbb{Z}^+$.*

Lemma 18 *For $k > -1$, $\sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} m^{-2}(t-T)^k(t-m-T)^k/T^{2k} = O(R^{2k+1}/T^{2k})$, where $0 < h < \infty$, $R \geq h$, $h \in \mathbb{Z}^+$ and $R \in \mathbb{Z}^+$.*

Lemma 19 *Given square matrices A and B, where A and $A+B$ are invertible. Then $(A+B)^{-1} = A^{-1} - (I + A^{-1}B)^{-1}A^{-1}BA^{-1}$.*

A.3 Proofs of Theorems

Proof of Theorem 1. It follows from Lemma 1 that

$$(\hat{\beta}_R(1) - \beta(1))' x_T x_T' (\hat{\beta}_R(1) - \beta(1)) \simeq (1/\sqrt{R} + R/T)^2 \simeq 1/R + R^2/T^2 \tag{22}$$

Differentiating $1/R + R^2/T^2$ with respect to R and set it to zero, the optimal window size is at the rate of $T^{2/3}$ in probability. Since the second order derivative of $1/R + R^2/T^2$ is always positive, the optimal window size minimizes the objective function. ■

Proof of Theorem 2. Let $a(R) \equiv \hat{\beta}_R(1) - \beta(1)$ and $b \equiv \tilde{\beta}(1) - \beta(1)$ where $\hat{\beta}_R(1)$ and $\tilde{\beta}(1)$ are the local constant and linear estimates of $\beta(1)$ based on the last R and R_0 observations, respectively. Then we can write the infeasible and approximate loss functions as $L(R) = a(R)' x_T x_T' a(R)$ and $A(R) = (a(R) - b)' x_T x_T' (a(R) - b)$. We choose the optimal window size $\hat{R}$ to minimize the approximate loss function $A(R)$, i.e. $\hat{R} = \arg\min_{R \in \Theta_R}[A(R)]$. Let $\hat{R}'$ denote the window size which minimizes the infeasible loss function $L(R)$, i.e. $\hat{R}' = \arg\min_{R \in \Theta_R}[L(R)]$.

By expanding $A(\hat{R})$, eq. (14) can be written as

$$\frac{A(\hat{R})}{\inf_{R\in\Theta_R} L(R)} = \frac{L(\hat{R})}{\inf_{R\in\Theta_R} L(R)} - 2\frac{a(\hat{R})'x_T x_T' b}{\inf_{R\in\Theta_R} L(R)} + \frac{b' x_T x_T' b}{\inf_{R\in\Theta_R} L(R)} = I_1 - 2I_2 + I_3 p$$

We show that $I_1 \xrightarrow{p} 1$, $I_2 \xrightarrow{p} 0$ and $I_3 \xrightarrow{p} 0$. Since $L(R) > 0$ for any R, we get

$$\sup_{R,R'\in\Theta_R}\left|\frac{L(R)-L(R')-(A(R)-A(R'))}{L(R)+L(R')}\right| \leq \sup_{R\in\Theta_R}\left|\frac{L(R)-A(R)}{L(R)}\right| + \sup_{R'\in\Theta_R}\left|\frac{L(R')-A(R')}{L(R')}\right|.$$

It follows from Lemma 2 that

$$\sup_{R,R'\in\Theta_R}\left|\frac{L(R)-L(R')-(A(R)-A(R'))}{L(R)+L(R')}\right| \xrightarrow{p} 0. \tag{23}$$

Then, for any $\epsilon > 0$, there is

$$P\left[\frac{L(\hat{R})-L(\hat{R}')-(A(\hat{R})-A(\hat{R}'))}{L(\hat{R})+L(\hat{R}')} \leq \epsilon\right] \to 1.$$

This implies that $(1-\epsilon)L(\hat{R}) - (1+\epsilon)L(\hat{R}') \leq A(\hat{R}) - A(\hat{R}') \leq 0$, with probability approaching one, which results in $1 \leq L(\hat{R})/L(\hat{R}') \leq (1+\epsilon)(1-\epsilon)$ with probability approaching one. Thus $I_1 \xrightarrow{p} 1$. The results, $I_2 \xrightarrow{p} 0$ and $I_3 \xrightarrow{p} 0$, follow from the proofs of Appendix A.4 and Appendix A.4. Combining these results completes the proof. ■

A.4 Proofs of Lemmas

Proof of Lemma 1. Expanding $\hat{\beta}_R(1) - \beta(1)$ pointwise for each $R \in [\underline{R}, \bar{R}]$, we have

$$\begin{aligned}\hat{\beta}_R(1) - \beta(1) &= \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'\right)^{-1}\left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t u_{t+h}\right) \\ &\quad + \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'\right)^{-1}\left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))\right) = E_1 + E_2.\end{aligned}$$

Let $B_R \equiv \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'\right)^{-1}$, $H_R \equiv \frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t u_{t+h}$, $Q_R \equiv \frac{1}{R}\sum_{t=T-R+1}^{T-h} x_t x_t'(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))$, $B_R^* \equiv \left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} E(x_t x_t')\right)^{-1}$ and $Q_R^* \equiv \frac{1}{R}\sum_{t=T-R+1}^{T-h} E(x_t x_t')(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))$. Then

$\hat{\beta}_R(1) - \beta(1)$ can be abbreviated as

$$\hat{\beta}_R(1) - \beta(1) = B_R H_R + B_R Q_R = E_1 + E_2 \tag{24}$$

First, we check the rate of the first term E_1. Write E_1 as $E_1 = B_R^* H_R + (B_R - B_R^*)H_R$. By Lemma 3, we have $\|vec(B_R^*)\| \simeq C$ for some C. By Lemma 4, we know $H_R \simeq R^{-1/2}$. Thus $B_R^* H_R \simeq R^{-1/2}$. By Lemma 5 and Lemma 6(b), we know that $\|B_R - B_R^*\| = O_p(R^{-1/2})$. Then $E_1 \simeq R^{-1/2} + O_p(R^{-1}) \simeq R^{-1/2}$.

Next we need to find the rate of the second term E_2. The rate of E_2 follows from

$$\begin{aligned} E_2 =& B_R Q_R = B_R^* Q_R^* + B_R^*(Q_R - Q_R^*) + (B_R - B_R^*)Q_R^* + (B_R - B_R^*)(Q_R - Q_R^*) \\ \simeq& cR/T + O_p(\sqrt{R}/T) + O_p(1/\sqrt{R})cR/T + O_p(1/\sqrt{R})O_p(\sqrt{R}/T) \simeq R/T, \end{aligned}$$

for some C, because by Lemma 7 $\|Q_R^*\| \simeq R/T$ and by Lemma 8(b) $Q_R - Q_R^* = O_p(\sqrt{R}/T)$. Combining the rate of E_1 and E_2 yields $\hat{\beta}_R(1) - \beta(1) \simeq 1/\sqrt{R} + R/T$ in probability. ■

Proof of Lemma 2. The distance $A(R)$ can be decomposed as

$$\begin{aligned} A(R) =& (a(R) - b)' x_T x_T' (a(R) - b) \\ =& a(R)' x_T x_T' a(R) - 2a(R)' x_T x_T' b + b' x_T x_T' b \\ =& L(R) - 2a(R)' x_T x_T' b + b' x_T x_T' b, \end{aligned}$$

so it is equivalent to show that

$$\sup_{R \in \Theta_R} \left| \frac{-2a(R)' x_T x_T' b + b' x_T x_T' b}{L(R)} \right| \xrightarrow{p} 0. \tag{25}$$

By the triangular inequality, we have

$$\sup_{R \in \Theta_R} \left| \frac{-2a(R)' x_T x_T' b + b' x_T x_T' b}{L(R)} \right| \leq \sup_{R \in \Theta_R} \left| \frac{-2a(R)' x_T x_T' b}{L(R)} \right| + \sup_{R \in \Theta_R} \left| \frac{b' x_T x_T' b}{L(R)} \right|. \tag{26}$$

Because $\sup_{R \in \Theta_R}(\min(R^{\frac{1}{2}-\delta}, T/R)a(R)' x_T) = O_p(1)$ and $\min(R_0^{\frac{1}{2}}, T^2/R_0^2) b' x_T = O_p(1)$ by

Lemma 9 and Lemma 13, we have

$$\begin{aligned}
\sup_R \left| \frac{a(R)' x_T x_T' b}{a(R)' x_T x_T' a(R)} \right| &= \sup_R \left| \frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R})}{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2})} \frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R}) a(R)' x_T \min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2}) b' x_T}{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R}) a(R)' x_T (R^{\frac{1}{2}-\delta}, \frac{T}{R}) a(R)' x_T} \right| \\
&= O_p\left(\sup_R \frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R})}{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2})} \right) = o_p(1), \qquad (27)
\end{aligned}$$

and

$$\begin{aligned}
\sup_R \left| \frac{b' x_T x_T' b}{a(R)' x_T x_T' a(R)} \right| &= \sup_R \left| \left(\frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R})}{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2})} \right)^2 \frac{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2}) b' x_T \min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2}) b' x_T}{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R}) a(R)' x_T \min(R^{\frac{1}{2}-\delta}, \frac{T}{R}) a(R)' x_T} \right| \\
&= O_p\left(\sup_R \left(\frac{\min(R^{\frac{1}{2}-\delta}, \frac{T}{R})}{\min(R_0^{\frac{1}{2}}, \frac{T^2}{R_0^2})} \right)^2 \right) = o_p(1), \qquad (28)
\end{aligned}$$

where the last equalities in (27) and (28) follow from Lemma 14. Equations (26), (27) and (28) completes the proof of eq. (25). ■

Proof of Lemma 3. $\|R^{-k-1} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t')(T-t)^k / T^k)\|$ is bounded as

$$\begin{aligned}
\left\| \inf_t \left(vec(E(x_t x_t')) \right) \frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} \left(\frac{T-t}{T} \right)^k \right\| &\leq \left\| \frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t')) \left(\frac{T-t}{T} \right)^k \right\| \\
&\leq \left\| \sup_t \left(vec(E(x_t x_t')) \right) \frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} \left(\frac{T-t}{T} \right)^k \right\| \qquad (29)
\end{aligned}$$

It follows from Lemma 16 that

$$\frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} \left(\frac{T-t}{T} \right)^k = \frac{1}{R^{k+1}} \sum_{j=h}^{R-1} \frac{j^k}{T^k} \simeq \frac{1}{R^{k+1}} \sum_{j=h}^{R} \frac{j^k}{T^k} \simeq \frac{C}{T^k} \qquad (30)$$

for some C. Then eq. (29) becomes

$$\left\| \inf_t \left(vec(E(x_t x_t')) \right) \right\| \frac{C}{T^k} \leq \left\| \frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t')) \left(\frac{T-t}{T} \right)^k \right\| \leq \left\| \sup_t \left(vec(E(x_t x_t')) \right) \right\| \frac{C}{T^k}.$$

It follows from Assumptions 2(ii) and 3 that $1/C < \|vec(E(x_t x_t'))\| < C$ uniformly in t for some

C. Thus

$$\left\| \frac{1}{R^{k+1}} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t')) \left(\frac{T-t}{T} \right)^k \right\| = \frac{C}{T^k},$$

for some C. ■

Proof of Lemma 4. **(a)** The long-run variance of S_U is given by

$$\begin{aligned}
vec(Var(S_U)) &= \sum_{t=T-R+1}^{T-h} \left[vec(E(U_t U_t')) \left(\frac{t-T}{T} \right)^{2k} \right] \\
&+ \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[vec\left(E(U_t U_{t-m}')\right) \left(\frac{t-T}{T} \right)^k \left(\frac{t-m-T}{T} \right)^k \right] \\
&+ \sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} \left[vec\left(E(U_t U_{t+m}')\right) \left(\frac{t-T}{T} \right)^k \left(\frac{t+m-T}{T} \right)^k \right] \\
&= A_1 + A_2 + A_3.
\end{aligned}$$

First we show the rate of A_1. A_1 is bounded as

$$\inf_t \left(vec(E(U_t U_t')) \right) \sum_{t=T-R+1}^{T-h} \left(\frac{t-T}{T} \right)^{2k} \leq A_1 \leq \sup_t \left(vec(E(U_t U_t')) \right) \sum_{t=T-R+1}^{T-h} \left(\frac{t-T}{T} \right)^{2k}. \quad (31)$$

Since it follows from Lemma 16 that

$$\sum_{t=T-R+1}^{T-h} \left(\frac{t-T}{T} \right)^{2k} = \sum_{j=h}^{R-1} \frac{j^{2k}}{T^{2k}} \simeq \sum_{j=h}^{R} \frac{j^{2k}}{T^{2k}} \simeq \frac{R^{2k+1}}{T^{2k}}, \quad (32)$$

eq. (31) becomes

$$\inf_t \left(vec(E(U_t U_t')) \right) \frac{CR^{2k+1}}{T^{2k}} \leq A_1 \leq \sup_t \left(vec(E(U_t U_t')) \right) \frac{CR^{2k+1}}{T^{2k}}.$$

It follows from Assumption 1 that $\|vec(E(U_t U_t'))\| \equiv \|vec(E(\sigma_t^2 x_t x_t'))\| \simeq \|vec(E(x_t x_t'))\|$. Since $1/C < \|vec(E(x_t x_t'))\| < C$ uniformly in t for some C by Assumptions 2(ii) and 3, $A_1 \simeq R^{2k+1}/T^{2k}$.

Next we find the rate of A_2. Let $U_{i,t}$ denote the ith element of U_t. Thus we have

$$A_2 \simeq \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\left\| vec\left(E(U_t U'_{t-m})\right) \right\| \left(\frac{t-T}{T}\right)^k \left(\frac{t-m-T}{T}\right)^k \right]$$
$$= \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\left(\sum_{i=1}^{p} \sum_{j=1}^{p} (E(U_{i,t}U_{j,t-m}))^2 \right)^{1/2} \left(\frac{t-T}{T}\right)^k \left(\frac{t-m-T}{T}\right)^k \right].$$

To show the rate of $E(U_{i,t}U_{j,t-m})$, we need to show all the conditions in Lemma 15 hold. First U_t is zero mean by Assumption 1. Next, the condition (i) in Lemma 15 holds because by Assumption 2(i), Theorem 17.9 in Davidson (1994, p.268), and the Lyapunov inequality (see 9.23 in Davidson, 1994, p.139), which ensures that Assumption 2(i) implies $U_t \equiv x_t u_{t+h}$ is $L_{2r/(r-1)}$-NED of size -2 on $\{V_t\}$, where $\{V_t\}$ is α-mixing of size $-2r/(r-2)$. Also by the Lyapunov inequality, $L_{2r/(r-1)}$-NED of size -2 implies $L_{r/(r-1)}$-NED of size -2 when $r > 2$. The condition (ii) of Lemma 15 directly is implied by Assumption 2(ii). Then applying Lemma 15, we know $\{U_t\}$ is $L_{r/(r-1)}$-mixingale of size -2. Thus, for each $i, j = 1, 2, \ldots, p$,

$$\begin{aligned} |E(U_{i,t}U_{j,t-m})| &\leq \|U_{i,t}\|_{\frac{1}{r-1}} \|E_{t-m}(U_{j,t-m})\|_{\frac{r}{r-1}} \\ &= O(\zeta_m) \ = \ O(m^{-2}). \end{aligned} \tag{33}$$

Since p is finite, we have

$$A_2 = O\left(\sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\frac{1}{m^2} \left(\frac{t-T}{T}\right)^k \left(\frac{t-m-T}{T}\right)^k \right] \right) = O\left(\frac{R^{2k+1}}{T^{2k}}\right), \tag{34}$$

where the last equality is derived by Lemma 18.

Finally we need to find the rate of A_3. Following similar arguments used to prove eq. (34), we can show that

$$A_3 = O\left(\frac{R^{2k+1}}{T^{2k}}\right).$$

Combining the results for A_1, A_2 and A_3, we have $vec(Var(S_U)) \simeq R^{2k+1}/T^{2k}$.

(b) The proof for the rate of $R^{-1}S_U$ follows from the CLT for NED processes, so we need to show that the conditions of Theorem 24.6 and Corollary 24.7 in Davidson (1994, p. 386) hold. Define $U_t^* \equiv (Var(S_U))^{1/2} U_t (t-T)^k / T^k$ and $S_U^* \equiv \sum_{t=T-R+1}^{T-h} U_t^*$. Denote $s \equiv T-t$, where $T-R+1 \leq t \leq T-h$, so s satisfies $h \leq s \leq R-1$. Let $\{Z_{Rs}, s=h,\ldots,R-1, R \geq h+1\}$ be a triangular stochastic array, such that $Z_{Rs} \equiv U_{T-s}^*$, then

$$Z_{Rs} \equiv (Var(S_U))^{1/2} x_{T-s} u_{T-s+h} \frac{s^k}{T^k} \simeq \frac{s^k}{R^{k+\frac{1}{2}}} x_{T-s} u_{T-s+h},$$

where the last expression follows from part (a). Let $S_R \equiv \sum_{s=h}^{R-1} Z_{Rs}$. Notice that $S_R \equiv S_U^*$.

First, we note that the condition (a) of Theorem 24.6 in Davidson (1994, p. 386) holds because $E(U_t^*) = 0$ and $E(S_U^* S_U^{*\prime}) = I_p$ implies that $E(Z_{Rs}) = 0$ and $E(S_R S_R') = I_p$. We next show that the condition (b) of Theorem 24.6 in Davidson (1994, p.386) and the condition (d') of Corollary 24.7 in Davidson (1994, p. 387) hold. Suppose a positive constant array $\{c_{Rs}\}$ satisfies $c_{Rs} = R^{-1/2}$, then it follows from Assumption 2(ii) we have

$$\sup_{R,s} \|Z_{Rs}/c_{Rs}\|_r \simeq \sup_{R,s} \left\| \frac{s^k}{R^{k+\frac{1}{2}} c_{Rs}} x_{T-s} u_{T-s+h} \right\|_r \leq \sup_{R,s} \frac{s^k}{R^{k+\frac{1}{2}} c_{Rs}} \cdot \sup_s \|x_{T-s} u_{T-s+h}\|_r < \infty.$$

Hence, condition (b) of Theorem 24.6 in Davidson (1994, p. 386) is satisfied. Define $M_R \equiv \max_{h \leq s \leq R-1}\{c_{Rs}\}$. Condition (d') of Corollary 24.7 in Davidson (1994, p. 387) holds because $M_R = R^{-1/2}$ and $\sup_R R M_R^2 = 1 < \infty$. We now show that condition (c') of Corollary 24.7 in Davidson (1994, p. 387) holds. We know from part (a) that U_t is $L_{2r/(r-1)}$-NED of size -2 on $\{V_t\}$, which is α-mixing of size $-2r/(r-2)$. When $r > 2$, we have $2 < 2r/(r-1) < 4$, thus $L_{2r/(r-1)}$-NED of size -2 implies L_2-NED of size -1. We also know that α-mixing of size $-2r/(r-2)$ implies α-mixing of size $-r/(r-2)$. Hence condition (c') of Corollary 24.7 in Davidson (1994, p.387) is satisfied. Thus we can conclude that $S_R = S_U^* \xrightarrow{d} N(0, I_p)$ and $S_U \simeq R^{k+1/2}/T^k$ in probability. Then $R^{-1} S_U \simeq (R^{(2k-1)/2}/T^k)$ in probability. ■

Proof of Lemma 5. **(a)** Note that v_t is $p \times p$ and S_C is $p^2 \times 1$. For $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$, let $v_{ij,t}$ denote the ith row and jth column element of v_t. For $l = 1, 2, \ldots, p^2$, let $C_{l,t}$ denote the lth element of C_t and let S_{Cl} denote the lth element of S_C. Since $\{C_t\}$ is zero-mean, the long-run variance of S_C is given by

$$\|vec(Var(S_C))\| = \|vec(E(S_C S_C'))\| \leq \sum_{t=T-R+1}^{T-h} \left[\|vec(E(C_t C_t'))\| \left(\frac{t-T}{T} \right)^{2k} \right]$$
$$+ \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\|vec\left(E(C_t C_{t-m}')\right)\| \left(\frac{t-T}{T} \right)^{k} \left(\frac{t-m-T}{T} \right)^{k} \right]$$
$$+ \sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} \left[\|vec\left(E(C_t C_{t+m}')\right)\| \left(\frac{t-T}{T} \right)^{k} \left(\frac{t+m-T}{T} \right)^{k} \right] = D_1 + D_2 + D_3$$

First, we show the rate of D_1. It follows from the Hölder inequality and Assumption 2(ii) that

$$D_1 = \sum_{t=T-R+1}^{T-h} \left[\left(\sum_{i=1}^{p^2} \sum_{j=1}^{p^2} (E(C_{i,t} C_{j,t}))^2 \right)^{1/2} \left(\frac{t-T}{T} \right)^{2k} \right] \leq \sum_{t=T-R+1}^{T-h} \left[p^2 C \left(\frac{t-T}{T} \right)^{2k} \right]. \quad (35)$$

It follows from Lemma 16, eq. (32) and the finiteness of p that $D_1 = O(R^{2k+1}/T^{2k})$.

