Testing for Changes in Forecasting Performance*

Pierre Perron† **Yohei Yamamoto**‡

Boston University **Hitotsubashi University**

May 24, 2018; Revised December 12, 2018

Abstract

We consider the issue of forecast failure (or breakdown) and propose methods to assess retrospectively whether a given forecasting model provides forecasts which show evidence of changes with respect to some loss function. We adapt the classical structural change tests to the forecast failure context. First, we recommend that all tests should be carried with a fixed scheme to have best power. This ensures a maximum difference between the fitted in and out-of-sample means of the losses and avoids contamination issues under the rolling and recursive schemes. With a fixed scheme, Giacomini and Rossi's (2009) (GR) test is simply a Wald test for a one-time change in the mean of the total (the in-sample plus out-of-sample) losses at a known break date, say m, the value that separates the in and out-of-sample periods. To alleviate this problem, we consider a variety of tests: maximizing the GR test over values of m within a pre-specified range; a Double sup-Wald (DSW) test which for each m performs a sup-Wald test for a change in the mean of the out-of-sample losses and takes the maximum of such tests over some range; we also propose to work directly with the total loss series to define the Total Loss sup-Wald (TLSW) and Total Loss UDmax (TLUD) tests. Using theoretical analyses and simulations, we show that with forecasting models potentially involving lagged dependent variables, the only tests having a monotonic power function for all data-generating processes considered are the DSW and TLUD tests, constructed with a fixed forecasting window scheme. Some explanations are provided and empirical applications illustrate the relevance of our findings in practice.

JEL Classification Number: C14, C22

Keywords: forecast breakdown, non-monotonic power, structural change, out-of-sample forecast

*We thank he Editor (Todd Clark), an associate editor and two referees for useful comments. Yamamoto acknowledges financial support from the MEXT Fund for the Promotion of Joint International Research.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

‡Department of Economics, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo, Japan, 186-8601 (yohei.yamamoto@econ.hit-u.ac.jp).

1 Introduction

We consider the issue of forecast failure (or breakdown) and propose methods to detect changes in the forecasting performance over time. The aim is to assess retrospectively whether a given forecasting model provides forecasts which show evidence of changes (improvements or deterioration) with respect to some loss function. Since the losses can change because of changes in the variance of the shocks (e.g., good luck), detection of a forecast failure does not necessarily mean that a forecast model should be abandoned. Care must be exercised to assess the source of the changes. But if a model is shown to provide stable forecasts, it can more safely be applied in real time. In practice, such forecasts are made at the time of the last available data, using a fixed, recursive or rolling window. Hence, there is a natural separation between the in and out-of-sample periods dictated by the last data point. Such is not the case when trying to assess retrospectively whether a given model provides stable forecasts. There is a need for a separation between the in and out-of-sample periods at some date labelled m, say. This date should be such that the model in the in-sample period is stable, i.e., yielding stable forecasts. This can create problems since one needs a truncation point m to assess forecast failures but the choice of this value is itself predicated on some knowledge of stability. An example of such test is that of Giacomini and Rossi (2009), GR hereafter. It is a global and retrospective test which compares the in and out-of-sample averages of the sequence of forecast losses. See Casini (2017) for an extension as well as Casini and Perron (2018) and Perron (2006) for a review of the relevant issues.

We adapt classical structural change tests to the forecast failure context. First, we recommend that all tests should be carried with a fixed scheme to have best power, which ensures the maximum difference between the fitted in and out-of-sample means of the losses; contamination issues under the rolling and recursive schemes induce power losses. With this fixed scheme, GR's test is simply a Wald test for a one-time change in the mean of the total (in-sample plus out-of-sample) losses at a known break date m. To alleviate this problem, which leads to power losses when the forecast breakdown is not at m, one can follow Inoue and Rossi (2012) and maximize the GR test over values of m within a pre-specified range; i.e., a sup-Wald test for a single change at a date constrained to separate the in and out-of-sample periods. The test is still not immune to non-monotonic power problems when multiple changes occur. Hence, we propose a Double sup-Wald (DSW) test which for each m performs a sup-Wald test for a change in the mean of the out-of-sample losses and takes the maximum over the range $m \in [m_0, m_1]$: $DSW = \max_{m\in[m_0,m_1]} SW_{L^o(m)}$, where $SW_{L^o(m)}$ is the sup-Wald test for a change in the mean of the out-of-sample losses for a forecast horizon τ, $L_t^o(\hat{\beta})$ for $t = m + \tau, \ldots, T$, defined by $SW_{L^o(m)} = \max_{T_b(m) \in [m+\tau+\epsilon n,\, m+\tau+(1-\epsilon)n]} [SSR_{L^o(m)} - SSR\left(T_b\left(m\right)\right)_{L^o(m)}]/\hat{V}_{L^o(m)}$, where $n = T - m - \tau + 1$, $SSR_{L^o(m)}$ is the restricted sum of squared residuals (SSR), $SSR\left(T_b\left(m\right)\right)_{L^o(m)}$ is the SSR with a change at $T_b\left(m\right)$, and $\hat{V}_{L^o(m)}$ is the long-run variance estimate of the out-of-sample losses (ϵ is a small trimming parameter set at 0.1 throughout). We also propose working directly with the total loss series $L\left(m\right)$ to define the Total Loss sup-Wald ($TLSW$) and UDmax tests ($TLUD$). Using simulations based on the original design of GR, we show that with forecasting models with lagged dependent variables, the only tests with monotonic power for all data-generating processes considered are the DSW and $TLUD$ tests, constructed with a fixed forecasting window scheme.