Next we check the rate of D_2. Expanding D_2 gives

$$D_2 = \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\left(\sum_{i=1}^{p^2} \sum_{j=1}^{p^2} (E(C_{i,t} C_{j,t-m}))^2 \right)^{1/2} \left(\frac{t-T}{T} \right)^{k} \left(\frac{t-m-T}{T} \right)^{k} \right].$$

Following the proof for the rate of A_2 in Lemma 4(a) and applying Lemma 15, we know $\{C_t\}$ is $L_{r/(r-1)}$-mixingale of size -2. Thus, for each $i, j = 1, 2, \ldots, p^2$

$$\begin{aligned} |E(C_{i,t} C_{j,t-m})| &\leq \|C_{i,t}\|_{\frac{1}{r-1}} \|E_{t-m}(C_{j,t-m})\|_{\frac{r}{r-1}} \\ &= O(\zeta_m) \ = \ O(m^{-2}). \end{aligned} \quad (36)$$

Then we have

$$D_2 = O\left(\sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left[\frac{1}{m^2} \left(\frac{t-T}{T} \right)^{k} \left(\frac{t-m-T}{T} \right)^{k} \right] \right) = O\left(\frac{R^{2k+1}}{T^{2k}} \right),$$

where the last equality is derived by Lemma 18.

Finally the rate of D_3 can be obtained by using the similar arguments for D_2: $D_3 = O\left(\frac{R^{2k+1}}{T^{2k}}\right)$.

Combining the results for D_1, D_2 and D_3, we have $\|vec(Var(S_C))\| = O(R^{2k+1}/T^{2k})$.

(b) By Chebyshev's inequality for vectors, we have that, for any real number $\epsilon > 0$

$$P\left[\left\|\frac{T^k}{R^{k-\frac{1}{2}}}\frac{1}{R}S_C\right\| > \epsilon\right] \leq \frac{\|vec(Var(\frac{T^k}{R^{k-\frac{1}{2}}}\frac{1}{R}S_C))\|}{\epsilon^2} = \frac{\frac{T^{2k}}{R^{2k+1}}\|vec(Var(S_C))\|}{\epsilon^2} = O(1),$$

where the last equality follows from the result in part (a). Therefore $\|R^{-1}S_C\| = O_p(R^{k-\frac{1}{2}}/T^k)$. ■

Proof of Lemma 6. **(a)**. First we note that there is a continuously differentiable function $f(\cdot)$ such that $f(vec(B_R^{-1})) - f(vec(B_R^{*-1})) = vec(B_R) - vec(B_R^*)$. Then there exists an open neighborhood $N(vec(B_R^{*-1}))$ of $vec(B_R^{*-1})$ such that $\sup_{v\in N(vec(B_R^{*-1}))}|f_\nu(\nu)| < d$, where d is a constant, $0 < d < \infty$, and $f_\nu(\nu) \equiv \partial f(\nu)/\partial\nu$. Next by taking the first order Taylor expansion of $f(\cdot)$ around $vec(B_R^{*-1})$, we get

$$\begin{aligned}
&\sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\,\|vec(B_R) - vec(B_R^*)\| = \sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\left\|f_\nu(\tilde{\nu})(vec(B_R^{-1}) - vec(B_R^{*-1}))\right\| \\
&\leq \left(\sup_{R\in\Theta_R}|f_\nu(\tilde{\nu})|\right)\left(\sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\left\|vec(B_R^{-1}) - vec(B_R^{*-1})\right\|\right) \qquad (37)
\end{aligned}$$

where $\tilde{\nu}$ satisfies that $\|\tilde{\nu} - vec(B_R^{*-1})\| \leq \|vec(B_R^{-1}) - vec(B_R^{*-1})\|$. Then it is sufficient to show that $\sup_{R\in\Theta_R}|f_\nu(\tilde{\nu})| = O_p(1)$ and $\sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\|vec(B_R^{-1}) - vec(B_R^{*-1})\| = O_p(1)$.

First we need to show that $\sup_{R\in\Theta_R}|f_\nu(\tilde{\nu})| = O_p(1)$. Following the notation in Lemma 5 and the result of Lemma 5(a), we have $\|vec(Var(S_C))\| = O(R^{2k+1}/T^{2k})$. So when $k = 0$,

$$\|vec(Var(S_C))\| \equiv \left\|vec\left(Var\left(\sum_{t=T-R+1}^{T-h}vec(x_tx_t' - E(x_tx_t'))\right)\right)\right\| = O(R).$$

Let $B \equiv vec(B_R^{-1}) - vec(B_R^{*-1})$, then we can write $\sum_{t=T-R+1}^{T-h}vec(x_tx_t' - E(x_tx_t')) \equiv RB$. Therefore, $\|vec(Var(B))\| = O(R^{-1})$. Then for any $\epsilon > 0$,

$$\begin{aligned}
&P\left(\sup_{R\in\Theta_R}\left\|R^{\frac{1}{2}-\delta}B\right\| > \epsilon\right) \leq \#\Theta_R\cdot\sup_{R\in\Theta_R}P\left(\left\|R^{\frac{1}{2}-\delta}B\right\| > \epsilon\right) \\
&\leq \#\Theta_R\cdot\sup_{R\in\Theta_R}\frac{\|vec(Var(R^{\frac{1}{2}-\delta}B))\|}{\epsilon^2} \leq \#\Theta_R\cdot\sup_{R\in\Theta_R}\frac{C}{R^{2\delta}\epsilon^2} = O(1),
\end{aligned}$$

for some C. Hence, $vec(B_R^{-1}) \xrightarrow{p} vec(B_R^{*-1})$ uniformly in R. Then for any $\epsilon > 0$, there exists sufficiently large R such that $P(\tilde{\nu} \in N(vec(B_R^{*-1}))) > 1 - \epsilon$. This implies that for sufficiently large

R, $|f_\nu(\tilde{\nu})| \leq \sup_{\nu \in N(vec[B_R^{-1}])} |f_\nu(\nu)| < C$ uniformly in R with probability greater than $1-\epsilon$, for some C. Thus $\sup_{R\in\Theta_R} |f_\nu(\tilde{\nu})| = O_p(1)$.

Following the proof above, we have just shown that

$$\sup_{R\in\Theta_R} \|R^{\frac{1}{2}-\delta} B\| \equiv \sup_{R\in\Theta_R} (R^{\frac{1}{2}-\delta}) \|vec(B_R^{-1}) - vec(B_R^{*\,-1})\| = O_p(1).$$

The product of the rate of $\sup_{R\in\Theta_R} |f_\nu(\tilde{\nu})|$ and $\sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\|vec(B_R^{-1}) - vec(B_R^{*\,-1})\|$ gives the result .

(b) As shown in eq. (37),

$$\begin{aligned}
&\|vec(B_R) - vec(B_R^*)\| = \left\| f_\nu(\tilde{\nu})(vec(B_R^{-1}) - vec(B_R^{*\,-1}))\right\| \\
&\leq |f_\nu(\tilde{\nu})| \cdot \left\| vec\left(\frac{1}{R}\sum_{t=T-R+1}^{T-h} (x_t x_t' - E(x_t x_t'))\right)\right\| = O_p(1/\sqrt{R}),
\end{aligned}$$

where the last equality follows from Lemma 5 and the proof in (a). ■

Proof of Lemma 7. Applying the Taylor expansion to $\beta(\frac{t}{T})$ around $\frac{T}{T}$ yields

$$\left\|\frac{1}{R^2}\sum_{t=T-R+1}^{T-h}\left[E(x_t x_t')\left(\beta(\frac{t}{T}) - \beta(\frac{T}{T})\right)\right]\right\| = \left\|\frac{1}{R^2}\sum_{t=T-R+1}^{T-h}\left[E(x_t x_t')\beta^{(1)}(c)\left(\frac{t-T}{T}\right)\right]\right\| \tag{38}$$

where $c = \lambda\frac{t}{T} + (1-\lambda)\frac{T}{T}$, for some $\lambda \in (0,1)$. Then eq. (38) is bounded as

$$\left\|\inf_t [E(x_t x_t')\beta^{(1)}(c)] \cdot \frac{1}{R^2}\sum_{t=T-R+1}^{T-h}(\frac{t-T}{T})\right\| \leq eq.\ (38) \leq \left\|\sup_t [E(x_t x_t')\beta^{(1)}(c)] \cdot \frac{1}{R^2}\sum_{t=T-R+1}^{T-h}(\frac{t-T}{T})\right\| \tag{39}$$

Based on Lemma 16, we derive

$$\frac{1}{R^2}\sum_{t=T-R+1}^{T-h}\frac{t-T}{T} = -\frac{1}{R^2}\sum_{j=h}^{R-1}\frac{j}{T} \simeq \frac{1}{R^2}\sum_{j=h}^{R}\frac{j}{T} = \frac{C}{T},$$

for some C. Then eq. (39) becomes

$$\left\|\inf_t [E(x_t x_t')\beta^{(1)}(c)] \cdot \frac{c_1}{T}\right\| \leq eq.\ (38) \leq \left\|\sup_t [E(x_t x_t')\beta^{(1)}(c)] \cdot \frac{c_1}{T}\right\|$$

From Assumption 2 and Assumption 4, we know that $\|E(x_t x_t')\beta^{(1)}(c)\| \leq C$ uniformly in t.

Based on Assumption 3 and Assumption 4, we also know that $\|E(x_t x_t')\beta^{(1)}(c)\| > 0$ uniformly in t. Thus eq. (38)$= C/T$ for some C. ■

Proof of Lemma 8. **(a)** Denote $v_t \equiv x_t x_t' - E(x_t x_t')$, which is $p \times p$. For $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$, $v_{ij,t}$ denotes the ith row and jth column element of v_t. Note that $\{G_t\}$ is zero-mean. Then applying the triangle inequality gives

$$\begin{aligned}
&\|vec(Var(S_G))\| = \left\| vec\left(E\left[\left(\sum_{t=T-R+1}^{T-h} G_t \right) \left(\sum_{t=T-R+1}^{T-h} G_t \right)' \right] \right) \right\| \\
&\leq \sum_{t=T-R+1}^{T-h} \left\| vec(E(G_t G_t')) \right\| + \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left\| vec(E(G_t G_{t-m}')) \right\| \\
&+ \sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} \left\| vec(E(G_t G_{t+m}')) \right\| = J_1 + J_2 + J_3
\end{aligned}$$

First we show the rate of J_1. Expanding $\|vec(E(G_t G_t'))\|$ gives

$$\begin{aligned}
&\left\| vec(E(G_t G_t')) \right\| = \left\| vec\left(E\left[v_t(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))' v_t' \right] \right) \right\| \\
&= \left\{ \sum_{i=1}^{p} \sum_{l=1}^{p} \left[E\left(\left(\sum_{j=1}^{p} v_{ij,t}(\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T})) \right) \left(\sum_{k=1}^{p} v_{lk,t}(\beta_k(\frac{t}{T}) - \beta_k(\frac{T}{T})) \right) \right) \right]^2 \right\}^{1/2}
\end{aligned}$$

It follows from Assumption 4 that there exists some C, such that $\left|\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T})\right| \leq \left|\frac{t-T}{T}\right| C$ for each j. Then

$$\begin{aligned}
\left\| vec(E(G_t G_t')) \right\| &\leq \left\{ \sum_{i=1}^{p} \sum_{l=1}^{p} \left[E\left(\left(\sum_{j=1}^{p} |v_{ij,t}| \left| \frac{t-T}{T} \right| C \right) \left(\sum_{k=1}^{p} |v_{lk,t}| \left| \frac{t-T}{T} \right| C \right) \right) \right]^2 \right\}^{1/2} \\
&= \left(\frac{t-T}{T} \right)^2 C^2 \cdot \left\{ \sum_{i=1}^{p} \sum_{l=1}^{p} \left[E\left(\left(\sum_{j=1}^{p} |v_{ij,t}| \right) \left(\sum_{k=1}^{p} |v_{lk,t}| \right) \right) \right]^2 \right\}^{1/2} \\
&= \left(\frac{t-T}{T} \right)^2 C^2 \cdot \left\{ \sum_{i=1}^{p} \sum_{l=1}^{p} \left[\sum_{j=1}^{p} \sum_{k=1}^{p} E(|v_{ij,t} v_{lk,t}|) \right]^2 \right\}^{1/2} \qquad (40)
\end{aligned}$$

By the Cauchy-Schwartz inequality, we know that

$$E(|v_{ij,t} v_{lk,t}|) \leq \|v_{ij,t}\|_2 \|v_{lk,t}\|_2.$$

Because $\{v_{ij,t}\}$ is L_r-bounded uniformly in t for some $r > 2$ by Assumption 2(ii), $\|v_{ij,t}\|_2$ is finite for each $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$. Thus since p is finite, we have

$$\left\{\sum_{i=1}^{p}\sum_{l=1}^{p}\left[\sum_{j=1}^{p}\sum_{k=1}^{p}E(|v_{ij,t}v_{lk,t}|)\right]^2\right\}^{1/2} \le \left\{\sum_{i=1}^{p}\sum_{l=1}^{p}\left[\sum_{j=1}^{p}\sum_{k=1}^{p}\|v_{ij,t}\|_2\|v_{lk,t}\|_2\right]^2\right\}^{1/2} = O(1).$$

Then it follows from eq. (40) that

$$J_1 = \sum_{t=T-R+1}^{T-h}\|vec(E(G_tG_t'))\| \le \sum_{t=T-R+1}^{T-h}\left(\frac{t-T}{T}\right)^2 M^2 \cdot O(1) = O(R^3/T^2).$$

The last equality follows from Lemma 16 and eq. (32). Hence, the rate of J_1 is $O(R^3/T^2)$.

Next we need to show the rate of J_2. Again expanding $\|vec(E(G_tG_{t-m}'))\|$ gives

$$\|vec(E(G_tG_{t-m}'))\| = \left\|vec\left(E\left[v_t(\beta(\frac{t}{T}) - \beta(\frac{T}{T}))(\beta(\frac{t-m}{T}) - \beta(\frac{T}{T}))'v_{t-m}'\right]\right)\right\|$$
$$=\left\{\sum_{i=1}^{p}\sum_{l=1}^{p}\left[E\left(\left(\sum_{j=1}^{p}v_{ij,t}(\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T}))\right)\left(\sum_{k=1}^{p}v_{lk,t-m}(\beta_k(\frac{t-m}{T}) - \beta_k(\frac{T}{T}))\right)\right)\right]^2\right\}^{1/2}$$

By the Lipschitz condition implied by Assumption 4, there exists some C such that for each $j = 1, 2, \ldots, p$, $\left|\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T})\right| \le \left|\frac{t-T}{T}\right| C$ and $\left|\beta_j(\frac{t-m}{T}) - \beta_j(\frac{T}{T})\right| \le \left|\frac{t-m-T}{T}\right| C$. Then we have

$$\begin{aligned}
&E\left[\left(\sum_{j=1}^{p}v_{ij,t}(\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T}))\right)\left(\sum_{k=1}^{p}v_{lk,t-m}(\beta_k(\frac{t-m}{T}) - \beta_k(\frac{T}{T}))\right)\right]\\
&\le \sum_{j=1}^{p}\sum_{k=1}^{p}|Ev_{ij,t}v_{lk,t-m}|\left|\left(\beta_j(\frac{t}{T}) - \beta_j(\frac{T}{T})\right)\left(\beta_k(\frac{t-m}{T}) - \beta_k(\frac{T}{T})\right)\right|\\
&\le \sum_{j=1}^{p}\sum_{k=1}^{p}|Ev_{ij,t}v_{lk,t-m}|\,C^2\left|\frac{t-T}{T}\right|\left|\frac{t-m-T}{T}\right|.
\end{aligned} \tag{41}$$

From Assumption 2, Theorem 17.9 in Davidson (1994, p.268) and the Lyapunov inequality (see 9.23 in Davidson (1994, p. 139)), we know that $\{vec(v_t)\}$ is $L_{2r/(r-1)}$-NED of size -2 on $\{V_t\}$, where $\{V_t\}$ is α-mixing of size $-2r/(r-2)$. So by the Lyapunov inequality, $L_{2r/(r-1)}$-NED of size -2 implies $L_{r/(r-1)}$-NED of size -2 when $r > 2$. Hence, $\{vec(v_t)\}$ is an $L_{r/(r-1)}$-NED process of size -2 on $\{V_t\}$. Also as stated in the proof of Lemma 4, it follows from Assumption 2 that

$\{v_{ij,t}\}$ is L_r-bounded uniformly in t for $r > 2$. So applying Lemma 15, we have $\{vec(v_t)\}$ is an $L_{r/(r-1)}$-mixingale of size -2. Combining with Theorem 17.5(i), Theorem 17.7(i) and equation (17.26) in Davidson (1994, p.267), there exists a sequence of non-negative constants $\{\zeta_m\}$, where $\zeta_m = O(m^{-2})$, such that

$$|Ev_{ij,t}v_{lk,t-m}| \;\leq\; \|v_{ij,t}\|_{\frac{1}{r-1}} \|v_{lk,t-m}\|_{\frac{r}{r-1}} \;=\; O(\zeta_m) \;=\; O(m^{-2}). \tag{42}$$

If follows from eq. (41) that

$$\begin{aligned}
\|vec(E(G_t G'_{t-m}))\| \leq & M^2 \left|\frac{t-T}{T}\right| \left|\frac{t-m-T}{T}\right| \left\{ \sum_{i=1}^{p}\sum_{l=1}^{p} \left[\left(\sum_{j=1}^{p}\sum_{k=1}^{p} |Ev_{ij,t}v_{lk,t-m}| \right) \right]^2 \right\}^{1/2} \\
\leq & M^2 \left|\frac{t-T}{T}\right| \left|\frac{t-m-T}{T}\right| \left\{ \sum_{i=1}^{p}\sum_{l=1}^{p} \left[\left(\sum_{j=1}^{p}\sum_{k=1}^{p} \|v_{ij,t}\|_r \|v_{lk,t-m}\|_r \zeta_m \right) \right]^2 \right\}^{1/2} \\
= & O\left(\left|\frac{t-T}{T}\right| \left|\frac{t-m-T}{T}\right| \frac{1}{m^2} \right)
\end{aligned}$$

where the last equality holds because p and M are both finite, $0 < p < \infty$, and $0 < M < \infty$. Then the rate of J_2 is given by

$$J_2 = \sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left\|vec(E(G_t G'_{t-m}))\right\| = O\left(\sum_{t=T-R+2}^{T-h} \sum_{m=1}^{t-(T-R+1)} \left|\frac{t-T}{T}\right| \left|\frac{t-m-T}{T}\right| \frac{1}{m^2} \right)$$

Using Lemma 18, we know $J_2 = O(R^3/T^2)$.

Finally, taking similar steps to find the rate of J_2 with Lemma 18 replaced by Lemma 17, it can be shown that $J_3 = O(R^3/T^2)$. By combining the results for J_1, J_2 and J_3, the rate of $\|vec(Var(S_G))\|$ is $O(R^3/T^2)$.

(b) By Chebyshev's inequality for vectors, we have for any real number $\epsilon > 0$

$$P\left[\left\| \frac{T}{R^{3/2}} S_G \right\| > \epsilon \right] \leq \frac{\|vec(Var(\frac{T}{R^{3/2}} S_G))\|}{\epsilon^2} = \frac{\frac{T^2}{R^3}\|vec(Var(S_G))\|}{\epsilon^2} = O(1),$$

where the last equality follows from the result in part (a). Therefore $R^{-1}S_G = O_p(\sqrt{R}/T)$.

(c) By Boole's inequality (Chung, 1974, p.20) and Chebyshev's inequality, it follows that for any

$\epsilon > 0$,

$$P\left(\sup_{R\in\Theta_R}\left\|\frac{T}{R^{\frac{1}{2}+\delta}}\left(\frac{1}{R}S_G\right)\right\| > \epsilon\right) \leq \sum_{R\in\Theta_R} P\left(\left\|\frac{T}{R^{\frac{1}{2}+\delta}}\left(\frac{1}{R}S_G\right)\right\| > \epsilon\right)$$

$$\leq \#\Theta_R \cdot \sup_{R\in\Theta_R} P\left(\left\|\frac{T}{R^{\frac{1}{2}+\delta}}\left(\frac{1}{R}S_G\right)\right\| > \epsilon\right) \leq \#\Theta_R \cdot \sup_{R\in\Theta_R} \frac{\|vec(Var(\frac{T}{R^{\frac{1}{2}+\delta}}\left(\frac{1}{R}S_G\right)))\|}{\epsilon^2}$$

$$\leq \#\Theta_R \cdot \sup_{R\in\Theta_R} \frac{\frac{T^2}{R^{3+2\delta}}\cdot\|vec(Var(S_G))\|}{\epsilon^2} \#\Theta_R \cdot \sup_{R\in\Theta_R} \frac{C}{R^{2\delta}\epsilon^2},$$

for some C. The last inequality follows from part (a). Hence

$$P\left(\sup_{R\in\Theta_R}\left\|\frac{T}{R^{\frac{1}{2}+\delta}}\left(\frac{1}{R}S_G\right)\right\| > \epsilon\right) \leq \#\Theta_R \cdot \frac{c}{\underline{R}^{2\delta}\epsilon^2} = O(1)$$

Therefore $\sup_{R\in\Theta_R} \frac{T}{R^{\frac{1}{2}+\delta}}\left\|\frac{1}{R}S_G\right\| = O_p(1)$, for some δ, $0 < \delta < 1/2$. The proof is complete. ∎

Proof of Lemma 9. Following the notation in Lemma 1 and the decomposition in eq. (24), we need to further show the rate of E_1 and E_2 uniformly in R, $R \in \Theta_R$. Hereafter denote $\sup_{R\in\Theta_R}$ by $\sup_R$ for simplicity. First we show the uniform rate of the first term E_1. Expanding E_1 gives

$$\sup_R \|R^{\frac{1}{2}-\delta}E_1\| = \sup_R \|R^{\frac{1}{2}-\delta}B_R H_R\| = \sup_R \|R^{\frac{1}{2}-\delta}(B_R^* H_R + (B_R - B_R^*)H_R)\|$$
$$\leq \left(\sup_R \|vec(B_R^*)\|\right)\left(\sup_R(R^{\frac{1}{2}-\delta}\|H_R\|)\right) + \left(\sup_R \|vec(B_R) - vec(B_R^*)\|\right)\left(\sup_R(R^{\frac{1}{2}-\delta})\|H_R\|\right) \quad (43)$$

It follows from Assumption 3 that $\sup_R \|vec(B_R^*)\| = O(1)$. Lemma 11 implies that $\sup_R(R^{\frac{1}{2}-\delta}\|H_R\|) = O_p(1)$. From Lemma 6(a), we know $\sup_{R\in\Theta_R}(R^{\frac{1}{2}-\delta})\,\|vec(B_R) - vec(B_R^*)\| = O_p(1)$. Then the rate of eq. (43) is given by

$$\sup_R \|R^{\frac{1}{2}-\delta}E_1\| \leq O(1)O_p(1) + \sup_R(R^{-\frac{1}{2}+\delta})O_p(1)O_p(1) = O_p(1).$$

Therefore the uniform rate of E_1 is given by $\sup_R \|R^{\frac{1}{2}-\delta}E_1\| = O_p(1)$. Next we need to find the uniform rate of the second term E_2. Expanding E_2 gives

$$\begin{aligned}
&\sup_R \left\| \frac{T}{R} E_2 \right\| = \sup_R \left\| \frac{T}{R} B_R Q_R \right\| \\
=& \sup_R \frac{T}{R} \| B_R^* Q_R^* + B_R^*(Q_R - Q_R^*) + (B_R - B_R^*) Q_R^* + (B_R - B_R^*)(Q_R - Q_R^*) \| \\
\leq& \left(\sup_R \| vec(B_R^*) \| \right) \left(\sup_R \| \frac{T}{R} Q_R^* \| \right) + \left(\sup_R \| vec(B_R^*) \| \right) \left(\sup_R \| \frac{T}{R} (Q_R - Q_R^*) \| \right) \\
&+ \left(\sup_R \| vec(B_R) - vec(B_R^*) \| \right) \left(\sup_R \| \frac{T}{R} Q_R^* \| \right) \\
&+ \left(\sup_R \| vec(B_R) - vec(B_R^*) \| \right) \left(\sup_R \| \frac{T}{R} (Q_R - Q_R^*) \| \right) \qquad (44)
\end{aligned}$$

Again it follows from Assumption 3 that $\sup_R \|vec(B_R^*)\| = O(1)$. From Lemma 12, we know that $\sup_R \|\frac{T}{R} Q_R^*\| = O(1)$. Lemma 8(c) implies that for some constant δ, $0 < \delta < 1/2$, $\sup_{R \in \Theta_R} \frac{T}{R^{\frac{1}{2}+\delta}} \|Q_R - Q_R^*\| = O_p(1)$. Again from Lemma 6(a), we know $\sup_{R \in \Theta_R} (R^{\frac{1}{2}-\delta}) \|vec(B_R) - vec(B_R^*)\| = O_p(1)$. Thus the rate of eq. (44) is given by

$$\begin{aligned}
\sup_R \left\| \frac{T}{R} E_2 \right\| \leq& O(1)O(1) + O(1) \sup_R (R^{-\frac{1}{2}+\delta}) O_p(1) \\
&+ \sup_R (R^{-\frac{1}{2}+\delta}) O_p(1) O(1) + \sup_R (R^{-\frac{1}{2}+\delta}) O_p(1) \sup_R (R^{-\frac{1}{2}+\delta}) O_p(1) = O(1).
\end{aligned}$$

Combining the uniform rates of E_1 and E_2 and Equation (24) yields $\sup_R \min(R^{\frac{1}{2}-\delta}, T/R) \|\hat{\beta}_R(1) - \beta(1)\| = O_p(1)$. Then using the Cauchy-Schwarz inequality, we get

$$\sup_{R \in \Theta_R} \min(R^{\frac{1}{2}-\delta}, T/R)(\hat{\beta}_R(1) - \beta(1))' x_T \leq \sup_{R \in \Theta_R} \min(R^{\frac{1}{2}-\delta}, T/R) \|\hat{\beta}_R(1) - \beta(1)\| \|x_T\| = O_p(1).$$

■

Proof of Lemma 10. Assumption 2(ii) implies that $\|vec(E(x_t x_t'))\|$ is bounded uniformly in t, thus $\|R^{-1} \sum_{t=T-R+1}^{T-h} vec(E(x_t x_t'))\|$ is $O(1)$ uniformly for all R. ■

Proof of Lemma 11. Following the notation in Lemma 4, we have $\sum_{t=T-R+1}^{T-h} x_t u_{t+h} \equiv S_U$, when the parameter k in Lemma 4 is set to zero. Then Lemma 4(a) implies that $\|vec(Var(S_U))\| \simeq R$.

By Boole's inequality and Chebyshev's inequality, for any $\epsilon > 0$,

$$P\left(\sup_{R\in\Theta_R}\left\|\frac{1}{R^{\frac{1}{2}+\delta}}S_U\right\| > \epsilon\right) \le \sum_{R\in\Theta_R} P\left(\left\|\frac{1}{R^{\frac{1}{2}+\delta}}S_U\right\| > \epsilon\right) \le \#\Theta_R \cdot \sup_{R\in\Theta_R} P\left(\left\|\frac{1}{R^{\frac{1}{2}+\delta}}S_U\right\| > \epsilon\right)$$
$$\le \#\Theta_R \cdot \sup_{R\in\Theta_R}\frac{\|vec(Var(\frac{1}{R^{\frac{1}{2}+\delta}}S_U))\|}{\epsilon^2} \simeq \#\Theta_R \cdot \sup_{R\in\Theta_R}\frac{C}{R^{2\delta}\epsilon^2},$$

for some C. Hence

$$P\left(\sup_{R\in\Theta_R}\left\|\frac{1}{R^{\frac{1}{2}+\delta}}S_U\right\| > \epsilon\right) \le \#\Theta_R \cdot \frac{c}{\underline{R}^{2\delta}\epsilon^2} = O(1).$$

Therefore $\sup_{R\in\Theta_R}\|R^{-\frac{1}{2}-\delta}\sum_{t=T-R+1}^{T-h}x_t u_{t+h}\| = O_p(1)$, where $0 < \delta < 1/2$. ■

Proof of Lemma 12. By the Cauchy-Schwarz inequality, we have

$$\sup_{R\in\Theta_R}\left\|\frac{T}{R^2}\sum_{t=T-R+1}^{T-h}E(x_t x_t')\left(\beta(\frac{t}{T})-\beta(\frac{T}{T})\right)\right\| \le \sup_{R\in\Theta_R}\left(\frac{T}{R^2}\sum_{t=T-R+1}^{T-h}\|vec(E(x_t x_t'))\|\|\beta(\frac{t}{T})-\beta(\frac{T}{T})\|\right).$$

Assumption 2(ii) implies that $\|vec(E(x_t x_t'))\|$ is bounded uniformly in t. Next Assumption 4(i) implies that $\|\beta(\frac{t}{T})-\beta(\frac{T}{T})\| \le C\left(\frac{T-t}{T}\right)$ for some C. Then it follows that

$$\sup_{R\in\Theta_R}\left(\frac{T}{R^2}\sum_{t=T-R+1}^{T-h}\|vec(E(x_t x_t'))\|\|\beta(\frac{t}{T})-\beta(\frac{T}{T})\|\right) \le \sup_{R\in\Theta_R}\left(\frac{CT}{R^2}\sum_{t=T-R+1}^{T-h}\left(\frac{T-t}{T}\right)\right) = O(1),$$

for some C, where the last equality follows from eq. (30) and Lemma 16. ■

Proof of Lemma 13. Define matrix $S_{R_0}(1)$ and vector $T_{R_0}(1)$ as

$$\begin{bmatrix}\tilde{\beta}(1)\\ \tilde{\beta}^{(1)}(1)\end{bmatrix} = \begin{bmatrix}\frac{1}{R_0}\sum x_s x_s' & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)\\ \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right) & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^2\end{bmatrix}^{-1}\begin{bmatrix}\frac{1}{R_0}\sum x_s y_{s+h}\\ \frac{1}{R_0}\sum x_s y_{s+h}\left(\frac{s-T}{T}\right)\end{bmatrix}$$
$$\equiv [S_{R_0}(1)]^{-1}T_{R_0}(1) \tag{45}$$

where the summation $\sum$ represents $\sum_{s=T-R_0+1}^{T-h}$. R_0 is the window size used for the local linear regression. Note that both $S_{R_0}(1)$ and $T_{R_0}(1)$ depend on R_0.

Applying Taylor's theorem to $\beta(\frac{s}{T})$ yields

$$\beta\left(\frac{s}{T}\right) = \beta(1) + \beta^{(1)}(1)\left(\frac{s-T}{T}\right) + \frac{\beta^{(2)}(c)}{2!}\left(\frac{s-T}{T}\right)^2 \tag{46}$$

where $c = \lambda \frac{s}{T} + (1-\lambda)\frac{T}{T}$, for $\lambda \in (0,1)$. By substituting eq. (46) into y_{s+h}, we can write

$$y_{s+h} = x_s'\beta\left(\frac{s}{T}\right) + u_{s+h} = x_s'\beta(1) + x_s'\left(\frac{s-T}{T}\right)\beta^{(1)}(1) + x_s'\left(\frac{s-T}{T}\right)^2\frac{\beta^{(2)}(c)}{2!} + u_{s+h}.$$

Then $T_{R_0}(1)$ can be expanded as follows:

$$\begin{aligned} T_{R_0}(1) = \begin{bmatrix} \frac{1}{R_0}\sum x_s y_{s+h} \\ \frac{1}{R_0}\sum x_s y_{s+h}\left(\frac{s-T}{T}\right) \end{bmatrix} &= \underbrace{\begin{bmatrix} \frac{1}{R_0}\sum x_s x_s' & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right) \\ \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right) & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^2 \end{bmatrix}}_{=S_{R_0}(1)} \begin{bmatrix} \beta(1) \\ \beta^{(1)}(1) \end{bmatrix} \\ &+ \begin{bmatrix} \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^2 \\ \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^3 \end{bmatrix} \frac{\beta^{(2)}(c)}{2!} + \begin{bmatrix} \frac{1}{R_0}\sum x_s u_{s+h} \\ \frac{1}{R_0}\sum x_s\left(\frac{s-T}{T}\right)u_{s+h} \end{bmatrix}. \end{aligned} \tag{47}$$

By substitiuting eq. (47) into eq. (45), we obtain

$$\begin{bmatrix} \tilde{\beta}(1) - \beta(1) \\ \tilde{\beta}^{(1)}(1) - \beta^{(1)}(1) \end{bmatrix} = [S_{R_0}(1)]^{-1} \begin{bmatrix} \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^2 \\ \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^3 \end{bmatrix} \frac{\beta^{(2)}(c)}{2!} + [S_{R_0}(1)]^{-1} \begin{bmatrix} \frac{1}{R_0}\sum x_s u_{s+h} \\ \frac{1}{R_0}\sum x_s\left(\frac{s-T}{T}\right)u_{s+h} \end{bmatrix} \tag{48}$$

$S_{R_0}(1)$ can be expanded as follows:

$$\begin{aligned} S_{R_0}(1) &= \begin{bmatrix} \frac{1}{R_0}\sum x_s x_s' & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right) \\ \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right) & \frac{1}{R_0}\sum x_s x_s'\left(\frac{s-T}{T}\right)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{R_0}\sum E(x_s x_s') & \frac{1}{R_0}\sum E(x_s x_s')\left(\frac{s-T}{T}\right) \\ \frac{1}{R_0}\sum E(x_s x_s')\left(\frac{s-T}{T}\right) & \frac{1}{R_0}\sum E(x_s x_s')\left(\frac{s-T}{T}\right)^2 \end{bmatrix} \\ &\quad + \begin{bmatrix} \frac{1}{R_0}\sum (x_s x_s' - E(x_s x_s')) & \frac{1}{R_0}\sum (x_s x_s' - E(x_s x_s'))\left(\frac{s-T}{T}\right) \\ \frac{1}{R_0}\sum (x_s x_s' - E(x_s x_s'))\left(\frac{s-T}{T}\right) & \frac{1}{R_0}\sum (x_s x_s' - E(x_s x_s'))\left(\frac{s-T}{T}\right)^2 \end{bmatrix} \\ &= K_1 + K_2 \end{aligned}$$

It follows from Assumption 3, Lemma 3 and Lemma 5 that

$$K_1 = \begin{bmatrix} O(1) & O\left(\frac{R_0}{T}\right) \\ O\left(\frac{R_0}{T}\right) & O\left(\frac{R_0^2}{T^2}\right) \end{bmatrix}, \quad K_2 = \begin{bmatrix} O_p\left(\frac{1}{\sqrt{R_0}}\right) & O_p\left(\frac{\sqrt{R_0}}{T}\right) \\ O_p\left(\frac{\sqrt{R_0}}{T}\right) & O_p\left(\frac{R_0^{3/2}}{T^2}\right) \end{bmatrix}.$$

Therefore, the rate of $S_{R_0}(1)$ is determinted by K_1. By using the inverse of partitioned matrices in Abadir and Magnus (2005, p106) and Lemma 19, we obtain

$$[S_{R_0}(1)]^{-1} = \begin{bmatrix} O(1) & O\left(\frac{T}{R_0}\right) \\ O\left(\frac{T}{R_0}\right) & O\left(\frac{T^2}{R_0^2}\right) \end{bmatrix}$$

Then the rate of eq. (48) is

$$\begin{aligned} \begin{bmatrix} \tilde{\beta}(1) - \beta(1) \\ \tilde{\beta}^{(1)}(1) - \beta^{(1)}(1) \end{bmatrix} &= [S_{R_0}(1)]^{-1} \begin{bmatrix} \frac{1}{R_0} \sum E(x_s x_s') \left(\frac{s-T}{T}\right)^2 \\ \frac{1}{R_0} \sum E(x_s x_s') \left(\frac{s-T}{T}\right)^3 \end{bmatrix} \frac{\beta^{(2)}(c)}{2!} \\ &+ [S_{R_0}(1)]^{-1} \begin{bmatrix} \frac{1}{R_0} \sum (x_s x_s' - E(x_s x_s')) \left(\frac{s-T}{T}\right)^2 \\ \frac{1}{R_0} \sum (x_s x_s' - E(x_s x_s')) \left(\frac{s-T}{T}\right)^3 \end{bmatrix} \frac{\beta^{(2)}(c)}{2!} \\ &+ [S_{R_0}(1)]^{-1} \begin{bmatrix} \frac{1}{R_0} \sum x_s u_{s+h} \\ \frac{1}{R_0} \sum x_s \left(\frac{s-T}{T}\right) u_{s+h} \end{bmatrix} = L_1 + L_2 + L_3 \end{aligned}$$

It follows from Lemma 3, Assumption 4(ii), Lemma 5 and Lemma 4 that

$$L_1 = \begin{bmatrix} O\left(\frac{R_0^2}{T^2}\right) \\ O\left(\frac{R_0}{T}\right) \end{bmatrix}, \quad L_2 = \begin{bmatrix} O_p\left(\frac{R_0^{3/2}}{T^2}\right) \\ O_p\left(\frac{\sqrt{R_0}}{T}\right) \end{bmatrix}, \quad L_3 = \begin{bmatrix} O_p\left(\frac{1}{\sqrt{R_0}}\right) \\ O_p\left(\frac{T}{R_0^{3/2}}\right) \end{bmatrix}.$$