The benefits of forecast breakdown tests over structural change tests applied to the forecasting model are the following, among others. As stated in GR, they can detect breaks in variance (with a squared error loss); they allow model misspecifications; breaks in coefficients and in variance may offset; they allow for instability in the distribution of the regressors. More interestingly, the two types of tests are complementary. Consider data generated with the predictor changing from x_t to w_t at some date; w_t is not accessible and a regression of $y_{t+\tau}$ on x_t is used. For a given variance of x_t, when the variance of w_t is small the power of the forecasting tests is small, while that of the structural change tests is large (and vice versa), since w_t is part of the error term. This holds for more general tests with unknown multiple breaks. It shows a complementarity between the power of the forecast breakdown and structural change tests. See Supplement (a) for theoretical and simulation analyses.

The paper is structured as follows. Section 2 introduces the statistical framework and some tests: 2.1 reviews the single break case at a known date; 2.2 discusses why using a fixed scheme is preferable; 2.3 considers unknown break dates and proposes new tests. Some limit distributions are stated in Section 3. Section 4 considers the size and power of the tests: 4.1 for the finite sample size of the proposed tests; 4.2 describes the setup to evaluate the power functions; 4.3 contains theoretical results about the shapes of the loss functions. Section 4.4 presents a summary of the results and 4.5 expands on the reasons for non-monotonic power functions. Section 5 provides empirical applications and Section 6 concluding remarks. A supplement contains technical derivations, additional discussions and results.

2 The framework and the tests

We have data (y_t, x_t) with y_t a scalar variable to be forecasted and x_t a q-dimensional vector of predictors for $t = 1, ..., T$. Consider a model forecasting $y_{t+\tau}$ at period t, a τ-period ahead forecast obtained using the direct method is $\hat{y}_{t+\tau} = f_t(\hat{\beta}_m; x_t)$, where f_t is a known function that defines the model, $\hat{\beta}_m$ is the estimate of the parameter β $(q \times 1)$ obtained from an in-sample window of size $m \geq q$; e.g., for the linear model, $f_t(\hat{\beta}_m; x_t) = \hat{\beta}_m x_t$ and $\hat{\beta}_m$ is the OLS estimate from a regression of $y_{t+\tau}$ on x_t using data from the in-sample window. The out-of-sample forecast procedure divides the full sample into an in-sample window of size m and an out-of-sample window of size $n = T - m - \tau + 1$. The model is estimated in the in-sample window and the out-of-sample window is used for forecast error evaluation. We consider three popular forecast schemes: fixed window, with the in-sample consisting of observations 1 to m; rolling window, with the in-sample consisting of observations $t - m + 1$ to t; recursive window, with the in-sample consisting of observations 1 to t. We denote the sequence of in-sample losses by $L_t^i(\hat{\beta}_m)$, defined by the in-sample fitted values $\hat{y}_t = f_t(\hat{\beta}_m; x_t)$, and the sequence of out-of-sample losses by $L_t^o(\hat{\beta}_m)$, defined by the forecast values $\hat{y}_{t+\tau}$. With the popular squared error loss function, $L_{t+\tau}^i(\hat{\beta}_m) = (y_{t+\tau} - f_t(\hat{\beta}_m; x_t))^2$ and $L_{t+\tau}^o(\hat{\beta}_m) = (y_{t+\tau} - \hat{y}_{t+\tau})^2$. We also define the in-sample loss sequence: $L^i(m) = (L_{\tau+1}^i(\hat{\beta}_m), ..., L_m^i(\hat{\beta}_m))$, the out-of-sample loss sequence: $L^o(m) = (L_{m+\tau}^o(\hat{\beta}_m), ..., L_T^o(\hat{\beta}_m))$ and the total loss series as the stacked vector of both, i.e., $L(m) = (L_{\tau+1}^i(\hat{\beta}_m), ..., L_m^i(\hat{\beta}_m),\ L_{m+\tau}^o(\hat{\beta}_m), ..., L_T^o(\hat{\beta}_m))$, a $(T - 2\tau + 1)$ vector. A time-indexed total loss series is denoted by $\{L_{t+\tau}\}_{t=1}^{T-2\tau+1}$. Note that we use a "direct forecast" method when $\tau > 1$. Supplement (g) shows that the main theoretical results also apply when using an indirect forecast method.

The goal is to assess whether there are instabilities in forecast accuracy; e.g., a deterioration usually referred to as a "forecast breakdown". This can occur because of a genuine change in the stability of the forecasting regression, via the conditional mean, or from changes in the variance of the errors. It can also occur if the forecasting model is misspecified in which case an over-fitting problem is possible, so that the out-of-sample losses are inflated relative to the in-sample losses irrespective of whether a change in the stability of the forecasts is present or not. We shall be concerned about the former case. If one wants to guard about potential changes related to over-fitting, one can simply adjust the out-of-sample losses by subtracting a correction factor (see GR for the exact expressions for a linear model and a quadratic loss function). The null hypothesis considered is $H_0 : E[L_t(\beta^*)] = \mu_0$, for all $t = \tau + 1, ..., T - \tau + 1$ for some $\beta^* = p\lim_{m\to\infty} \hat{\beta}_m$, which implicitly assumes that the probability limit of $\hat{\beta}_m$ is the same for all m under the null hypothesis. The alternative hypothesis is $H_1 : E[L_t(\beta^*)] \neq E[L_{t+1}(\beta^*)]$ for at least one $t = \tau + 1, ..., T - \tau$. Hence, we are concerned with testing the stability of the forecast performance in population as opposed to in finite samples using the terminology of Clark and McCracken (2013).