Therefore we obtain the convergence rates of $\tilde{\beta}(1)$ and $\tilde{\beta}^{(1)}(1)$ as

$$\begin{bmatrix} \tilde{\beta}(1) - \beta(1) \\ \tilde{\beta}^{(1)}(1) - \beta^{(1)}(1) \end{bmatrix} = \begin{bmatrix} O_p\left(\frac{R_0^2}{T^2}\right) + O_p\left(\frac{1}{\sqrt{R_0}}\right) \\ O_p\left(\frac{R_0}{T}\right) + O_p\left(\frac{T}{R_0^{3/2}}\right) \end{bmatrix}.$$

■

Proof of Lemma 14. If $R_0^{1/2} \leq T^2/R_0^2$,

$$\frac{\min(R^{\frac{1}{2}-\delta}, T/R)}{\min(R_0^{\frac{1}{2}}, T^2/R_0^2)} = \frac{\min(R^{\frac{1}{2}-\delta}, T/R)}{R_0^{\frac{1}{2}}} \leq \frac{R^{\frac{1}{2}-\delta}}{R_0^{\frac{1}{2}}} = o(1), \tag{49}$$

where the inequality follows because $\min(A, B) \leq A$ and the last equality follows since $R_0 >> R$ by Assumption 5 and $0 < \delta < 1/2$. If $R_0^{1/2} \geq T^2/R_0^2$,

$$\frac{\min(R^{\frac{1}{2}-\delta}, T/R)}{\min(R_0^{\frac{1}{2}}, T^2/R_0^2)} = \frac{\min(R^{\frac{1}{2}-\delta}, T/R)}{T^2/R_0^2} \leq \frac{T/R}{T^2/R_0^2} = o(1) \tag{50}$$

where the last equality follows because $R_0^2 R/T = o(1)$ in Assumption 5. ■

Proof of Lemma 15. Given that $\{U_t\}$ is an $L_{r/(r-1)}$-NED process of size -2 on $\{V_t\}$ with constants $\{d_t\}$, where $\{V_t\}$ is α-mixing of size $-2r/(r-2)$, then it follows from Theorem 17.5 in Davidson (1994, p.264) that $\{U_t, \mathcal{F}_{-\infty}^t\}$ is an $L_{r/(r-1)}$ mixingale of size $-\min\{2, (2r/(r-2))((r-1)/r - 1/r)\} = -2$ with constants $c_t = O(max\{\|U_t\|_r, d_t\})$. ■

Proof of Lemma 16. Using Theorem 2.27 in Davidson (1994, p.32), we know that $\sum_{j=1}^R j^k \simeq R^{k+1}$ when $k > -1$. Then $B_k \simeq \sum_{j=1}^R j^k/R^{k+1} \simeq \sum_{j=h}^R j^k/R^{k+1} \simeq C$, for some C, $0 < c < \infty$. ■

Proof of Lemma 17. Notice that

$$\begin{aligned}
&\sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} \left[\frac{1}{m^2}\left(\frac{t-T}{T}\right)^k \left(\frac{t+m-T}{T}\right)^k\right] \leq \sum_{t=T-R+1}^{T-h-1} \sum_{m=1}^{T-h-t} \left[\frac{1}{m^2}\left(\frac{T-t}{T}\right)^{2k}\right] \\
&= \sum_{t=T-R+1}^{T-h-1} \left(\frac{T-t}{T}\right)^{2k} + \sum_{t=T-R+1}^{T-h-2} \sum_{m=2}^{T-h-t} \left[\frac{1}{m^2}\left(\frac{T-t}{T}\right)^{2k}\right] = H_1 + H_2.
\end{aligned}$$

$H_1 \simeq R^{2k+1}/T^{2k}$ by eq. (32). H_2 follows

$$\begin{aligned}
H_2 &\leq \sum_{t=T-R+1}^{T-h-2} \left[\left(\frac{T-t}{T}\right)^{2k} \sum_{m=2}^{T-h-t} \int_{m-1}^{m} \frac{1}{x^2} \mathrm{d}x\right] = \sum_{t=T-R+1}^{T-h-2} \left[\left(\frac{T-t}{T}\right)^{2k} \int_{1}^{T-h-t} \frac{1}{x^2} \mathrm{d}x\right] \\
&= \sum_{t=T-R+1}^{T-h-2} \left[\left(\frac{T-t}{T}\right)^{2k} \left(1 - \frac{1}{T-h-t}\right)\right] \leq \sum_{t=T-R+1}^{T-h-2} \left(\frac{T-t}{T}\right)^{2k} = O\left(\frac{R^{2k+1}}{T^{2k}}\right).
\end{aligned} \tag{51}$$

The first inequality in eq. (51) holds because for $m = 2, 3, ..., R$, $R \in \mathbb{Z}^+$, we always have

$$\frac{1}{m^2} \leq \frac{1}{m(m-1)} = \int_{m-1}^{m} \frac{1}{x^2} \mathrm{d}x \tag{52}$$

Combining the rates of H_1 and H_2 gives the result. $\blacksquare$

Proof of Lemma 18. The expression is bounded by:

$$\left|\sum_{t=T-R+2}^{T-h}\sum_{m=1}^{t-(T-R+1)}\left[\frac{1}{m^2}\left(\frac{t-T}{T}\right)^k\left(\frac{t-m-T}{T}\right)^k\right]\right| \leq \sum_{t=T-R+2}^{T-h}\sum_{m=1}^{t-(T-R+1)}\left[\frac{1}{m^2}\left(\frac{T-t}{T}\right)^k\left(\frac{R}{T}\right)^k\right]$$
$$=\left(\frac{R}{T}\right)^k\sum_{t=T-R+2}^{T-h}\left(\frac{T-t}{T}\right)^k+\left(\frac{R}{T}\right)^k\sum_{t=T-R+2}^{T-h}\left[\left(\frac{T-t}{T}\right)^k\sum_{m=2}^{t-(T-R+1)}\frac{1}{m^2}\right]=I_1+I_2.$$

First, we want to show that $I_1 \simeq R^{2k+1}/T^{2k}$. According to Lemma 16, we have

$$\sum_{t=T-R+2}^{T-h}\left(\frac{T-t}{T}\right)^k=\sum_{j=h}^{R-2}\frac{j^k}{T^k}\simeq\sum_{j=h}^{R}\frac{j^k}{T^k}\simeq\frac{R^{k+1}}{T^k}. \tag{53}$$

Hence it follows that $I_1 = CR^{2k+1}/T^{2k}$ for some C.

Next, we want to show the rate of I_2. It follows from eq. (52) that

$$\sum_{m=2}^{t-(T-R+1)}\frac{1}{m^2}\leq\sum_{m=2}^{t-(T-R+1)}\int_{m-1}^{m}\frac{1}{x^2}\mathrm{d}x=\int_{1}^{t-(T-R+1)}\frac{1}{x^2}\mathrm{d}x=1-\frac{1}{t-(T-R+1)}. \tag{54}$$

Then the rate of I_2 is given by

$$I_2\leq\left(\frac{R}{T}\right)^k\sum_{t=T-R+2}^{T-h}\left[\left(\frac{T-t}{T}\right)^k\left(1-\frac{1}{t-(T-R+1)}\right)\right]\leq\left(\frac{R}{T}\right)^k\sum_{t=T-R+2}^{T-h}\left(\frac{T-t}{T}\right)^k.$$

By using arguments similar to those used in eq. (53), $I_2 \simeq R^{2k+1}/T^{2k}$, we obtain $\sum_{t=T-R+2}^{T-h}\sum_{m=1}^{t-(T-R+1)} m^{-2}(t-T)^k(t-m-T)^k/T^{2k} = O(R^{2k+1}/T^{2k})$. $\blacksquare$

Proof of Lemma 19. Suppose that $(A+B)^{-1} = A^{-1}+X$. Then

$$\begin{aligned}(A^{-1}+X)(A+B) =& I\\ X(A+B) =& -A^{-1}B\\ X =& -A^{-1}B(A+B)^{-1} = -A^{-1}B(A^{-1}+X) = -(I+A^{-1}B)^{-1}A^{-1}BA^{-1}\end{aligned}$$

Hence $(A+B)^{-1} = A^{-1}-(I+A^{-1}B)^{-1}A^{-1}BA^{-1}$. $\blacksquare$

References

Abadir, M.K., and J.R. Magnus, 2005, Matrix algebra, Cambridge University Press, Cambridge.

Anatolyev, S. and V. Kitov, 2007, Using all observations when forecasting under structural breaks. Finnish Economic Papers 20, 166-176.

Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Bai, J. and P. Perron, 1998, Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

Bacchetta, P., E. van Wincoop and T. Beutler, 2010, Can parameter instability explain the Meese-Rogoff puzzle? in: L. Reichlin and K. West, (Eds.), NBER international seminar on macroeconomics, University of Chicago Press, Chicago.

Cai, Z., 2007, Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Carriero, A., G. Kapetanios and M. Marcellino, 2009, Forecasting exchange rates with a large Bayesian VAR. International Journal of Forecasting 25, 400–417.

Chen, B. and Y. Hong, 2012, Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80, 1157–1183.

Cheung, Y.W., M.D. Chinn and A.G. Pascual, 2005, Empirical exchange rate models of the nineties: Are Any Fit to Survive? Journal of International Money and Finance 24, 1150–1175.

Chung, K.L., 1974, A course in probability theory, second edition. Academic Press, San Diego.

Clark, T.E. and M.W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85-110.

Clements, M. P. and D. F. Hendry, 1998, Intercept corrections and structural change. Journal of Applied Econometrics 11, 475–495.

Davidson, J., 1994, Stochastic limit theory: An introduction for econometricicans. Oxford University Press, Oxford.

Della Corte, P., L. Sarno and G. Sestieri, 2012, The Predictive information content of external imbalances for exchange rate returns: How much is it worth? Review of Economics and Statistics 94, 100–115.

Fan, J. and I. Gijbels, 1996, Local polynomial modelling and its applications. Chapman and Hall, London.

Faust, J., J.H. Rogers and J.H. Wright, 2003, Exchange rate forecasting: The errors we've really made. Journal of International Economics 60, 35–59.

Giacomini, R. and B. Rossi, 2009, Detecting and predicting forecast breakdowns. Review of Economic Studies 76(2, 669-705.

Giraitis, L., G. Kapetanios and S. Price, 2013, Adaptive forecasting in the presence of recent and ongoing structural change. Journal of Econometrics 177, 153–170.

Goyal, A. and I. Welch, 2003, Predicting the equity premium with dividend ratios. Management Science 49, 639-654.

Härdle, W. and J.S. Marron, 1985, Optimal bandwidth selection in nonparametric regression function estimation. Annals of Statistics 13, 1465–1481.

Inoue, A. and B. Rossi, 2012, Out-of-sample forecast tests robust to the choice of window size. Journal of Business and Economic Statistics 30, 432–453.

Koop, G. and S.M. Potter, 2004, Forecasting in large macroeconomic panels using Bayesian model averaging. Econometrics Journal 7, 550-565.

Laurent, S., J.V.K. Rombouts and F. Violante, 2012, On the forecasting accuracy of multivariate GARCH models. Journal of Applied Econometrics, 27, 934–955.

Lu, Z. and O. Linton, 2007, Local linear fitting under near epoch dependence. Econometric Theory 23, 37–70.

Marron, J.S., 1985, An asymptotically efficient solution to the bandwidth problem of kernel density estimation. Annals of Statistics 14, 1011–1023.

Marron, J.S. and W. Härdle, 1986, Random approximations to some measures of accuracy in nonparametric curve estimation. Journal of Multivariate Analysis 20, 91-113.

Meese, R. and K.S. Rogoff, 1983a, Exchange rate models of the seventies. Do they fit out of sample? Journal of International Economics 14, 3–24.

Meese, R.A. and K.S. Rogoff, 1983b, The out-of-sample failure of empirical exchange rate models: Sampling error or mis-specification? in: Jacob Frenkel (Eds.), Exchange rates and international macroeconomics, NBER and University of Chicago Press, Chicago.

Molodtsova T., and D.H. Papell, 2009, Out-of-sample exchange rate predictability with Taylor rule fundamentals. Journal of International Economics 77, 167–180.

Molodtsova, T., and D.H. Papell, 2012, Taylor rule exchange rate forecasting during the financial crisis. NBER international seminar on macroeconomics 9, 55–97.

Nyblom, J., 1989, Testing for the constance of parameters over time. Journal of the American Statistical Association 84, 223-230.

Patton, A.J., 2015, Evaluating and comparing possibly misspeficied models. unpublished manuscript, Duke University.

Paye, B. and A. Timmermann, 2006, Instability of return prediction models. Journal of Empirical Finance 13, 274–315

Pesaran, M.H., and A. Pick, 2011, Forecast combination across estimation windows. Journal of Business and Economic Statistics 29, 307–318.

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.

Robinson, P.M., 1989, Nonparamatric estimation of time-varying parameters. Hackl, P., Ed., Statistical analysis and forecasting of economic structural change, Springer, Berlin, 253-264.

Rossi, B., 2013, Advances in forecasting under model instability. in: G. Elliott and A. Timmermann, (Eds.), Handbook of economic forecasting volume 2B, Elsevier Publications, Amsterdam, 1203–1324.

Schinasi, G. and P. Swamy, 1989, The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. Journal of International Money and Finance 8, 375-390.

Stock, J.H. and M.W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11–30.

Stock, J.H. and M.W. Watson, 1999, Forecasting inflation. Journal of Monetary Economics 44, 293–335.

Stock, J.H. and M.W. Watson, 2003, Forecasting output and inflation: The role of asset prices. Journal of Economic Literature 41, 788–829.

Stock, J.H. and M.W. Watson, 2007, Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking 39, 3–34.

Swanson, N.R., 1998, Money and output viewed through a rolling window. Journal of Monetary Economics 41, 455–474.

Welch, I. and A. Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

West, K.D., 1996, Asymptotic inference about predictive ability. Econometrica 64, 1067-1084.

Wolff, C., 1987, Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models. Journal of Business and Economic Statistics 5, 87-97.

Table 1: Description of DGPs

$$\begin{bmatrix} y_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} \beta_t \\ 0 \end{bmatrix} + \begin{bmatrix} a_t & b_t \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} + \begin{bmatrix} u_{y,t+1} \\ u_{x,t+1} \end{bmatrix},$$

$$\text{where } \begin{bmatrix} u_{y,t+1} \\ u_{x,t+1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

DGP	β_t	σ^2	a_t	b_t	Comments
1	0	1	0	0	Constant parameter
2	$I(t \geq 0.25T + 1)$	$\frac{3}{16}$	0	0	Break at date $0.25T$
3	$I(t \geq 0.5T + 1)$	$\frac{1}{4}$	0	0	Break at date $0.5T$
4	$I(t \geq 0.75T + 1)$	$\frac{3}{16}$	0	0	Break at date $0.75T$
5	t/T	$\frac{1}{12}$	0	0	Linearly time-varying parameter
6	$(t/T)^2$	$\frac{9}{100}$	0	0	Quadratically time-varying parameter
7	$\beta_t = \beta_{t-1} + \sqrt{\frac{2}{T}}\epsilon_t$	1	0	0	
	$\epsilon_t \sim N(0,1)$, $\beta_0 = 0$				β_t follows a random walk
8	0	1	0.9	1	Constant parameters
9	0	1	$0.9 - 0.4I(t \geq 0.25T + 1)$	1	Break in a_t at date $0.25T$
10	0	1	$0.9 - 0.4I(t \geq 0.5T + 1)$	1	Break in a_t at date $0.5T$
11	0	1	$0.9 - 0.4I(t \geq 0.75T + 1)$	1	Break in a_t at date $0.75T$
12	0	1	$0.9 - 0.4I(t \geq 0.95T + 1)$	1	Break in a_t at date $0.95T$
13	0	1	0.9	$1 + I(t \geq 0.25T + 1)$	Break in b_t at date $0.25T$
14	0	1	0.9	$1 + I(t \geq 0.5T + 1)$	Break in b_t at date $0.5T$
15	0	1	0.9	$1 + I(t \geq 0.75T + 1)$	Break in b_t at date $0.75T$
16	0	1	0.9	$1 + I(t \geq 0.95T + 1)$	Break in b_t at date $0.95T$
17	0	1	$0.9 - 0.4(t/T)$	1	Linearly time-varying a_t
18	0	1	0.9	$1 + (t/T)$	Linearly time-varying b_t
19	0	1	$0.9 - 0.4(t/T)^2$	1	Quardratically time-varying a_t
20	0	1	0.9	$1 + (t/T)^2$	Quardratically time-varying b_t
21	0	1	$a_t = a_{t-1} + \frac{0.1}{\sqrt{T}}\epsilon_t$	1	a_t follows a random walk
			$\epsilon_t \sim N(0,1)$, $a_0 = 0.9$		
22	0	1	0.9	$b_t = b_{t-1} + \frac{1}{\sqrt{T}}\epsilon_t$	b_t follows a random walk
				$\epsilon_t \sim N(0,1)$, $b_0 = 1$	

Table 2: Root MSFE (T=100, h=1)

	Estimated break date ($\hat{T}_1$)						Unknown break date			Cai's methods							New methods			
DGP	Postbk	CV	WA	Pooled	Troff	WLS	CV	WA	Pooled	Cai1	Cai2	LL1	LL2	LQ1	LQ2	PPP	OptR1	OptR2	OptR3	True
1	1.003	1.001	**1.000**	1.000	1.003	1.003	1.007	1.002	1.004	1.011	1.006	1.066	1.050	1.131	1.124	1.005	1.002	1.002	1.002	1.000
2	0.874	0.880	0.916	0.911	0.874	**0.874**	0.883	0.881	0.878	0.904	0.901	1.010	0.999	1.240	1.215	0.879	0.879	0.879	0.879	0.867
3	**0.718**	0.722	0.859	0.831	0.718	**0.718**	0.723	0.823	0.755	0.740	0.737	0.819	0.806	1.002	0.965	0.749	0.722	0.722	0.722	0.711
4	0.515	0.625	0.875	0.799	0.515	**0.514**	0.625	0.875	0.727	0.526	0.527	0.587	0.583	0.719	0.702	0.692	0.596	0.596	0.596	0.505
5	0.663	0.655	0.8500	0.822	0.663	0.655	0.599	0.813	0.698	0.706	0.740	0.543	**0.535**	0.576	0.572	0.671	0.588	0.588	0.588	0.522
6	0.603	0.621	0.860	0.818	0.603	0.592	0.602	0.850	0.719	0.598	0.667	0.458	**0.450**	0.500	0.484	0.681	0.586	0.586	0.586	0.447
7	0.878	0.883	0.948	0.933	0.878	0.877	0.873	0.938	0.894	**0.858**	0.867	0.897	0.884	1.027	1.003	0.879	0.881	0.881	0.881	0.816
8	1.007	1.003	**1.001**	**1.001**	1.007	1.419	1.018	1.011	1.019	1.030	1.022	1.213	1.191	1.619	1.496		1.008	1.008	1.008	0.999
9	0.806	**0.797**	0.939	0.925	0.808	1.236	0.804	0.831	0.806	0.844	0.825	1.168	1.079	3.396	2.6687		0.821	0.822	0.822	0.772
10	0.715	**0.697**	0.925	0.904	0.722	1.087	0.701	0.872	0.763	0.753	0.750	1.087	0.975	3.114	2.123		0.723	0.723	0.723	0.668
11	**0.708**	0.909	0.950	0.912	0.747	1.060	0.909	0.949	0.839	0.790	0.827	0.990	0.921	5.342	2.602		0.832	0.831	0.831	0.614
12	1.059	0.946	0.964	0.936	1.029	1.411	0.937	0.957	**0.903**	1.064	1.015	1.777	1.591	5.114	3.536		0.917	0.917	0.917	0.639
13	**0.881**	0.885	0.938	0.928	0.883	1.636	0.893	0.892	0.891	0.938	0.914	1.294	1.155	2.884	2.015		0.917	0.919	0.919	0.857
14	**0.731**	0.734	0.889	0.860	0.741	1.246	0.738	0.849	0.773	0.803	0.792	1.112	0.998	2.653	2.344		0.758	0.760	0.760	0.696
15	**0.650**	0.785	0.894	0.831	0.673	1.040	0.784	0.894	0.772	0.752	0.791	0.919	0.848	2.842	1.541		0.716	0.716	0.716	0.584
16	0.890	0.903	0.945	0.907	0.892	1.240	0.887	0.932	0.851	0.895	0.911	0.898	0.873	1.409	1.815		0.845	**0.845**	**0.845**	0.577
17	0.849	0.848	0.927	0.916	0.851	1.242	0.830	0.892	0.845	0.874	0.887	0.882	0.860	1.102	1.027		**0.824**	0.825	0.825	0.768
18	0.847	0.846	0.914	0.901	0.848	1.301	0.830	0.884	0.839	0.872	0.882	0.896	0.868	1.135	1.047		**0.825**	0.826	0.826	0.761
19	0.798	0.809	0.921	0.902	0.801	1.177	0.792	0.901	0.822	0.855	0.885	0.790	0.786	1.003	0.920		**0.763**	0.764	0.764	0.676
20	0.776	0.789	0.900	0.874	0.781	1.195	0.775	0.881	0.802	0.851	0.870	0.787	0.757	1.129	0.920		**0.756**	0.756	0.756	0.663
21	1.039	1.030	**0.999**	1.000	1.000	4.594	1.030	0.999	1.005	1.312	1.126	1.034	1.475	1.945	1.261		1.038	1.038	1.038	0.790
22	0.823	0.835	0.928	0.902	0.828	1.245	0.826	0.912	0.839	0.865	0.873	1.085	0.993	3.086	1.652		**0.822**	0.823	0.823	0.709