2.1 The case with a single break occurring at a known date

We consider first a single break in forecast accuracy at a known date T_b, say, so that the alternative is $E[L_t(\beta^*)] = \mu_1$ for $t \leq T_b$, μ_2 for $t > T_b$. The obvious thing is then to apply a test for a change in the mean of the total loss series at date T_b, fixed for any given T but increasing with T so that $\lim_{T\to\infty} T_b/T = \lambda$, say, when performing an asymptotic analysis. This is achieved by setting $m = T_b$ and assessing whether the averages of the in and out-of-sample losses are different. This is the test proposed by GR. Define the surprise loss as the out-of-sample loss minus the mean of the in-sample losses, i.e., $SL_{t+\tau}(\hat{\beta}_m) \equiv L^o_{t+\tau}(\hat{\beta}_m) - \bar{L}^i(\hat{\beta}_m)$ where $\bar{L}^i(\hat{\beta}_m) \equiv (m-\tau)^{-1}\sum_{s=\tau+1}^{m} L^i_s(\hat{\beta}_m)$, with the out-of-sample losses adjusted for over-fitting if desired. The test they propose is $GR_m = (n^{-1/2}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\hat{\beta}_m))/\hat{\Omega}^{1/2}$, where $n = T - m - \tau + 1$ is the size of the out-of-sample window and $\hat{\Omega}$ is an estimate of the long-run variance of the loss sequence (see GR for the exact form suggested). It is easy to verify that the square of this test is equivalent to an F-test for a change in mean occurring at date m when applied to the total loss series $L(m)$. This simple observation leads to the following comments. First, it is really inconsequential if we work with the original loss functions (in-sample plus out-of-sample) with a given value of m that assumes a pre-sample with no change or with the "surprise loss functions" for which the average of the in-sample losses is subtracted from the out-of-sample losses. Hence, below, we shall also consider tests constructed via the original (not-demeaned) out-of sample losses, again assuming the model to be stable prior to date m (the demarcation between the in and out-of-sample). Second, the test of GR is problematic since the true break date is unknown in practice even if only one is present. This makes the test sensitive to the choice of m. As will be shown via simulations, it can have non-monotonic power (decreasing as the mean-change of the losses increases) for a range of choices for m. Hence, we also consider tests that allow m to vary within some pre-specified range, whether with the surprise or original losses.

2.2 The choice of the forecasting scheme

Before considering tests that do not assume a known break date, we discuss the merits and drawbacks of the fixed, rolling or recursive forecasting schemes. To get better forecasts it is, in general, better to adopt a recursive or rolling forecasting scheme in the presence of instabilities, or a combination of both; see, e.g., Clark and McCracken (2009). The parameter estimates then adapt to the data-generating process to fit the data better and provide more accurate forecasts. A fixed forecasting scheme fails to provide such adjustments. A rolling one provides adjustments but at the expense of increased variability due to a smaller in-sample window. However, when trying to detect retrospectively whether a change has occurred the opposite ranking applies. The best scheme to adopt is a fixed one. Suppose that the break date is known and m is set accordingly. A fixed scheme ensures the maximum difference between the fitted in and out-of-sample means of the losses. With a recursive scheme, the in-sample fitted mean of the loss series is pushed towards the fitted mean of the out-of-sample losses inducing a loss of power. With a rolling window scheme, the same occurs but in a more pronounced way since the in-sample fitted mean can eventually reach the post-break mean if the window is small. Hence, when testing for changes in forecast accuracy, it is preferable to use a fixed window scheme. This will remain true with one or multiple breaks occurring at unknown dates. We provide explanations with some theory and simulations later.

2.3 The case with unknown break dates

A simple method to alleviate the dependence of the GR_m test on m is to take the supremum over a range of m, say $[m_0, m_1]$, a version denoted by $SGR = \max_{m \in [m_0, m_1]} |GR_m|$. This test is tailored to the alternative hypothesis with T_b unknown. The limit distribution and critical values of SGR are in Inoue and Rossi (2012) for typical values of m_0 and m_1. Alternatively, one could use $SGR^2 = \max_{m \in [m_0, m_1]} GR_m^2$, which is equivalent to a sup-Wald test for a change in mean and use the critical values in Andrews (1993). This modification will, however, not be immune from power problems when multiple changes occur. To see why, consider the case with two breaks. Then for any choice of m in the range $[m_0, m_1]$ at least one segment will be contaminated due to biased parameter estimates and the average loss will be reduced thereby decreasing the power of the test. As we shall see, this problem can be especially severe when the range $[m_0, m_1]$ is large. To avoid it, one can perform a sup-Wald test for a change in the mean of the out-of-sample losses for each value of m and take the maximum of such tests over a range $m \in [m_0, m_1]$, labelled as the Double sup-Wald (DSW) test, defined by $DSW = \max_{m \in [m_0, m_1]} SW_{L^0(m)}$, where $SW_{L^o(m)}$ is the sup-Wald test for a change in the mean of the out-of-sample loss series $L_t^o(\hat{\beta})$ for $t = m + \tau, ..., T$, defined by

$$SW_{L^o(m)} = \max_{T_b(m) \in [m+\tau+\epsilon n, m+\tau+(1-\epsilon)n]} [SSR_{L^o(m)} - SSR(T_b(m))_{L^o(m)}] / \hat{V}_{L^o(m)}, \qquad (1)$$

where $SSR_{L^o(m)}$ is the restricted SSR, $SSR(T_b(m))_{L^o(m)}$ is the SSR with a change at time $T_b(m)$, and $\hat{V}_{L^o(m)}$ is the long-run variance estimate of the demeaned out-of-sample loss series with the mean changing at date $T_b(m)$, obtained using the method of Andrews (1991). The parameter ϵ is a small trimming value set at 0.1. Too small a value leads to size distortions and large ones to power problems. The results are not sensitive to minor variations, e.g., $\epsilon = 0.05$ or 0.15; see Supplement (c). The limit distribution of the DSW test is stated in Section 3. To construct it: 1) Start with an out-of-sample method with an in-sample length m_0 small but large enough to estimate the model. Let $n \equiv T - m - \tau + 1$. 2) Compute the out-of-sample loss series $\{L^o_{t+\tau}\}_{t=m}^{T-\tau}$. 3) Consider a regression with only a constant: $L^o_{t+\tau} = \gamma + e_{t+\tau}$. Apply a sup-Wald test for the constancy of γ, with a HAC variance estimate if there is evidence of serial correlation in the losses. Store the value as $SW_{L^o(m)}$. 4) Update m to $m+1$ and repeat Steps 2-3 up to $m = m_1$. The range of m, $m_1 - m_0$, is some fraction of n, denoted by $\bar{\mu}$. 5) Take the maximum of the sequence of $\{SW_{L^o(m)}\}_{m=m_0}^{m_1}$.

The reason why the DSW test improves upon the SGR test is because it produces three segments instead of only two, which is beneficial with more than one break. Two are defined by m and the other by the date at which the Wald test is maximized in the range $[m+\tau+\epsilon n, m+\tau+(1-\epsilon)n]$. Hence, three segments can be inserted within the total sample, which guarantees that the two segments with the largest difference can be separated by a break, thereby increasing power. With a single break the mean of at least one segment is contaminated by the values of the means in the other segments, reducing power. The idea is akin to that of Qu (2007) who showed that when searching whether any part of a sample is stationary all one needs is a search with two breaks or three segments. To go further, one can also consider a test similar to the UDmax test for multiple changes of Bai and Perron (1998). However, the size distortions were rather high and we shall not consider it further. An alternative is to work directly with the total loss series $L(m)$ instead of only using the out-of-sample losses. This can yield higher power given that more information is used. We consider two tests following this approach: the Total Loss sup-Wald test ($TLSW$) and the Total Loss UDmax test ($TLUD$). More precisely, $TLSW = \max_{m\in[m_0,m_1]} SW_{L(m)}$, where $SW_{L(m)}$ is the sup-Wald test applied for a mean-change in the total loss series $L(m)$:

$$SW_{L(m)} = \max_{T_b(m)\in[\tau+\epsilon(T-2\tau+1),\tau+(1-\epsilon)(T-2\tau+1)]} [SSR_{L(m)} - SSR(T_b(m))_{L(m)}]/\hat{V}_{L(m)},$$

where $SSR_{L(m)}$ is the restricted SSR, $SSR(T_b(m))_{L(m)}$ is the SSR assuming a one-time change at time $T_b(m)$ and $\hat{V}_{L(m)}$ is the long-run variance estimate of the total loss series using demeaned total loss series with the mean changing at date $T_b(m)$; ϵ is some small trimming parameter set at 0.1. Also, $TLUD(k) = \max_{m\in[m_0,m_1]} UD^k_{L(m)}$, where

$$UD^k_{L(m)} = \max_{i=1,\dots,k} \max_{\{T^i_b(m)\}_{i=1}^k\in\Lambda_\epsilon} [SSR_{L(m)} - SSR(\{T^i_b(m)\}_{i=1}^k)_{L(m)}]/\hat{V}_{L(m)}$$

with $\Lambda_\epsilon = \{(T^1_b(m), ..., T^k_b(m)) : \left|T^i_b(m) - T^{i-1}_b(m)\right| \geq \tau+\epsilon(T-2\tau+1),\, T^0_b(m) = 1, T^k_b(m) \leq$

$\tau + (1-\epsilon)(T - 2\tau + 1)\}$ and $\hat{V}_{L(m)}$ the long-run variance estimate of the total loss series using the demeaned total loss series with the mean changing at dates $T_b^1(m), ..., T_b^k(m)$. We set $k = 5$. To summarize the construction of the TL tests: 1) start with a value of the in-sample length; e.g., $m_0 = \lfloor 0.15T \rfloor$; 2) compute the total loss series $\{L_{t+\tau}\}_{t=1}^{T-2\tau+1}$; 3) for the regression with only a constant, $L_{t+\tau} = \gamma + e_{t+\tau}$, apply the sup-Wald or UDmax test for the constancy of γ with a HAC variance estimate if there is evidence of serial correlation in the losses. Store the value as $SW_{L(m)}$ or $UD_{L(m)}^k$. 4) Update m to $m+1$ and repeat Steps 2-3. Take a choice of $m_1 > m_0$, say $m_1 = \lfloor 0.85T \rfloor$, and continue up to $m = m_1$. The choice of m_0 and m_1 does not affect the asymptotic critical values (Theorem 2 below). 5) Take the maximum of the sequence of $\{SW_{L(m)}\}_{m=m_0}^{m_1}$ or $\{UD_{L(m)}^k\}_{m=m_0}^{m_1}$.