Notes: Postbk: Pesaran and Timmermann's (2007) post-break method; CV: PT's cross validation; WA: PT's weighted average of forecasts; Pooled: PT's pooled forecast combination; Troff: PT's trade-off method; WLS: Anatolyev and Kitov's (2007); Cai1: Cai's (2007) AIC and the rolling OLS estimator; Cai2: Cai's (2007) AIC and local constant regressins on the Epanechnikov kernel; LL1: Cai's (2007) AIC and local linear regressions with the uniform kernel; LL2: Cai's (2007) AIC and local linear regressions with the Epanechnikov kernel; LQ1: Cai's (2007) AIC and local quadratic regressions with the uniform kernel; LQ2: Cai's (2007) AIC and local quadratic regressions with the Epanechnikov kernel; PPP: Pesaran, Pick and Pranovich's (2013) robust optimal weights (eq. 48 in their paper) that integrate the break date over the entire sample; OptR1: R0=CV(unknown break date), $\underline{R} = \max(1.5T^{2/3}, 20)$, $\overline{R} = \min(4T^{2/3}, T-h)$; OptR2: R0=CV(unknown break date), $\underline{R} = \max(1.5T^{2/3}, 20)$, $\overline{R} = \min(5T^{2/3}, T-h)$; OptR3: R0=CV(unknown break date), $\underline{R} = \max(1.5T^{2/3}, 20)$, $\overline{R} = \min(6T^{2/3}, T-h)$; True: the infeasible MSFE criterion. The estimated break date $\hat{T}_1$ is obtained using Bai and Perron (1998) with $[0.15T, 0.85T]$ trimming range for possible break dates at the 5% significance level.

Table 3: Root MSFE (T=200, h=1)

	Estimated break date ($\hat{T}_1$)						Unknown break date			Cai's methods							New method			
DGP	Postbk	CV	WA	Pooled	Troff	WLS	CV	WA	Pooled	Cai1	Cai2	LL1	LL2	LQ1	LQ2	PPP	OptR1	OptR2	OptR3	True
1	1.001	1.000	**1.000**	**1.000**	1.001	1.001	1.003	1.001	1.001	1.006	1.004	1.032	1.024	1.070	1.066	1.002	1.000	1.000	1.000	1.000
2	0.866	0.869	0.914	0.908	0.866	**0.866**	0.871	0.875	0.870	0.880	0.880	0.944	0.943	1.058	1.060	0.870	0.870	0.870	0.870	0.864
3	0.709	0.711	0.857	0.830	0.709	**0.709**	0.711	0.819	0.749	0.719	0.718	0.768	0.767	0.858	0.854	0.741	0.711	0.711	0.711	0.706
4	0.504	0.611	0.872	0.798	0.504	**0.504**	0.611	0.872	0.723	0.508	0.509	0.546	0.545	0.608	0.612	0.687	0.508	0.508	0.508	0.500
5	0.658	0.653	0.847	0.821	0.658	0.654	0.590	0.810	0.694	0.694	0.731	0.518	**0.514**	0.537	0.535	0.665	0.553	0.553	0.553	0.511
6	0.597	0.614	0.857	0.817	0.597	0.592	0.595	0.848	0.716	0.524	0.587	0.434	**0.430**	0.461	0.452	0.675	0.529	0.529	0.529	0.436
7	0.862	0.874	0.946	0.930	0.862	0.861	0.868	0.938	0.893	**0.836**	0.839	0.853	0.853	0.926	0.924	0.878	0.859	0.859	0.859	0.805
8	1.005	1.008	**1.000**	**1.000**	1.004	1.418	1.005	1.001	1.005	1.009	1.006	1.082	1.067	1.188	1.162		1.003	1.003	1.003	1.000
9	0.750	**0.744**	0.924	0.905	0.751	1.140	0.748	0.791	0.755	0.780	0.772	0.965	0.921	1.420	1.314		0.754	0.755	0.755	0.738
10	0.670	**0.663**	0.914	0.891	0.672	1.006	0.664	0.859	0.736	0.693	0.690	0.847	0.813	1.274	1.168		0.677	0.678	0.678	0.654
11	**0.625**	0.833	0.936	0.893	0.635	0.938	0.833	0.936	0.814	0.646	0.677	0.769	0.750	1.145	1.046		0.685	0.685	0.685	0.595
12	0.904	0.924	0.966	0.944	0.895	1.284	0.918	0.962	0.920	0.925	0.931	0.932	0.928	1.359	1.335		**0.882**	**0.882**	**0.882**	0.685
13	**0.858**	0.861	0.922	0.912	0.858	1.407	0.863	0.872	0.864	0.905	0.895	1.119	1.055	1.582	1.427		0.874	0.875	0.876	0.851
14	**0.685**	0.688	0.868	0.835	0.689	1.065	0.689	0.827	0.736	0.723	0.715	0.910	0.867	1.311	1.187		0.704	0.706	0.706	0.676
15	**0.567**	0.707	0.891	0.821	0.576	0.879	0.707	0.891	0.753	0.597	0.619	0.718	0.690	1.052	0.930		0.592	0.594	0.594	0.545
16	0.814	0.891	0.948	0.907	0.817	1.153	0.882	0.942	0.868	0.829	0.849	**0.726**	0.729	0.828	0.807		0.819	0.818	0.818	0.575
17	0.813	0.810	0.913	0.902	0.816	1.181	0.775	0.871	0.812	0.838	0.854	0.782	0.769	0.855	0.839		**0.763**	0.764	0.764	0.734
18	0.781	0.782	0.894	0.879	0.783	1.159	0.762	0.865	0.803	0.835	0.852	0.757	0.748	0.831	0.815		**0.741**	0.742	0.743	0.705
19	0.769	0.768	0.911	0.893	0.773	1.122	0.739	0.890	0.802	0.801	0.837	0.704	**0.690**	0.787	0.752		0.700	0.700	0.700	0.654
20	0.706	0.723	0.885	0.855	0.708	1.056	0.712	0.873	0.778	0.801	0.835	0.655	**0.643**	0.730	0.705		0.664	0.664	0.664	0.605
21	1.240	1.020	1.006	1.030	**1.000**	1.097	1.020	1.006	1.043	1.004	1.006	1.540	1.547	1.689	1.586		1.052	1.052	1.052	0.998
22	0.829	0.841	0.934	0.909	0.834	1.217	0.830	0.920	0.848	0.817	0.822	0.934	0.909	1.293	1.195		0.815	**0.814**	0.814	0.707

Notes: See the notes to Table 2.

Table 4: Root MSFE (T=100, h=2)

DGP	Estimated break date ($\hat{T}_1$)						Unknown break date			Cai's methods						New methods			
	Postbk	CV	WA	Pooled	Troff	WLS	CV	WA	Pooled	Cai1	Cai2	LL1	LL2	LQ1	LQ2	OptR1	OptR2	OptR3	True
8	1.118	1.024	**1.007**	1.008	1.097	1.048	1.031	1.017	1.031	1.441	1.478	3.244	3.365	9.637	10.161	1.077	1.077	1.077	0.978
9	**0.791**	0.793	0.924	0.908	0.796	0.868	0.798	0.809	0.795	0.975	0.961	2.304	2.164	10.152	8.590	0.818	0.820	0.820	0.751
10	0.659	**0.652**	0.903	0.873	0.674	0.721	0.655	0.837	0.720	0.817	0.808	2.007	1.888	10.133	9.164	0.681	0.682	0.682	0.607
11	**0.643**	0.866	0.927	0.884	0.695	0.683	0.865	0.925	0.796	0.721	0.718	1.794	1.808	8.358	8.344	0.778	0.777	0.777	0.528
12	1.229	0.970	0.976	0.961	1.090	1.094	0.968	0.971	**0.947**	1.507	1.501	3.375	3.505	6.837	6.774	0.971	0.971	0.971	0.634
13	0.935	**0.929**	0.949	0.944	0.935	1.084	0.938	0.931	0.940	1.354	1.342	3.366	3.381	8.167	8.950	0.975	0.977	0.977	0.888
14	**0.816**	0.818	0.903	0.888	0.835	0.908	0.822	0.878	0.847	1.157	1.153	2.875	2.908	7.110	7.512	0.861	0.864	0.864	0.758
15	**0.789**	0.892	0.918	0.878	0.828	0.834	0.891	0.918	0.846	1.001	0.997	2.494	2.521	6.682	6.721	0.823	0.824	0.824	0.669
16	1.063	0.959	0.951	0.918	1.040	0.978	0.957	0.948	**0.888**	1.264	1.266	2.571	2.629	7.128	7.396	0.912	0.912	0.912	0.677
17	0.822	0.825	0.920	0.909	0.832	0.859	0.804	0.865	0.818	0.910	0.908	1.950	1.993	7.921	7.585	**0.794**	0.796	0.796	0.710
18	0.945	0.915	0.930	0.924	0.945	0.966	0.910	0.919	**0.904**	1.197	1.201	2.863	2.962	6.517	7.470	0.907	0.909	0.909	0.814
19	0.767	0.775	0.910	0.892	0.780	0.807	0.749	0.873	0.786	0.812	0.821	1.731	1.757	7.336	7.099	**0.716**	0.717	0.717	0.604
20	0.888	0.872	0.918	0.901	0.898	0.909	0.868	0.909	0.868	1.082	1.089	2.549	2.654	5.917	6.596	**0.849**	0.850	0.850	0.739
21	1.113	0.991	1.024	1.029	1.000	3.426	0.991	1.024	1.045	1.117	1.148	1.184	**0.911**	7.240	8.066	1.033	1.033	1.033	0.975
22	0.883	0.873	0.934	0.913	0.885	0.915	**0.861**	0.917	0.865	1.078	1.094	2.437	2.482	6.962	7.412	0.873	0.874	0.874	0.702

Notes: See the notes to Table 2.

Table 5: Root MSFE (T=200, h=2)

DGP	Estimated break date ($\hat{T}_1$)						Unknown break date			Cai's methods						New methods			
	Postbk	CV	WA	Pooled	Troff	WLS	CV	WA	Pooled	Cai1	Cai2	LL1	LL2	LQ1	LQ2	OptR1	OptR2	OptR3	True
8	1.051	1.009	**1.002**	1.003	1.043	1.031	1.013	1.005	1.012	1.187	1.217	1.854	2.017	3.252	3.614	1.028	1.028	1.028	0.991
9	0.757	**0.753**	0.918	0.900	0.760	0.803	0.759	0.791	0.764	0.848	0.851	1.295	1.314	2.595	2.638	0.764	0.768	0.769	0.737
10	0.639	**0.633**	0.894	0.865	0.644	0.669	0.637	0.836	0.710	0.711	0.715	1.106	1.135	2.216	2.296	0.656	0.657	0.657	0.617
11	**0.568**	0.788	0.913	0.862	0.588	0.594	0.788	0.913	0.776	0.609	0.612	0.945	0.980	1.937	2.037	0.634	0.635	0.635	0.528
12	0.922	0.946	0.974	0.958	0.921	0.920	0.931	0.964	0.924	0.967	0.967	1.203	1.236	2.326	2.451	**0.896**	**0.896**	**0.896**	0.651
13	**0.914**	0.918	0.947	0.943	0.917	1.003	0.923	0.922	0.923	1.096	1.105	1.812	1.897	3.329	3.615	0.941	0.942	0.943	0.900
14	**0.785**	0.788	0.896	0.880	0.796	0.843	0.789	0.866	0.823	0.930	0.939	1.543	1.611	2.898	3.092	0.813	0.817	0.818	0.765
15	**0.694**	0.798	0.912	0.862	0.719	0.740	0.798	0.912	0.818	0.781	0.792	1.302	1.351	2.418	2.567	0.718	0.719	0.720	0.650
16	0.941	0.943	0.963	0.934	0.937	0.919	0.933	0.956	0.900	0.919	0.923	1.185	1.235	2.032	2.175	0.882	0.882	**0.882**	0.682
17	0.804	0.801	0.915	0.904	0.809	0.823	0.763	0.859	0.799	0.809	0.817	1.057	1.153	1.952	2.210	**0.750**	0.752	0.752	0.707
18	0.862	0.861	0.918	0.908	0.868	0.890	0.854	0.905	0.872	0.941	0.957	1.463	1.594	2.555	2.883	**0.841**	0.842	0.843	0.795
19	0.750	0.749	0.906	0.888	0.759	0.765	0.709	0.873	0.775	0.714	0.715	0.956	1.014	1.819	2.010	**0.666**	0.666	0.666	0.605
20	0.799	0.813	0.909	0.886	0.813	0.832	0.808	0.901	0.842	0.847	0.853	1.289	1.395	2.254	2.527	**0.774**	0.776	0.777	0.713
21	1.019	1.036	1.018	1.028	1.000	5.852	1.036	1.018	1.027	1.048	1.048	**0.836**	**0.836**	2.776	3.031	1.038	1.038	1.038	0.984
22	0.837	0.852	0.934	0.911	0.846	0.864	0.843	0.921	0.857	0.862	0.873	1.356	1.417	2.471	2.582	**0.832**	0.834	0.834	0.699

Notes: See the notes to Table 2.

Table 6: Root MSFE (T=100, h=1): break tests with different trimming range and significance level

	$\alpha = 0.05$, $trim = 0.05$				$\alpha = 0.1$, $trim = 0.15$			
DGP	OptR1	OptR2	OptR3	True	OptR1	OptR2	OptR3	True
1	1.0013	1.0013	1.0013	0.9999	1.0028	1.0028	1.0028	0.9997
2	0.8790	0.8790	0.8790	0.8671	0.8790	0.8790	0.8790	0.8671
3	0.7218	0.7218	0.7218	0.7105	0.7218	0.7218	0.7218	0.7105
4	0.5959	0.5959	0.5959	0.5047	0.5959	0.5959	0.5959	0.5047
5	0.5879	0.5879	0.5879	0.5223	0.5879	0.5879	0.5879	0.5223
6	0.5855	0.5855	0.5855	0.4470	0.5855	0.5855	0.5855	0.4470
7	0.8816	0.8816	0.8816	0.8175	0.8795	0.8795	0.8795	0.8103
8	1.0075	1.0075	1.0075	0.9993	1.0127	1.0127	1.0127	0.9983
9	0.8208	0.8222	0.8222	0.7724	0.8204	0.8218	0.8218	0.7722
10	0.7228	0.7232	0.7232	0.6679	0.7228	0.7232	0.7232	0.6679
11	0.8322	0.8311	0.8311	0.6139	0.8319	0.8308	0.8308	0.6127
12	0.9170	0.9167	0.9167	0.6394	0.9170	0.9167	0.9167	0.6394
13	0.9170	0.9183	0.9183	0.8571	0.9177	0.9191	0.9191	0.8570
14	0.7581	0.7595	0.7595	0.6963	0.7581	0.7595	0.7595	0.6963
15	0.7161	0.7160	0.7160	0.5848	0.7155	0.7153	0.7153	0.5825
16	0.8448	0.8446	0.8446	0.5774	0.8448	0.8446	0.8446	0.5774
17	0.8260	0.8264	0.8264	0.7701	0.8225	0.8230	0.8230	0.7633
18	0.8265	0.8274	0.8274	0.7644	0.8184	0.8192	0.8192	0.7525
19	0.7640	0.7643	0.7643	0.6787	0.7611	0.7613	0.7613	0.6724
20	0.7582	0.7588	0.7588	0.6667	0.7516	0.7523	0.7523	0.6559
21	1.0376	1.0375	1.0375	0.7881	1.0376	1.0375	1.0375	0.7881
22	0.8224	0.8228	0.8228	0.7108	0.8213	0.8217	0.8217	0.7015

Notes: α is the significance level and $trim$ is the trimming rate for Bai and Perron's (1998) test. When $trim = 0.15$, for example, $[0.15T, 0.85T]$ is used. See also the notes to Table 2.

Table 7: Root MSFE (T=200, h=1): break tests with different trimming range and significance level

	$\alpha = 0.05$, $trim = 0.05$				$\alpha = 0.1$, $trim = 0.15$			
DGP	OptR1	OptR2	OptR3	True	OptR1	OptR2	OptR3	True
1	1.0004	1.0004	1.0004	0.9999	1.0017	1.0017	1.0017	0.9999
2	0.8696	0.8698	0.8698	0.8641	0.8696	0.8698	0.8698	0.8641
3	0.7108	0.7108	0.7108	0.7055	0.7108	0.7108	0.7108	0.7055
4	0.5079	0.5079	0.5079	0.4995	0.5079	0.5079	0.5079	0.4995
5	0.5533	0.5533	0.5533	0.5113	0.5533	0.5533	0.5533	0.5113
6	0.5291	0.5291	0.5291	0.4362	0.5291	0.5291	0.5291	0.4362
7	0.8591	0.8591	0.8591	0.8033	0.8586	0.8586	0.8586	0.8020
8	1.0028	1.0028	1.0028	1.0002	1.0046	1.0047	1.0047	0.9990
9	0.7535	0.7547	0.7554	0.7376	0.7535	0.7547	0.7554	0.7376
10	0.6771	0.6778	0.6781	0.6538	0.6771	0.6778	0.6781	0.6538
11	0.6849	0.6848	0.6847	0.5952	0.6849	0.6848	0.6847	0.5952
12	0.8818	0.8818	0.8818	0.6851	0.8818	0.8818	0.8818	0.6851
13	0.8740	0.8754	0.8759	0.8513	0.8740	0.8754	0.8759	0.8513
14	0.7039	0.7057	0.7063	0.6764	0.7039	0.7057	0.7063	0.6764
15	0.5924	0.5936	0.5939	0.5453	0.5924	0.5936	0.5939	0.5453
16	0.8185	0.8181	0.8180	0.5751	0.8185	0.8181	0.8180	0.5751
17	0.7633	0.7640	0.7643	0.7361	0.7632	0.7639	0.7642	0.7359
18	0.7414	0.7422	0.7429	0.7049	0.7414	0.7422	0.7429	0.7050
19	0.6998	0.7002	0.7004	0.6541	0.6998	0.7002	0.7004	0.6541
20	0.6635	0.6642	0.6643	0.6053	0.6636	0.6642	0.6643	0.6053
21	1.0530	1.0530	1.0530	0.9975	1.0530	1.0530	1.0530	0.9975
22	0.8139	0.8130	0.8131	0.7051	0.8142	0.8133	0.8134	0.7032

Notes: α is the significance level and $trim$ is the trimming rate for Bai and Perron's (1998) test. When $trim = 0.15$, for example, $[0.15T, 0.85T]$ is used. See also the notes to Table 2.