3 Asymptotic distributions of the proposed tests

This section discusses the asymptotic distributions of the proposed test statistics under the null hypothesis. We let "$\xrightarrow{p}$" denote convergence in probability and "$\Rightarrow$" denote weak convergence in distribution. We first require the following assumption. Throughout, we assume that T, m, and n go to infinity at the same rate unless otherwise stated.

Assumption 1 *Under no change in forecast accuracy: i) $\hat{\beta}_m \xrightarrow{p} \beta^*$ for all $m \in [m_0, m_1]$, with m_0 and m_1 the smallest and largest values of the in-sample lengths; ii) For $L_{t+\tau} \equiv \{L_{t+\tau}(\beta^*)\}_{t=1}^{T-2\tau+1}$, $E[L_{t+\tau}] = \mu$ for all t and $T^{-1}E[\sum_{t=1}^{\lfloor r(T-2\tau+1) \rfloor}(L_{t+\tau} - \mu)]^2 \xrightarrow{p} r\Omega$, as $T \to \infty$, for $r \in [0,1]$ with τ fixed, Ω a non-random matrix and $T^{-1/2}\sum_{t=1}^{\lfloor r(T-2\tau+1) \rfloor}(L_{t+\tau} - \mu) \Rightarrow \Omega^{1/2}W(r)$, with $W(r)$ a standard Wiener process defined on $r \in [0,1]$.*

These high level assumptions characterize the properties of the loss series under the null hypothesis. It is informative to see what they imply for the linear forecasting model $y_{t+\tau} = x_t'\beta + e_{t+\tau}$. Then Assumption 1 basically requires that β is stable over time under the null hypothesis of no change in forecast accuracy and the loss sequence satisfies a standard functional limit theorem with long-run variance Ω. Another important feature is that the loss series do not depend on m when evaluated at the limit value β^*. The relevance of this assumption is examined using the same example of a correctly specified linear model. Suppose we compute loss series using two distinct in-sample lengths m_1 and m_2. The coefficient estimates are denoted by $\hat{\beta}^1$ and $\hat{\beta}^2$, say, and the forecasting errors are $y_{t+\tau} - x_t'\hat{\beta}_1$ and $y_{t+\tau} - x_t'\hat{\beta}_2$, respectively. Under the null hypothesis, these series are asymptotically equivalent since, roughly speaking, both estimators converge to a unique limit value β^* for all m. In general, under the null hypothesis of no change in forecast accuracy, in large samples, and under a quadratic loss function, the losses are proportional to e_t^2 so that, if e_t has constant unconditional variance σ_e^2, Assumption 1 is satisfied if $T^{-1/2}\sum_{t=1}^{\lfloor r(T-2\tau+1)\rfloor}(e_{t+\tau}^2-\sigma_e^2)\Rightarrow\Omega^{1/2}W(r)$, where $\Omega = p\lim_{T\to\infty}T^{-1}E[\sum_{t=1}^{T-2\tau+1}(e_{t+\tau}^2-\sigma_e^2)]^2$ is the long-run variance of the centered values of the squared errors. This allows considerable forms of dependence in the higher moments of e_t, in particular the second moment so that conditional heteroskedasticity or serial correlation in the squared errors is allowed, in which case Ω is different from σ_e^2 and the test statistics need to be scaled by an estimate of the long-run variance of e_t^2. Hence, the conditions are quite general when the model is stable. Under Assumption 1, $SW_{L(m)}$ and $UD_{L(m)}$ have the same null limiting distribution as the sup-Wald test for a change in mean (Andrews, 1993) and the UD max test of Bai and Perron (1998), respectively. We next present the asymptotic distribution of the DSW test, whose proof is in Supplement (b).

Theorem 1 *Under Assumption 1, the limit distribution of DSW is given by*

$$DSW\Rightarrow\sup_{\mu\in[0,\bar{\mu}]}\sup_{\lambda\in[\mu+\epsilon(1-\mu),1-\epsilon(1-\mu)]}\frac{[(\lambda-\mu)W(1)+(1-\lambda)W(\mu)-(1-\mu)W(\lambda)]^2}{(1-\lambda)(1-\mu)(\lambda-\mu)},$$

as $T,m,n\to\infty$ at the same rate, where $W(r)$ is a standard Wiener process defined on $r\in[0,1]$, ϵ is the trimming parameter and $\bar{\mu}=\lim_{T\to\infty}(m_1-m_0)/n_0$, with $n_0=T-m_0-\tau+1$.

The critical values of the DSW test were tabulated using 5,000 replications with 5,000 steps to approximate the Wiener process as partial sums of $i.i.d.$ $N(0,1)$ random variables. We report results for a grid of values for $\bar{\mu}$ in the range $[0.20, 0.80]$ and we set $\epsilon=0.1$ (used throughout in the simulations and applications). The results are presented in Table 1 (see Supplement (c) for $\epsilon=0.05$ and 0.15). We next consider the limit distribution of the $TLSW$ and $TLUD$ tests. Exploiting the fact that the loss series $\{L_{t+\tau}(\beta^*)\}_{t=1}^{T-2\tau+1}$ does not asymptotically depend on m under the null hypothesis, we obtain the following theorem.