Table 8: Root MSFE (T=100, h=2): break tests with different trimming range and significance level

	$\alpha = 0.05$, $trim = 0.05$				$\alpha = 0.1$, $trim = 0.15$			
DGP	OptR1	OptR2	OptR3	True	OptR1	OptR2	OptR3	True
8	1.0846	1.0846	1.0846	0.9748	1.0814	1.0815	1.0815	0.9783
9	0.8179	0.8204	0.8204	0.7517	0.8178	0.8203	0.8203	0.7514
10	0.6809	0.6824	0.6824	0.6070	0.6809	0.6824	0.6824	0.6070
11	0.7776	0.7768	0.7768	0.5281	0.7776	0.7768	0.7768	0.5281
12	0.9707	0.9707	0.9707	0.6341	0.9707	0.9707	0.9707	0.6341
13	0.9753	0.9772	0.9772	0.8874	0.9753	0.9772	0.9772	0.8875
14	0.8605	0.8640	0.8640	0.7584	0.8605	0.8640	0.8640	0.7584
15	0.8225	0.8229	0.8229	0.6655	0.8224	0.8229	0.8229	0.6672
16	0.9117	0.9116	0.9116	0.6766	0.9117	0.9116	0.9116	0.6766
17	0.7940	0.7957	0.7957	0.7096	0.7941	0.7959	0.7959	0.7099
18	0.9065	0.9087	0.9087	0.8126	0.9064	0.9086	0.9086	0.8131
19	0.7159	0.7167	0.7167	0.6039	0.7159	0.7168	0.7168	0.6040
20	0.8479	0.8492	0.8492	0.7370	0.8484	0.8496	0.8496	0.7383
21	1.0326	1.0329	1.0329	0.9746	1.0326	1.0329	1.0329	0.9746
22	0.8715	0.8728	0.8728	0.6985	0.8725	0.8738	0.8738	0.7012

Notes: α is the significance level and $trim$ is the trimming rate for Bai and Perron's (1998) test. When $trim = 0.15$, for example, $[0.15T, 0.85T]$ is used. See also the notes to Table 2.

Table 9: Root MSFE (T=200, h=2): break tests with different trimming range and significance level

	$\alpha = 0.05$, $trim = 0.05$				$\alpha = 0.1$, $trim = 0.15$			
DGP	OptR1	OptR2	OptR3	True	OptR1	OptR2	OptR3	True
8	1.0391	1.0391	1.0389	0.9877	1.0312	1.0312	1.0310	0.9906
9	0.7643	0.7679	0.7687	0.7372	0.7643	0.7679	0.7687	0.7372
10	0.6556	0.6566	0.6568	0.6171	0.6556	0.6566	0.6568	0.6171
11	0.6342	0.6346	0.6346	0.5277	0.6342	0.6346	0.6346	0.5277
12	0.8962	0.8962	0.8962	0.6505	0.8962	0.8962	0.8962	0.6505
13	0.9410	0.9423	0.9433	0.8998	0.9410	0.9423	0.9433	0.8998
14	0.8133	0.8165	0.8176	0.7649	0.8133	0.8165	0.8176	0.7649
15	0.7174	0.7187	0.7194	0.6497	0.7174	0.7187	0.7194	0.6499
16	0.8822	0.8816	0.8815	0.6823	0.8822	0.8816	0.8815	0.6823
17	0.7499	0.7518	0.7524	0.7067	0.7499	0.7518	0.7524	0.7067
18	0.8404	0.8418	0.8424	0.7950	0.8404	0.8418	0.8425	0.7950
19	0.6655	0.6662	0.6663	0.6054	0.6655	0.6662	0.6663	0.6054
20	0.7742	0.7757	0.7764	0.7128	0.7742	0.7757	0.7764	0.7128
21	1.0381	1.0381	1.0381	0.9836	1.0381	1.0381	1.0381	0.9836
22	0.8319	0.8332	0.8333	0.6941	0.8322	0.8335	0.8336	0.6967

Notes: α is the significance level and $trim$ is the trimming rate for Bai and Perron's (1998) test. When $trim = 0.15$, for example, $[0.15T, 0.85T]$ is used. See also the notes to Table 2.

Table 10: Data Description

Mnemonics	Description	Transformation	Other Information: Seasonal Adjustment, Frequency, Units, Source
	Asset Prices		
fedfunds	Effective Federal Funds Rate	level	NSA, M, Percent, FRED
tb3ms	3-Month Treasury Bill: Secondary Market Rate	level	NSA, M, Percent, FRED
t10yr	10-Year Treasury Constant Maturity Rate	level	NSA, Q, Percent, FRED
termspread	Term Spread: t10yr−fedfunds	level	
sp500	S&P 500 Stock Price Index	$\Delta\ln$	NSA, M, Index, Yahoo
rfedfunds	Real Federal Funds Rate: fedfunds−CPI inflation rate	level	
rtb3ms	Real 3-month Treasury Bill: tb3ms−CPI inflation rate	level	
rt10yr	Real 10-Year Treasury Constant Maturity Rate: t10yr−CPI inflation rate	level	
rsp500	Real Stock Price Index: sp500×100/CPI	$\Delta\ln$	
	Real Economic Activity		
rgdp	Real Gross Domestic Product	$\Delta\ln$	SAAR, Q, Billions of Chained 2009 Dollars, FRED
rdpi	Real disposable personal income: Per capita	$\Delta\ln$	SAAR, Q, Chained 2009 Dollars, FRED
rgpdi	Real Gross Private Domestic Investment	$\Delta\ln$	SAAR, Q, Billions of Chained 2009 Dollars, FRED
ip	Industrial Production Index	$\Delta\ln$	SA, Q, Index 2007=100, FRED
emp	Civilian Employment-Population Ratio	$\Delta\ln$	SA, Q, Percent, FRED
unemp	Civilian Unemployment Rate	Δ	SA, Q, Percent, FRED
unempwomen	Unemployment Level - Women	Δ	SA, M, Percent, FRED
houst	Housing Starts: Total: New Privately Owned Housing Units Started	$\Delta\ln$	SAAR, Q, Thousands of Units, FRED
buildpermits	New Private Housing Units Authorized by Building Permits	$\Delta\ln$	SAAR, M, Thousands of Units, FRED
	Commodity Prices and Price Indices		
gdpdef	Gross Domestic Product: Implicit Price Deflator	$\Delta\ln$	SA, Q, 2009=100, FRED

Continued on the next page

ID	Description	Transformation	Other Information: Seasonal Adjustment, Frequency, Units, Source
cpi	Consumer Price Index for All Urban Consumers: All Items	$\Delta \ln$	SA, Q, Index 1982-84=100, FRED
cpiappsl	Consumer Price Index for All Urban Consumers: Apparel	$\Delta \ln$	SA, Q, Index 1982-84=100, FRED
cpiengsl	Consumer Price Index for All Urban Consumers: Energy	$\Delta \ln$	SA, M, Index 1982-84=100, FRED
ppi	Producer Price Index: All Commodities	$\Delta \ln$	NSA, Q, Index 1982=100, FRED
	Monetary Measure		
m0	St. Louis Adjusted Reserves	$\Delta \ln$	SA, M, Billions of Dollars, FRED
m1	M1 Money Stock	$\Delta \ln$	SA, Q, Billions of Dollars, FRED
m2	M2 Money Stock	$\Delta \ln$	SA, M, Billions of Dollars, FRED
rm0	Real M0: M0×100/CPI	$\Delta \ln$	
rm1	Real M1: M1×100/CPI	$\Delta \ln$	
rm2	Real M2: M2×100/CPI	$\Delta \ln$	

Notes: The following abbreviations appear in the table: SA: seasonally adjusted; NSA: not seasonally adjusted; SAAR: seasonally adjusted at an annual rate; Q: quarterly; M: monthly. Let S_t denote the original series and X_t denote the series used in regressions. The transformations are: (1) level: $X_t = S_t$; (2) $\Delta \ln$: $X_t = \ln S_t - \ln S_{t-1}$; (3)$\Delta$: $X_t = S_t - S_{t-1}$. Series that are not seasonally adjusted are transformed into seasonally adjusted series by the X-11-ARIMA method for estimation (SAS PROC X11).

Table 11: Pseudo Rolling Out-of-Sample Forecasts of the Real GDP Growth Rate: $y_{t+1} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+1}$.

	1984:Q1-2014:Q3										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
Univariate Models: $y_{t+1} = \mu_t + \beta_t(L)y_t + u_{t+1}$											
AR(1)	0.582	0.964	0.970	0.989	1.031	0.995	0.970	0.984	0.950	0.950	0.949
	–	0.014	0.036	0.201	0.834	0.397	0.012	0.347	0.001	0.001	0.001
AR(AIC)	0.593	0.969	0.958	0.978	1.011	1.010	0.960	0.965	0.960	0.960	0.959
	–	0.131	0.069	0.109	0.662	0.674	0.024	0.188	0.000	0.000	0.000
AR(BIC)	0.604	0.989	0.970	0.999	0.980	1.013	0.976	0.976	0.974	0.973	0.973
	–	0.324	0.110	0.481	0.274	0.733	0.113	0.255	0.016	0.016	0.015
ADL(BIC) Models: $y_{t+1} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+1}$											
fedfunds	0.639	1.009	0.977	0.923	1.026	1.062	0.945	0.939	0.974	0.973	0.973
	–	0.636	0.182	0.001	0.675	0.911	0.005	0.028	0.031	0.029	0.029
tb3ms	0.636	0.977	0.958	0.926	1.061	1.107	0.949	0.951	0.983	0.982	0.982
	–	0.153	0.041	0.002	0.812	0.962	0.007	0.065	0.129	0.122	0.122
t10yr	0.604	0.985	0.964	0.968	1.083	1.114	0.969	0.983	0.987	0.986	0.986
	–	0.281	0.084	0.115	0.931	0.992	0.065	0.309	0.194	0.182	0.182
termspread	0.621	1.016	0.982	0.955	1.001	1.011	0.990	0.972	0.967	0.967	0.967
	–	0.734	0.223	0.029	0.508	0.710	0.302	0.208	0.002	0.002	0.002
sp500	0.602	0.994	0.983	0.983	0.998	0.995	0.987	0.980	0.977	0.977	0.975
	–	0.344	0.136	0.323	0.475	0.394	0.206	0.319	0.060	0.061	0.042
rfedfunds	0.639	1.009	0.977	0.924	1.026	1.062	0.945	0.939	0.974	0.973	0.973
	–	0.636	0.181	0.002	0.677	0.913	0.005	0.028	0.031	0.028	0.028
rtb3ms	0.636	0.978	0.958	0.926	1.061	1.107	0.949	0.951	0.983	0.982	0.982
	–	0.164	0.041	0.002	0.813	0.963	0.007	0.066	0.127	0.120	0.120
rt10yr	0.604	0.985	0.964	0.968	1.083	1.112	0.969	0.983	0.986	0.986	0.985
	–	0.283	0.086	0.115	0.930	0.992	0.066	0.311	0.189	0.177	0.174
rsp500	0.613	0.992	0.983	0.976	0.999	1.009	0.990	0.969	0.981	0.982	0.981
	–	0.273	0.133	0.238	0.485	0.689	0.258	0.222	0.092	0.097	0.090
rdpi	0.635	0.980	0.960	0.926	0.999	1.018	0.973	0.929	0.981	0.981	0.980
	–	0.181	0.041	0.040	0.487	0.784	0.080	0.055	0.054	0.057	0.053
rgpdi	0.601	0.980	0.979	0.989	0.977	1.017	0.985	0.982	0.978	0.977	0.977
	–	0.202	0.229	0.335	0.252	0.793	0.253	0.298	0.031	0.031	0.029
ip	0.581	0.984	0.967	1.008	1.030	1.011	0.973	1.015	0.958	0.958	0.957
	–	0.257	0.118	0.583	0.754	0.765	0.123	0.632	0.004	0.003	0.003
emp	0.615	0.975	0.967	0.966	0.976	1.011	0.970	0.936	0.973	0.975	0.974
	–	0.146	0.090	0.119	0.269	0.652	0.064	0.029	0.020	0.028	0.026
unemp	0.616	0.994	0.975	0.951	1.040	1.073	0.980	0.930	0.976	0.975	0.974
	–	0.338	0.068	0.027	0.740	0.902	0.126	0.035	0.031	0.028	0.026
unempwomen	0.592	0.964	0.943	1.006	1.053	1.010	0.960	0.977	0.968	0.967	0.967
	–	0.073	0.012	0.573	0.835	0.733	0.048	0.291	0.005	0.005	0.004
houst	0.611	0.987	0.986	1.001	1.000	1.030	0.985	0.973	0.987	0.986	0.985
	–	0.229	0.228	0.514	0.502	0.773	0.201	0.293	0.155	0.142	0.123
buildpermits	0.604	0.997	0.988	0.959	0.983	1.024	0.993	0.954	1.003	1.004	1.003
	–	0.440	0.257	0.213	0.299	0.831	0.354	0.110	0.599	0.604	0.584
gdpdef	0.610	0.988	0.970	0.955	1.032	1.028	0.978	0.977	0.980	0.980	0.980
	–	0.326	0.142	0.049	0.770	0.892	0.170	0.263	0.037	0.043	0.043
cpi	0.610	1.014	0.992	0.969	1.081	1.055	0.995	0.988	1.000	1.000	1.000
	–	0.688	0.387	0.157	0.959	0.965	0.418	0.378	0.507	0.502	0.502
cpiengsl	0.642	0.952	0.941	0.911	0.941	0.958	0.914	0.912	0.924	0.924	0.923
	–	0.188	0.138	0.063	0.117	0.172	0.060	0.095	0.050	0.048	0.047
ppi	0.607	0.985	0.966	1.029	0.980	1.010	0.973	0.971	0.971	0.971	0.970
	–	0.266	0.081	0.772	0.279	0.688	0.083	0.209	0.010	0.010	0.009
m0	1.430	0.855	0.807	0.424	1.277	1.355	0.855	0.408	1.035	1.035	1.036

Continued on the next page

Table 11 – *Continued from previous page*

	1984:Q1-2014:Q3										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
	–	0.172	0.111	0.079	0.748	0.835	0.105	0.075	0.852	0.853	0.856
m1	0.678	0.986	0.974	0.954	0.963	1.010	0.958	0.906	0.970	0.952	0.952
	–	0.269	0.116	0.018	0.160	0.613	0.014	0.012	0.042	0.025	0.024
m2	0.619	1.004	0.988	0.977	0.979	1.011	0.971	0.965	0.968	0.968	0.968
	–	0.592	0.245	0.167	0.262	0.709	0.095	0.178	0.006	0.006	0.005
rm0	1.315	0.956	0.939	0.465	1.263	1.204	0.919	0.440	1.034	1.034	1.061
	–	0.093	0.277	0.100	0.846	0.727	0.171	0.092	0.864	0.866	0.859
rm1	0.668	1.020	1.009	0.985	1.046	1.020	0.992	0.922	0.964	0.964	0.963
	–	0.790	0.619	0.268	0.719	0.693	0.334	0.027	0.001	0.001	0.001
rm2	0.645	0.975	0.957	0.924	1.007	0.997	0.984	0.934	0.959	0.959	0.959
	–	0.181	0.058	0.049	0.551	0.462	0.189	0.050	0.059	0.057	0.057
cpiappsl	0.644	0.975	0.936	0.935	1.048	1.056	0.943	0.939	0.961	0.960	0.960
	–	0.098	0.008	0.026	0.773	0.899	0.034	0.091	0.008	0.008	0.008

Notes: Fixed: fixed window size with $R = 40$; Cai1: Cai's (2007) method based on the uniform kernel; Cai2: Cai's (2007) method based on the Epanechnikov kernel; CV: Pesaran and Timmermann's (2007) cross validation method with unknown break; LL1: local linear regression using window of Cai1; LL2: local linear regression using window of Cai2; AveW: Pesaran and Pick's (2011) AveW method with $w_{min} = 0.2$ and $m = 10$; PPP: Pesaran, Pick and Pranovich's (2013) robust optimal weights in equation (48) that integrate the break date over the entire sample. OptR1: R0=CV with unknown break date, $\underline{R} = \max(1.5T^{2/3}, 20)$ and $\underline{R} = \min(4T^{2/3}, T-h)$; OptR2: R0=CV with unknown break date, $\underline{R} = \max(1.5T^{2/3}, 20)$ and $\underline{R} = \min(5T^{2/3}, T-h)$; OptR3: R0=CV with unknown break date; $\underline{R} = \max(1.5T^{2/3}, 20)$ and $\underline{R} = \min(6T^{2/3}, T-h)$. In column "Fixed", the numbers are RMSFEs. In the other columns, the first number is the RMSFE ratio relative to the RMSFE based on the fixed window size, and the second number is the p-value of the DM test against the model based on the fixed window size.