Theorem 2 *Under Assumption 1 and the null hypothesis: a) the limit distribution of $TLSW$ is the same as the sup-Wald test for a change in mean (Andrews, 1993) for any m_0 and m_1 ($1\le m_0\le m_1\le T$); b) the limit distribution of $TLUD$ is the same as the UDmax test of Bai and Perron (1998) for any m_0 and m_1 ($1\le m_0\le m_1\le T$).*

Under Assumption 1, Theorem 2 follows trivially since $\{L_{t+\tau}(\beta^*)\}_{t=1}^{T-2\tau+1}$ does not depend on m under the null hypothesis and the tests computed with different m's are asymptotically perfectly correlated. This implies no effect of taking the maximum of the statistics over m on the limiting distribution. Note also that, unlike for the DSW test, the choices of m_0 and m_1 do not affect the limiting distribution of the $TLSW$ and $TLUD$ tests.

4 Analysis of the size and power of the tests

We present simulation and theoretical results to address the following issues: 1) the finite sample size of the tests proposed (Section 4.1); 2) the power function of the tests: Section 4.2 describes the experimental design, Section 4.3 provides theoretical results useful to understand the main features of the power functions and Section 4.4 provides a summary of the main results. Section 4.5 expands on the causes of various non-monotonic power functions.

4.1 Finite sample size of the proposed tests

We first examine the size of the $TLSW$ and $TLUD$ tests using the asymptotic distribution of Theorem 2. The DGP specifies $y_t \sim i.i.d.N(0,1)$ of lengths $T = 150, 300$. We consider the squared error loss for the static model: $y_{t+\tau} = c + e_{t+\tau}$, and the dynamic model: $y_{t+\tau} = c + \alpha y_t + e_{t+\tau}$, with $\tau = 1$ both estimated by OLS and tests with or without a HAC correction for serial correlation in the losses. The HAC variance estimate is constructed using Andrew's (1991) data dependent method with an AR1 approximation and the Bartlett kernel. For all cases, we consider a fixed, rolling and recursive forecasting scheme. The exact sizes of the tests are presented in Table 2 for $T = 300$ and $\bar{\mu} = 0.25$, 0.5, and 0.75 (for $T = 150$, see Supplement (d)). The number of replications is 1,000. We label the test without the HAC variance estimate by "non-robust" and with it by "robust". We also set $\epsilon = 0.1$, $m_0 = \lfloor 0.15T \rfloor$ and $m_1 = \lfloor 0.85T \rfloor$; any reasonable variations of these choices do not change the results qualitatively. The exact size is, in general, close to the nominal size. Some distortions are present with the robust version, which decrease as T increases. The results for the size of the DSW tests are in Table 3, with the same specifications as above (for $T = 150$, see Supplement (e)). For $\bar{\mu} = 0.25$ and 0.5, the test shows little size distortions, if any, for all cases. Some liberal size distortions are present when $\bar{\mu}$ is as large as 0.75 (caused by parameter uncertainty from unreported simulations). With larger sample sizes, the distortions somewhat decrease but remain substantial for $\bar{\mu} = 0.75$. Hence, we recommend using $\bar{\mu} = 0.25$ or $\bar{\mu} = 0.5$.

4.2 The experimental design for the power analysis

In order to ensure that our simulation design is not biased in favor of the tests we propose, we adopt the same design as in GR. Note, however, that we do not set m to be equal to the date of the first break. GR mention that this corresponds to the worst case scenario from a forecasting point of view. But what is more relevant in the context of assessing the presence of forecast instabilities is the fact that it corresponds to the best case possible for the power of the tests. Hence, such a choice can distort the power properties of the tests which are relevant in practice, given that the date of the break is unknown. There are five different DGPs involving single or multiple changes in level or in variance. They are:

DGP1: (single variance shift): $y_t = \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0, \sigma_t^2)$, with $\sigma_t^2 = 1 + \beta_A I(t > T/2)$;

DGP2: (multiple variance shifts): $y_t = \varepsilon_t \sim i.i.d.N(0, \sigma_t^2)$, $\sigma_t^2 = 1 + \beta_A I(t \notin \Lambda_0)$;

DGP3: (single mean shift): $y_t = \beta_A I(t > T/2) + \varepsilon_t$, with $\varepsilon_t \sim i.i.d.N(0, 1)$;

DGP4: (multiple mean shifts): $y_t = -\beta_A + \varepsilon_t$ if $t \in \Lambda_0$, $\beta_A + \varepsilon_t$ otherwise, $\varepsilon_t \sim i.i.d.N(0, 1)$;

DGP5: (mean shifts at unequal intervals): $y_t = \beta_A I(t \leq T/4) - 3\beta_A I(T/4 < t \leq T/2) + \sqrt{\beta_A} I(t > T/2) + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0, 1)$.