Table 12: Pseudo Rolling Out-of-Sample Forecasts of Inflation: $\pi_{t+1}-\pi_t = \mu_t+\alpha_t(L)x_t+\beta_t(L)\Delta\pi_t+u_{t+1}$

	1984:Q1-2014:Q3										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
Univariate Models: $\pi_{t+1} - \pi_t = \mu_t + \beta_t(L)\Delta\pi_t + u_{t+1}$											
AR(1)	0.778	0.972	0.982	1.007	1.047	1.020	0.984	1.132	0.989	0.989	0.989
	–	0.065	0.179	0.748	1.000	0.919	0.151	0.961	0.324	0.324	0.324
AR(AIC)	0.792	0.964	0.968	0.994	1.036	1.025	0.974	1.120	1.004	1.003	1.003
	–	0.023	0.053	0.285	0.991	0.943	0.040	0.956	0.602	0.592	0.590
AR(BIC)	0.819	0.964	0.967	1.002	1.040	1.028	0.975	1.083	0.990	0.990	0.990
	–	0.035	0.061	0.561	0.999	0.981	0.051	0.905	0.329	0.328	0.327
ADL(BIC) Models: $\pi^h_{t+h} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+h}$											
fedfunds	0.820	0.959	0.962	0.997	1.050	1.050	0.971	1.080	1.008	1.009	1.008
	–	0.022	0.039	0.393	0.999	0.999	0.033	0.894	0.779	0.805	0.797
tb3ms	0.822	0.956	0.959	0.996	1.043	1.043	0.969	1.077	1.005	1.006	1.005
	–	0.015	0.028	0.344	0.999	0.999	0.021	0.888	0.686	0.714	0.703
t10yr	0.819	0.961	0.964	0.988	1.034	1.034	0.971	1.079	1.010	1.011	1.011
	–	0.029	0.047	0.142	0.997	0.997	0.033	0.888	0.871	0.884	0.879
termspread	0.821	0.962	0.965	1.000	1.040	1.040	0.973	1.080	1.000	1.001	1.001
	–	0.027	0.048	0.512	0.999	0.999	0.041	0.898	0.511	0.550	0.536
sp500	0.813	0.963	0.966	1.003	1.042	1.042	0.976	1.101	1.003	1.004	1.003
	–	0.055	0.082	0.585	1.000	1.000	0.102	0.943	0.602	0.635	0.620
rfedfunds	0.820	0.959	0.962	0.997	1.050	1.050	0.971	1.080	1.008	1.009	1.008
	–	0.022	0.039	0.393	0.999	0.999	0.033	0.894	0.780	0.806	0.797

Continued on the next page

Table 12 – *Continued from previous page*

	1984:Q1-2014:Q3										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
rtb3ms	0.819	0.959	0.962	0.996	1.044	1.044	0.971	1.080	1.006	1.007	1.007
	–	0.020	0.037	0.370	0.999	0.999	0.029	0.894	0.733	0.758	0.749
rt10yr	0.819	0.961	0.964	0.988	1.034	1.034	0.971	1.079	1.012	1.013	1.012
	–	0.029	0.048	0.143	0.997	0.997	0.033	0.888	0.897	0.908	0.904
rsp500	0.803	0.970	0.974	1.005	1.042	1.042	0.984	1.105	1.007	1.007	1.007
	–	0.103	0.143	0.609	0.999	0.999	0.219	0.945	0.705	0.715	0.706
rdpi	0.830	0.959	0.963	0.992	1.037	1.037	0.968	1.083	0.998	0.999	0.999
	–	0.019	0.038	0.251	0.998	0.998	0.026	0.887	0.439	0.480	0.465
rgpdi	0.822	0.958	0.960	0.995	1.043	1.043	0.967	1.088	1.001	1.002	1.001
	–	0.017	0.030	0.325	1.000	1.000	0.018	0.914	0.531	0.568	0.534
ip	0.798	0.977	0.980	1.004	1.036	1.036	0.977	1.123	1.013	1.013	1.012
	–	0.177	0.227	0.584	0.994	0.994	0.113	0.958	0.867	0.871	0.862
emp	0.815	0.961	0.962	0.989	1.046	1.046	0.959	1.146	0.999	1.000	1.000
	–	0.094	0.108	0.233	0.999	0.999	0.023	0.960	0.480	0.516	0.500
unemp	0.795	0.968	0.971	0.996	1.044	1.044	0.967	1.155	1.007	1.009	1.009
	–	0.098	0.158	0.401	0.995	0.995	0.046	0.980	0.737	0.775	0.764
unempwomen	0.821	0.954	0.956	0.995	1.035	1.035	0.958	1.117	0.994	0.996	0.995
	–	0.022	0.035	0.340	0.996	0.996	0.010	0.952	0.279	0.341	0.326
houst	0.853	0.933	0.927	0.973	1.025	1.025	0.941	1.062	0.993	0.993	0.993
	–	0.020	0.015	0.081	0.936	0.936	0.014	0.836	0.249	0.245	0.247
buildpermits	0.832	0.950	0.953	0.985	1.037	1.037	0.961	1.072	1.000	1.000	1.000
	–	0.021	0.020	0.160	0.998	0.998	0.021	0.870	0.494	0.493	0.494
cpi	0.873	0.923	0.923	0.947	1.023	1.023	0.932	1.023	0.989	0.990	0.990
	–	0.128	0.129	0.144	0.977	0.977	0.113	0.617	0.280	0.304	0.298
cpiengsl	0.889	0.955	0.951	0.968	1.083	1.083	0.939	1.011	0.996	0.997	0.997
	–	0.024	0.025	0.081	0.955	0.955	0.005	0.553	0.406	0.430	0.421
ppi	0.819	0.964	0.967	1.002	1.040	1.040	0.975	1.083	1.003	1.004	1.004
	–	0.035	0.061	0.561	0.999	0.999	0.051	0.905	0.638	0.674	0.661
m0	0.819	0.964	0.967	1.002	1.040	1.040	0.975	1.083	1.003	1.004	1.004
	–	0.035	0.061	0.561	0.999	0.999	0.051	0.905	0.638	0.674	0.661
m1	0.819	0.964	0.967	1.002	1.040	1.040	0.975	1.083	1.003	1.004	1.004
	–	0.035	0.061	0.561	0.999	0.999	0.051	0.905	0.638	0.674	0.661
m2	0.775	0.999	1.001	1.030	1.052	1.052	1.000	1.171	1.019	1.019	1.019
	–	0.492	0.512	0.934	0.998	0.998	0.496	0.984	0.846	0.852	0.847
rm0	0.819	0.964	0.967	1.002	1.040	1.040	0.975	1.083	1.003	1.004	1.004
	–	0.035	0.061	0.561	0.999	0.999	0.051	0.905	0.638	0.674	0.661
rm1	0.819	0.964	0.967	1.002	1.040	1.040	0.975	1.083	1.003	1.004	1.004
	–	0.035	0.061	0.561	0.999	0.999	0.051	0.905	0.638	0.674	0.661
rm2	0.814	0.966	0.971	1.005	1.042	1.042	0.977	1.109	1.005	1.006	1.005
	–	0.044	0.087	0.663	0.999	0.999	0.068	0.944	0.683	0.716	0.703
cpiappsl	0.821	0.976	0.980	1.006	1.047	1.047	0.981	1.085	1.010	1.010	1.010
	–	0.128	0.182	0.697	0.999	0.999	0.116	0.912	0.825	0.834	0.827
rgdp	0.819	0.962	0.967	1.005	1.051	1.051	0.971	1.092	0.992	0.991	0.991
	–	0.033	0.070	0.666	1.000	1.000	0.046	0.915	0.246	0.233	0.231

Notes: See the notes to Table 11.

Table 13: Pseudo Rolling Out-of-Sample Forecasts of the Real GDP Growth Rate: $y_{t+h}^{h} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+h}$.

		1984:Q1-2014:Q3			
		h=2	h=4	h=8	h=12
Univariate Models: $y_{t+h}^{h} = \mu_t + \beta_t(L)y_t + u_{t+h}$					
AR(1)	Fixed	0.472	0.414	0.367	0.350
	OptR1	0.957	0.971	0.952	0.932
AR(AIC)	Fixed	0.495	0.419	0.377	0.352
	OptR1	0.963	0.975	0.951	0.947
AR(BIC)	Fixed	0.497	0.441	0.378	0.352
	OptR1	0.963	0.968	0.955	0.947
ADL(BIC) Models: $y_{t+h}^{h} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+h}$					
fedfunds	Fixed	0.518	0.454	0.414	0.372
	OptR1	0.990	1.000	0.939	0.935
tb3ms	Fixed	0.521	0.457	0.414	0.373
	OptR1	0.991	1.008	0.940	0.927
t10yr	Fixed	0.535	0.475	0.409	0.375
	OptR1	0.981	0.994	0.951	0.905
termspread	Fixed	0.511	0.476	0.417	0.364
	OptR1	0.970	0.965	0.970	0.959
sp500	Fixed	0.517	0.446	0.382	0.354
	OptR1	0.974	0.977	0.944	0.945
rfedfunds	Fixed	0.518	0.459	0.414	0.372
	OptR1	0.990	0.995	0.940	0.935
rtb3ms	Fixed	0.521	0.457	0.414	0.373
	OptR1	0.990	1.007	0.941	0.928
rt10yr	Fixed	0.535	0.475	0.409	0.375
	OptR1	0.981	0.993	0.951	0.908
rsp500	Fixed	0.502	0.458	0.379	0.352
	OptR1	0.972	0.998	0.953	0.951
rdpi	Fixed	0.498	0.441	0.382	0.352
	OptR1	0.961	0.973	0.949	0.947
rgpdi	Fixed	0.518	0.452	0.379	0.352
	OptR1	0.961	0.973	0.954	0.948
ip	Fixed	0.503	0.432	0.356	0.345
	OptR1	0.959	0.963	0.963	0.949
emp	Fixed	0.505	0.452	0.379	0.354
	OptR1	0.954	0.964	0.952	0.948
unemp	Fixed	0.495	0.451	0.379	0.354
	OptR1	0.967	0.963	0.953	0.949
unempwomen	Fixed	0.498	0.440	0.379	0.354
	OptR1	0.960	0.975	0.953	0.948
houst	Fixed	0.502	0.453	0.411	0.354
	OptR1	0.968	0.958	0.923	0.948
buildpermits	Fixed	0.495	0.452	0.421	0.356
	OptR1	0.965	0.961	0.908	0.944
gdpdef	Fixed	0.515	0.512	0.445	0.415
	OptR1	1.015	0.996	0.946	0.920
cpi	Fixed	0.522	0.511	0.424	0.416
	OptR1	1.020	1.029	1.001	0.962
cpiengsl	Fixed	0.520	0.461	0.381	0.353
	OptR1	0.937	0.936	0.950	0.945
ppi	Fixed	0.546	0.504	0.407	0.402
	OptR1	0.951	0.968	0.987	0.957
m0	Fixed	0.539	0.480	0.400	0.369

Continued on the next page

Table 13 – *Continued from previous page*

		1984:Q1-2014:Q3			
		h=2	h=4	h=8	h=12
	OptR1	0.983	0.951	1.013	0.962
m1	Fixed	0.572	0.448	0.449	0.417
	OptR1	0.929	0.976	0.931	0.898
m2	Fixed	0.529	0.466	0.422	0.386
	OptR1	0.964	0.968	0.974	0.983
rm0	Fixed	0.545	0.511	0.423	0.374
	OptR1	0.987	0.940	0.963	0.945
rm1	Fixed	0.583	0.488	0.480	0.401
	OptR1	0.929	0.947	0.936	0.906
rm2	Fixed	0.520	0.476	0.404	0.358
	OptR1	0.968	0.982	0.952	0.963
cpiappsl	Fixed	0.535	0.462	0.420	0.402
	OptR1	0.983	0.978	0.964	0.940

Notes: For each model, the first row labeled with "fixed" reports the RMSFE if the fixed rolling window of size 40 is used; the second row labeled with "optR1" reports the ratio of the RMSFE when using the optimal window size over the RMSFE when using the fixed window size. See also the notes to Table 11.

Table 14: Pseudo Rolling Out-of-Sample Forecasts of Inflation: $\pi^h_{t+h} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+h}$

		1984:Q1-2014:Q3			
		h=2	h=4	h=8	h=12
Univariate Models: $\pi^h_{t+h} - \pi_t = \mu_t + \beta_t(L)\Delta\pi_t + u_{t+h}$					
AR(1)	Fixed	0.688	0.688	0.826	0.955
	OptR1	1.019	1.030	1.055	1.053
AR(AIC)	Fixed	0.685	0.691	0.830	0.974
	OptR1	1.031	1.017	1.058	1.047
AR(BIC)	Fixed	0.704	0.707	0.843	0.966
	OptR1	1.018	1.015	1.042	1.056
ADL(BIC) Models: $\pi^h_{t+h} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+h}$					
fedfunds	FixedR	0.717	0.752	0.994	1.179
	OptR1	1.018	1.015	1.048	1.067
tb3ms	FixedR	0.711	0.757	0.994	1.169
	OptR1	1.029	1.021	1.021	1.004
t10yr	FixedR	0.703	0.734	1.004	1.286
	OptR1	1.038	1.036	1.040	1.035
termspread	FixedR	0.714	0.729	0.893	1.063
	OptR1	1.013	1.015	0.992	0.997
sp500	FixedR	0.687	0.694	0.842	0.966
	OptR1	1.020	1.024	1.042	1.056
rfedfunds	FixedR	0.717	0.752	0.995	1.180
	OptR1	1.018	1.014	1.047	1.066
rtb3ms	FixedR	0.711	0.757	0.994	1.170
	OptR1	1.029	1.021	1.019	1.002
rt10yr	FixedR	0.703	0.735	1.004	1.280
	OptR1	1.039	1.034	1.038	1.040

Continued on the next page

Table 14 – *Continued from previous page*

		1984:Q1-2014:Q3			
		h=2	h=4	h=8	h=12
rsp500	FixedR	0.684	0.689	0.842	0.966
	OptR1	1.023	1.026	1.042	1.056
rdpi	FixedR	0.706	0.720	0.923	1.055
	OptR1	1.015	1.029	1.010	1.002
rgpdi	FixedR	0.710	0.723	0.885	1.056
	OptR1	1.011	1.008	1.006	0.987
ip	FixedR	0.663	0.683	0.886	1.015
	OptR1	1.033	1.028	1.024	0.993
emp	FixedR	0.693	0.776	0.883	0.998
	OptR1	0.997	0.992	0.954	0.912
unemp	FixedR	0.656	0.732	0.900	1.065
	OptR1	1.022	1.013	0.976	0.950
unempwomen	FixedR	0.698	0.749	0.902	1.023
	OptR1	1.040	0.972	0.961	0.994
houst	FixedR	0.689	0.691	0.857	0.977
	OptR1	1.011	1.026	1.041	1.041
buildpermits	FixedR	0.711	0.690	0.858	0.962
	OptR1	1.013	1.028	1.037	1.038
cpi	FixedR	0.733	0.776	0.912	1.186
	OptR1	0.989	1.000	1.008	0.975
cpiengsl	FixedR	0.725	0.704	0.847	0.970
	OptR1	1.007	1.017	1.043	1.080
ppi	FixedR	0.704	0.742	0.854	1.112
	OptR1	1.019	0.995	1.043	1.046
m0	FixedR	0.713	0.718	0.921	1.090
	OptR1	1.014	1.012	1.026	1.041
m1	FixedR	0.722	0.705	0.883	1.095
	OptR1	1.004	1.013	1.034	1.017
m2	FixedR	0.656	0.693	0.842	0.961
	OptR1	1.019	1.018	1.034	1.031
rm0	FixedR	0.713	0.716	0.939	1.252
	OptR1	1.015	1.014	1.044	1.079
rm1	FixedR	0.709	0.707	0.917	1.199
	OptR1	1.006	1.024	1.020	0.995
rm2	FixedR	0.693	0.714	0.851	1.001
	OptR1	1.018	1.005	1.039	0.986
cpiappsl	FixedR	0.716	0.753	0.973	1.107
	OptR1	1.014	1.000	1.026	0.923
rgdp	FixedR	0.673	0.695	0.926	1.022
	OptR1	1.024	1.029	1.033	0.998

Notes: See the notes to Table 13.

Table 15: The Great Recession: Pseudo Rolling Out-of-Sample Forecasts of the Real GDP Growth Rate: $y^h_{t+h} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+h}$.

	2007:Q4-2009:Q2										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
Univariate Models: $y^h_{t+h} = \mu_t + \beta_t(L)y_t + u_{t+h}$											
AR(1)	1.234	1.006	0.982	0.997	0.948	0.997	0.998	1.126	0.944	0.944	0.944
	–	0.619	0.177	0.368	0.286	0.483	0.452	0.956	0.078	0.078	0.078

Continued on the next page

Table 15 – *Continued from previous page*

	2007:Q4-2009:Q2										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
AR(AIC)	1.272	0.996	0.981	0.953	1.011	0.997	0.961	1.049	0.982	0.983	0.983
	–	0.434	0.419	0.149	0.601	0.485	0.270	0.683	0.016	0.022	0.022
AR(BIC)	1.370	1.026	1.005	0.999	0.934	0.984	1.001	1.065	1.019	1.019	1.019
	–	0.874	0.524	0.492	0.205	0.414	0.504	0.801	0.812	0.812	0.812
ADL(BIC) Models: $y^h_{t+h} = \mu_t + \alpha_t(L)x_t + \beta_t(L)y_t + u_{t+h}$											
fedfunds	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
tb3ms	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
t10yr	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
termspread	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
sp500	1.328	1.047	1.028	0.961	0.926	0.969	1.049	1.099	1.053	1.053	1.049
	–	0.986	0.980	0.253	0.218	0.059	0.980	0.906	0.962	0.962	0.965
rfedfunds	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
rtb3ms	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
rt10yr	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
rsp500	1.340	1.046	1.026	0.947	0.892	0.971	1.048	1.076	1.046	1.046	1.046
	–	0.984	0.972	0.099	0.087	0.071	0.982	0.876	0.949	0.949	0.949
rdpi	1.529	0.988	0.983	0.823	0.927	0.997	1.005	0.892	1.033	1.033	1.033
	–	0.416	0.381	0.044	0.204	0.480	0.539	0.155	0.968	0.968	0.968
rgpdi	1.395	0.975	1.015	0.951	0.924	1.024	1.021	1.062	1.021	1.021	1.021
	–	0.358	0.569	0.218	0.214	0.677	0.628	0.800	0.857	0.857	0.857
ip	1.227	0.972	0.962	1.051	0.945	0.981	0.966	1.097	0.973	0.973	0.973
	–	0.370	0.349	0.666	0.033	0.267	0.347	0.797	0.169	0.169	0.169
emp	1.436	0.974	0.983	0.878	0.898	0.984	0.996	0.996	1.016	1.016	1.016
	–	0.342	0.398	0.017	0.163	0.416	0.469	0.465	0.770	0.770	0.770
unemp	1.393	1.036	1.020	0.928	1.075	1.138	1.013	0.996	1.013	1.013	1.013
	–	0.960	0.772	0.084	0.638	0.770	0.612	0.472	0.678	0.678	0.678
unempwomen	1.301	0.944	0.921	0.967	1.161	0.997	0.951	1.080	0.997	0.997	0.997
	–	0.233	0.135	0.352	0.829	0.425	0.258	0.840	0.456	0.456	0.456
houst	1.171	1.019	1.030	1.045	0.926	1.108	1.038	1.182	1.035	1.035	1.032
	–	0.952	0.988	0.757	0.166	0.797	0.853	0.986	0.837	0.837	0.816
buildpermits	1.306	1.036	1.017	0.919	0.877	1.002	1.051	1.020	1.048	1.050	1.050
	–	0.950	0.829	0.316	0.051	0.525	0.976	0.608	0.976	0.981	0.981
gdpdef	1.370	1.003	0.991	0.918	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.078	0.264	0.637	0.505	0.801	0.841	0.841	0.841
cpi	1.370	1.003	0.991	0.885	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.061	0.264	0.637	0.505	0.801	0.841	0.841	0.841
cpiengsl	1.386	0.988	0.980	0.906	0.970	1.021	0.992	1.051	1.012	1.012	1.012
	–	0.439	0.391	0.052	0.370	0.651	0.445	0.740	0.686	0.684	0.684
ppi	1.370	1.003	0.991	1.052	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.670	0.264	0.637	0.505	0.801	0.841	0.841	0.841
m0	5.524	0.838	0.784	0.209	1.314	1.405	0.840	0.222	1.054	1.054	1.055
	–	0.187	0.118	0.063	0.736	0.833	0.113	0.063	0.933	0.936	0.938
m1	1.370	1.003	0.991	0.919	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.074	0.264	0.637	0.505	0.801	0.841	0.841	0.841
m2	1.384	1.048	1.034	0.914	0.918	1.022	0.985	1.045	1.007	1.007	1.007
	–	0.907	0.823	0.091	0.144	0.647	0.418	0.692	0.610	0.610	0.610
rm0	4.976	0.958	0.938	0.234	1.314	1.248	0.912	0.245	1.053	1.053	1.085

Continued on the next page

Table 15 – *Continued from previous page*

	2007:Q4-2009:Q2										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
	–	0.146	0.319	0.091	0.847	0.722	0.205	0.090	0.942	0.944	0.906
rm1	1.370	1.003	0.991	0.914	0.941	1.020	1.001	1.065	1.021	1.021	1.021
	–	0.514	0.454	0.130	0.264	0.637	0.505	0.801	0.841	0.841	0.841
rm2	1.571	0.928	0.920	0.810	0.936	0.959	1.001	0.915	0.948	0.948	0.948
	–	0.158	0.121	0.069	0.200	0.314	0.507	0.181	0.250	0.250	0.250
cpiappsl	1.474	0.979	0.942	0.854	1.102	1.092	0.944	0.982	0.974	0.974	0.974
	–	0.319	0.212	0.058	0.698	0.751	0.287	0.433	0.287	0.287	0.287

Notes: See the notes to Table 11.

Table 16: The Great Recession: Pseudo Rolling Out-of-Sample Forecasts of Inflation: $\pi_{t+1} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+1}$

	2007:Q4-2009:Q2										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
Univariate Models: $\pi_{t+1} - \pi_t = \mu_t + \beta_t(L)\Delta\pi_t + u_{t+1}$											
AR(1)	0.981	0.988	0.991	0.989	0.989	0.992	0.998	1.426	1.024	1.024	1.024
	–	0.294	0.391	0.280	0.280	0.105	0.467	0.963	0.687	0.687	0.687
AR(AIC)	1.135	0.923	0.925	0.977	1.019	1.018	0.961	1.236	1.009	1.008	1.008
	–	0.117	0.137	0.120	0.725	0.754	0.143	0.997	0.695	0.682	0.680
AR(BIC)	1.128	0.928	0.931	0.973	1.020	1.018	0.966	1.244	0.997	0.996	0.996
	–	0.133	0.154	0.101	0.740	0.759	0.163	0.998	0.458	0.445	0.443
ADL(BIC) Models: $\pi_{t+1} - \pi_t = \mu_t + \alpha_t(L)x_t + \beta_t(L)\Delta\pi_t + u_{t+1}$											
fedfunds	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
tb3ms	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
t10yr	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
termspread	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
sp500	1.167	0.898	0.901	0.944	1.012	1.012	0.936	1.235	0.989	0.988	0.988
	–	0.059	0.072	0.046	0.654	0.654	0.067	0.998	0.344	0.331	0.330
rfedfunds	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rtb3ms	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rt10yr	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rsp500	1.034	0.966	0.974	0.942	1.016	1.016	1.020	1.317	1.046	1.045	1.045
	–	0.266	0.326	0.115	0.668	0.668	0.598	1.000	0.748	0.743	0.742
rdpi	1.060	0.937	0.946	0.925	1.018	1.018	0.982	1.438	1.012	1.011	1.010
	–	0.117	0.167	0.096	0.759	0.759	0.239	0.999	0.674	0.662	0.660
rgpdi	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
ip	0.868	1.111	1.122	1.036	1.020	1.020	1.058	1.647	1.065	1.058	1.057
	–	0.765	0.789	0.644	0.724	0.724	0.789	0.999	0.761	0.758	0.752
emp	0.718	1.267	1.267	1.079	0.955	0.955	1.106	1.996	1.075	1.076	1.076
	–	0.829	0.829	0.750	0.264	0.264	0.749	0.996	0.858	0.860	0.860
unemp	0.811	1.037	1.123	1.005	1.023	1.023	1.015	1.776	1.073	1.075	1.075
	–	0.569	0.697	0.529	0.595	0.595	0.556	0.998	0.828	0.835	0.835

Continued on the next page

Table 16 – *Continued from previous page*

	2007:Q4-2009:Q2										
	Fixed	Cai1	Cai2	CV	LL1	LL2	AveW	PPP	OptR1	OptR2	OptR3
unempwomen	0.928	1.053	1.054	1.007	0.975	0.975	1.023	1.576	1.019	1.023	1.023
	–	0.740	0.742	0.545	0.261	0.261	0.714	0.996	0.731	0.765	0.765
houst	1.145	0.914	0.917	0.961	1.021	1.021	0.953	1.242	1.002	1.001	1.001
	–	0.086	0.104	0.041	0.756	0.756	0.088	0.998	0.543	0.528	0.526
buildpermits	1.055	0.945	0.954	0.950	1.006	1.006	0.974	1.306	0.998	0.997	0.997
	–	0.244	0.284	0.266	0.548	0.548	0.387	0.986	0.485	0.477	0.477
cpi	1.735	0.612	0.617	0.738	1.031	1.031	0.698	0.803	0.937	0.937	0.937
	–	0.127	0.129	0.135	0.954	0.954	0.133	0.271	0.152	0.152	0.152
cpiengsl	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
ppi	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
m0	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
m1	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
m2	0.973	1.034	1.045	1.018	1.129	1.129	1.067	1.434	1.121	1.120	1.120
	–	0.695	0.723	0.609	0.843	0.843	0.766	0.988	0.806	0.804	0.803
rm0	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rm1	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rm2	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
cpiappsl	1.128	0.928	0.931	0.973	1.020	1.020	0.966	1.244	1.007	1.006	1.006
	–	0.133	0.154	0.101	0.740	0.740	0.163	0.998	0.626	0.612	0.610
rgdp	0.937	0.996	1.039	1.042	1.100	1.100	1.001	1.487	0.963	0.959	0.957
	–	0.489	0.602	0.725	0.919	0.919	0.505	0.986	0.268	0.247	0.238

Notes: See the notes to Table 11.

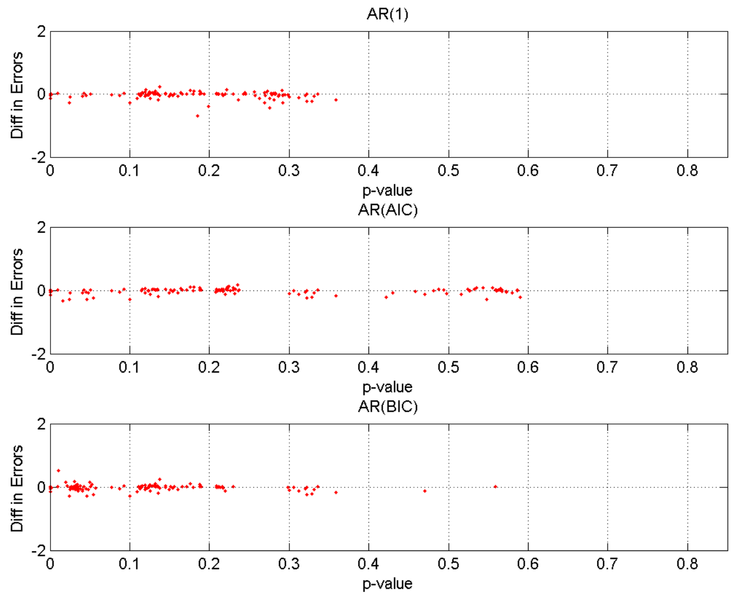

Figure 1: The QLR test for GDP forecasting AR

Notes: The x-axis reports the p-value of the QLR test for parameter constancy; the y-axis reports the squared forecast error of the optimal window size minus that based on the fixed window of size 40 for each time period between 1984:Q1 and 2014:Q3.

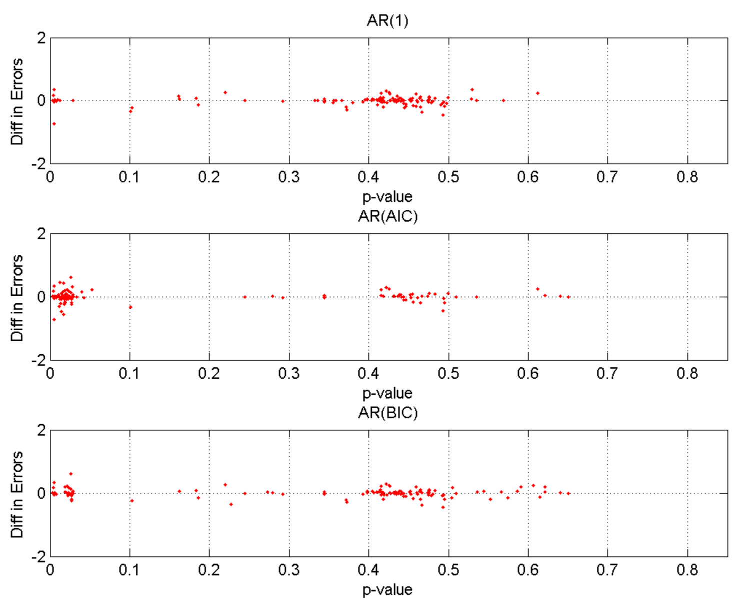

Figure 2: QLR test for inflation forecasting AR

Notes: The x-axis reports the p-value of the QLR test for parameter constancy; the y-axis reports the squared forecast error of the optimal window size minus that based on the fixed window of size 40 for each time period between 1984:Q1 and 2014:Q3.

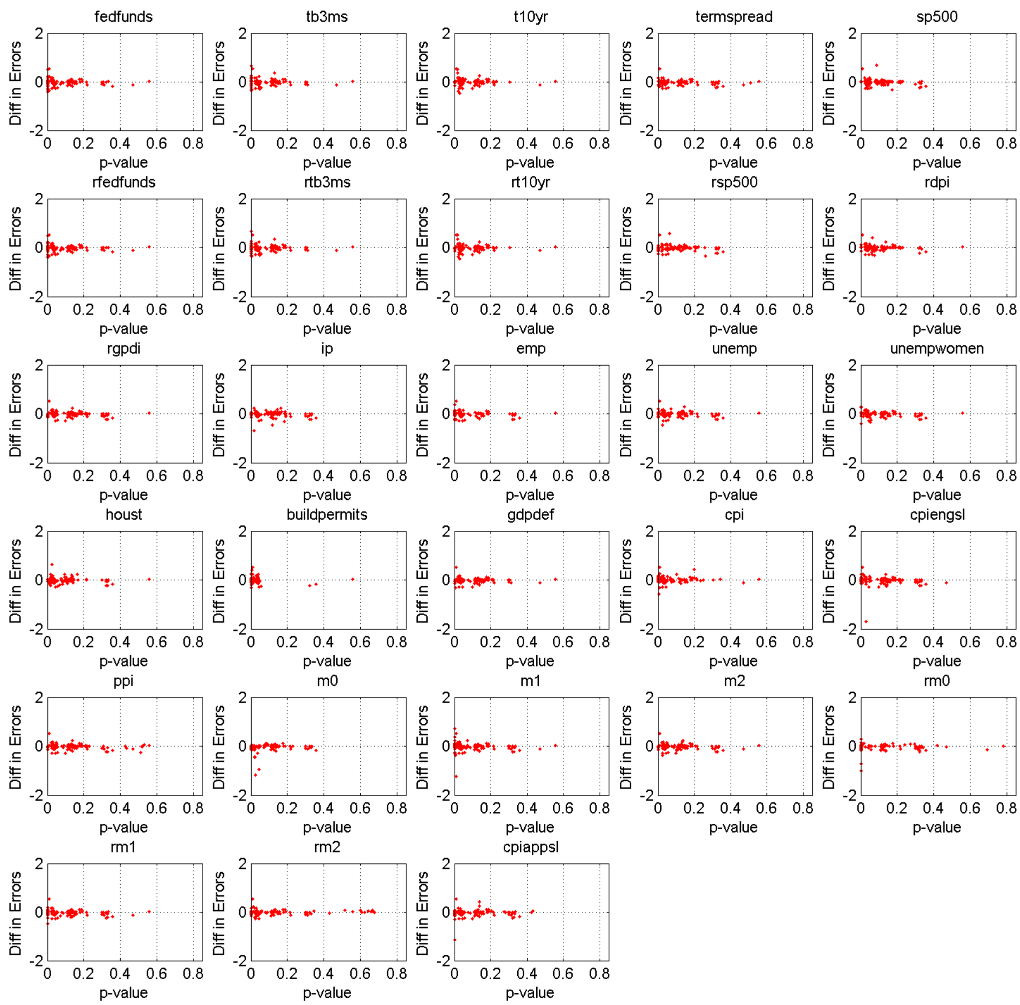

Figure 3: QLR test for GDP forecasting ADL(BIC)

Notes: The x-axis reports the *p*-value of the QLR test for parameter constancy; the y-axis reports the squared forecast error of the optimal window size minus that based on the fixed window of size 40 for each time period between 1984:Q1 and 2014:Q3.

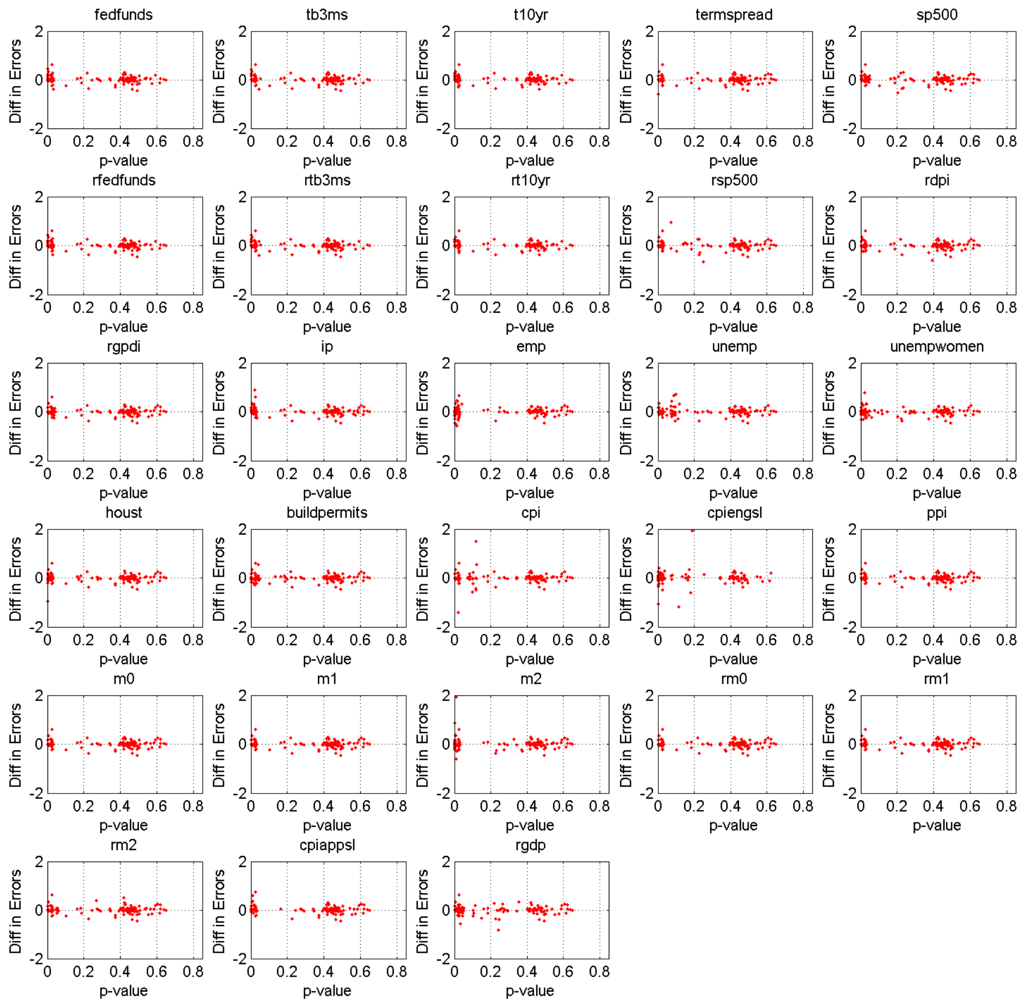

Figure 4: QLR test for inflation forecasting ADL(BIC)

Notes: The x-axis reports the *p*-value of the QLR test for parameter constancy; the y-axis reports the squared forecast error of the optimal window size minus that based on the fixed window of size 40 for each time period between 1984:Q1 and 2014:Q3.

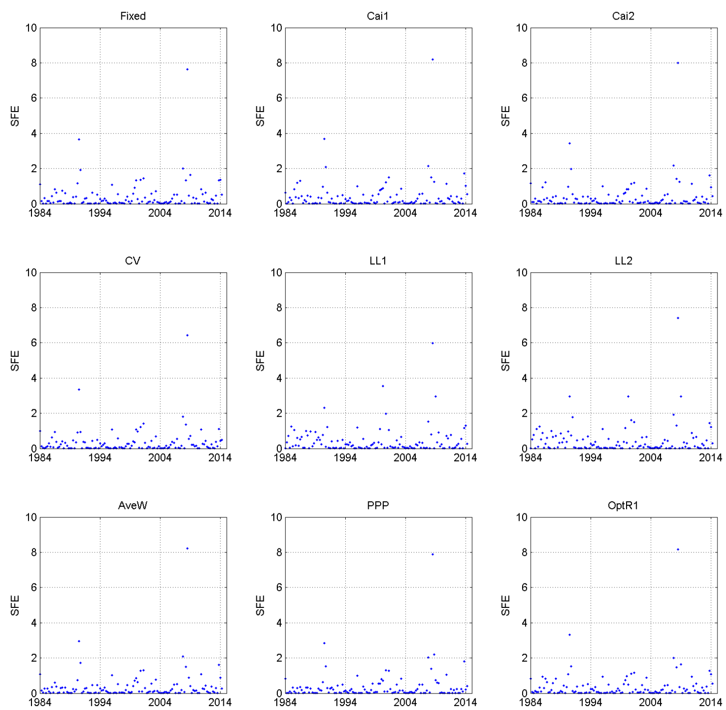

Figure 5: Real GDP forecasting with Fed Funds rate

Notes: The y-axis reports the squared forecast error (SFE) of the optimal window size minus that based on the fixed window of size 40.

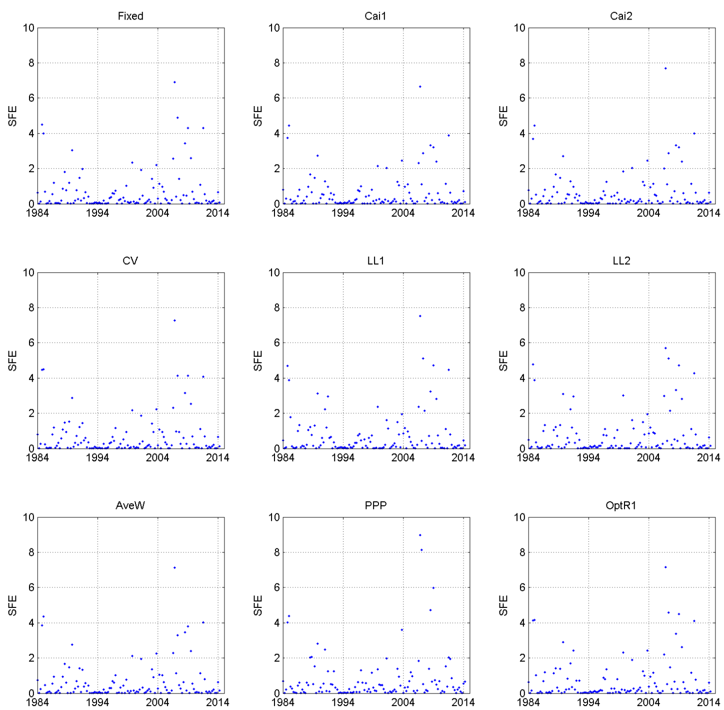

Figure 6: Inflation forecasting with Fed Funds rate

Notes: The y-axis reports the squared forecast error (SFE) of the optimal window size minus that based on the fixed window of size 40.