We set $T = 150$ and with multiple breaks as in DGPs 2 and 4, these occur every 50 periods, i.e. $\Lambda_0 \in \{(1, 50), (101, 150)\}$. The following tests are considered with the squared error loss function: the GR_m test of GR with $m = 40$ ($GR1$) and $m = 100$ ($GR2$); the supremum of the GR_m tests, labelled SGR over $0.2T \leq m \leq 0.8T$ ($SGR1$) and $0.3T \leq m \leq 0.7T$ ($SGR2$), the Double sup-Wald test DSW with $\bar{\mu} = 0.25$ ($DSW1$) and $\bar{\mu} = 0.50$ ($DSW2$) with $m_0 = 0.2T$ for both cases; the sup-Wald test for a single break and the UDmax test for multiple breaks (up to 5) applied to the total loss series, denoted $TLSW$ and $TLUD$, with $0.15T \leq m \leq 0.85T$. Two forecasting models are used: a static model: $y_{t+\tau} = c + e_{t+\tau}$, and the dynamic one: $y_{t+\tau} = c + \alpha y_t + e_{t+\tau}$, with $\tau = 1$, both estimated by OLS, and for each, two versions with or without a HAC correction for serial correlation in the losses. In the dynamic model an irrelevant lagged dependent variable is included (i.e., $\alpha = 0$), which is completely inconsequential. We could extend the DGPs to include genuine dynamics with $\alpha \neq 0$. The qualitative features would remain the same. The non-monotonicities reported would simply be more severe (e.g., Section 4.5). The static model with no HAC correction is labelled "static, non-robust" and with a HAC correction "static, robust". The dynamic model with no HAC correction is labelled "dynamic, non-robust" and the one with a HAC correction "dynamic, robust". We consider forecasting schemes using a fixed, rolling or recursive window. The number of replications is 1,000. The results are presented in Tables 4-1 to 4-5. The foremost criterion adopted to compare the tests is whether the power function is monotonically increasing as the magnitude of the change(s) in forecast accuracy increases. We view this as an essential feature for any reasonable test. For tests with monotonically increasing power, we compare the relative power functions. We start with some theoretical results about the limit value of the expected loss function that will help understand the sources of the power differences across various forecasting schemes.

4.3 Theoretical results about the limit of the loss function

We consider the limit of the loss sequence for a single coefficient or variance break. Our results pertain to the loss sequence in large samples, which are used for all tests. The surprise losses are used for the GR_m and SGR tests but since these simply differ by subtracting the average of the in-sample losses, the dynamics of both sequences are similar. Hence, we consider only the expected values of $p\lim_{T\to\infty} L_t = L^*(r)$, defined on the unit interval $r \in [0,1]$ where $r = \lim(t/T)$. The squared loss function is applied. For simplicity, and without substantive loss of generality, we consider the single break model:

$$y_{t+\tau} = x_t\beta_t + e_{t+\tau}, \quad \text{for } t = 1, ..., T-\tau, \tag{2}$$

where $\beta_t = \beta_1$ for $t \leq [T\lambda_0]$ and $\beta_t = \beta_2$ for $t > [T\lambda_0]$. Again for simplicity, the predictor x_t is a scalar that satisfies $E(x_t^2) = \sigma_x^2$ and $E(x_t x_{t-j}) = \sigma_{xj}$. Also, $e_t \sim i.i.d.(0, \sigma_t^2)$, with $\sigma_t^2 = \sigma_1^2$ for $t \leq [T\lambda_0]$ and $\sigma_t^2 = \sigma_2^2$ for $t > [T\lambda_0]$. The in-sample length is $m = [T\lambda]$ chosen so that $\lambda \leq \lambda_0$. We consider the following two cases: a coefficient change, i.e., $\beta_1 = 0$ and $\beta_2 = \Delta_\beta$ with $\sigma_1^2 = \sigma_2^2 = \sigma^2$; a variance change, i.e., $\sigma_1^2 = \sigma^2$ and $\sigma_2^2 = \sigma^2 + \Delta_{\sigma^2}$ with $\beta_1 = \beta_2 = \beta$. Suppose we use the static regression model of $y_{t+\tau}$ on x_t to produce τ-period ahead forecasts at time t. For the out-of-sample procedure, we use the estimate of β obtained from the in-sample information, labelled as $\hat{\beta}_{[1,t]}$. We consider the three window schemes and estimate the coefficient using OLS for the sample period $[1, m-\tau]$ with the fixed scheme, $[t-m+1, t-\tau]$ with the rolling scheme, and $[1, t-\tau]$ with the recursive scheme. Hence, $\hat{\beta}_{[1,t]}$ with the same t can be different depending on the window scheme. When the static regression is used, $E[L^*(r)] = E[\lim_{T\to\infty}(y_{t+\tau} - x_t\hat{\beta}_{[1,t]})^2] = \sigma_r^2 + \sigma_x^2(\beta_r - \beta_{[0,r]}^*)^2$, where $\beta_{[0,r]}^* = p\lim_{T\to\infty}\hat{\beta}_{[1,t]}$. We also denote the limit true value of β_t and σ_t^2, defined on the interval $r \in [0,1]$, by β_r and σ_r^2. Next, we consider a dynamic regression, i.e., $y_{t+\tau} = \alpha y_t + x_t\beta_t + e_{t+\tau}$. Including y_t as a predictor while the true model is (2) is inconsequential in a stable environment since the dynamic model nests the DGP. Things are quite different when instabilities are present. For the dynamic model, with $\alpha_{[0,r]}^* = p\lim_{T\to\infty}\hat{\alpha}_{[1,t]}$, and $\beta_{[0,r]}^* = p\lim_{T\to\infty}\hat{\beta}_{[1,t]}$, $E[L^*(r)] = (1 + \alpha_{[0,r]}^{*2})\sigma_r^{*2} + \sigma_x^2[(1 - \alpha_{[0,r]}^*)\beta_r - \beta_{[0,r]}^*]^2$ (see Supplement (f)). Figures S.1-S.2 present $E[L^*(r)]$ for the variance change case using the static and dynamic regressions, respectively. The limit of the loss sequence always has a stepwise change, whose magnitude depends on Δ_{σ^2}. Hence, all tests should have power monotonically increasing in Δ_{σ^2}.

Since the expressions for the case of a coefficient change are quite complex and yields little insights per se, we only present numerical values in the text for $E[L^*(r)]$. We set $\sigma^2 = 1$, $\lambda = 0.1$ and $\lambda_0 = 0.5$ and a small $(\Delta_\beta, \Delta_{\sigma^2} = 1)$ and large break $(\Delta_\beta, \Delta_{\sigma^2} = 5)$

to investigate how the size of the break affects the shape of the loss sequence. Figures 1-2 present $E[L^*(r)]$ for the coefficient change case for the static and dynamic regressions, respectively. The upper (lower) panels report the case of a small (large) break. When the static model is used (Figure 1), we see a stepwise change in the loss sequence when the fixed scheme is used. However, the same change in coefficient translates into a spiked shape with the rolling scheme and a triangular shape with the recursive scheme. This explains why the fixed scheme is to be preferred and why using the rolling or recursive scheme induces a loss of power. Note, however, that the break magnitude only changes the height of the change in the loss sequence, not the shapes of the loss sequences. Hence, increasing the magnitude of the break size should still increase power under all schemes. Things are different when considering the dynamic model. We still have the same general shapes for the loss sequences. However, the spikes and the triangular shape for the rolling and the recursive schemes become more narrow and closer to an outlier as the break becomes larger. This explains why, when using a dynamic model, the use of a rolling or recursive scheme leads to a non-monotonic power function, i.e., the power decreases as the break magnitude increases. We shall use the insights provided by these results to explain the power differences across the tests and the forecasting schemes. We show in Supplement (g) that the main theoretical results continue to hold using an indirect forecast method. The results are also similar with multiple breaks, in which case the shapes essentially repeat themselves for each break date; see Figures S.5 and S.6 in Supplement (h), which present a "typical" realization of the loss sequences for DGPs 3-5 under the three forecasting schemes for the tests $SGR2$, $DSW2$ and $TLSW$.

4.4 Summary of the main power results

The main findings of interest can be illustrated by the results for DGPs 4-5 for a dynamic forecasting model with a correction for serial correlation in the loss sequence. Only three tests have a monotonically increasing power function: the two versions of DSW and the $TLUD$ tests using a fixed forecasting window. All the other tests have a power function that eventually decreases to zero as the magnitude of changes increases in at least one and most often many cases. One exception is the SGR test with a rolling window, whose power appears high because of large size-distortions. The distortions are reduced as m increases as in $GR2$ and $SGR2$ and the power is then decreasing to 0 as the magnitude of the change increases. Cases with tests having a power function that eventually reaches zero as the magnitude of the change(s) increases can also be found when dealing with other DGPs with shifts in the conditional mean (i.e., not DGPs 1-2, which only affects the variance) and other forecasting methods. The GR_m tests can have zero or trivial power even in the "static, non-robust" case; see DGP4 (fixed and rolling), and DGP5 (recursive).

To compare the power of the tests, we focus on $\beta_A = 0.5$, a value with power not close to one or zero. We disregard the SGR test with the rolling window given the large size-distortions. For a single break (DGPs 1 and 3) the $TLUD$ and SGR tests have equally the highest power. However, with multiple breaks (DGPs 2, 4 and 5), the $TLUD$ and DSW tests outweigh the SGR tests in all cases. In summary, the test with highest power is the $TLUD$ (for DGP1 the $TLUD$ and DSW tests have nearly the same power). The $TLUD$ test has a monotonically increasing power and also highest power for small values of the alternative. Hence, we recommend using the $TLUD$ test followed by the DSW, both with a fixed forecasting scheme. The loss in power when using the DSW test instead of the $TLUD$ test may, however, be DGP-specific since the changes involved, i.e., recurrent regimes, are those most prone to cause power problems for the DSW tests. With non-recurrent regimes, the power of the DSW tests would be closer to that of $TLUD$.

4.5 Explanations for the power properties

DGPs 1-2 are cases with single and multiple variance changes (Tables 4-1 and 4-2). For such type of instability the forecast model is unaffected by the choice of T_m or the forecasting scheme. The forecast model is still consistently estimated because the conditional mean of the variable to be forecasted is unchanged. All tests have nontrivial power in all cases. The substantial difference between $GR1$ and $GR2$ is caused solely by the choice of m, showing its importance. SGR resolves this problem by maximizing the test statistics over all permissible m and achieves a reasonably high power in all cases. All tests proposed (DSW, $TLSW$ and $TLUD$) have, overall, high power. Under DGP1, DSW has a slightly lower power than $TLSW$ and $TLUD$, while under DGP2, DSW and $TLUD$ have a higher power than $TLSW$, because the latter accounts for a single break. Also, under DGP2, the power of these tests does not reach one because of the nature of the breaks; i.e., two breaks with the first and last regimes being the same, which is the most difficult case to detect with a single break test (e.g., Bai and Perron, 2006). The problem is alleviated allowing for multiple changes so that the $TLUD$ and DSW tests are the most powerful in this setting. These observations are consistent with the theoretical results reported in Section 4.3.

We now turn to models with mean breaks so that the conditional mean of the variable to be forecasted changes, i.e., DGPs 3-5 (results reported in Tables 4-3 to 4-5). The power functions exhibit non-monotonicity or a significant power loss because of three potential sources. The first is the "robust effect" indicated with an "R" in the last row of a case with non-monotonic power. The second is the "window effect" indicated with a "W". The third cause is the "dynamic effect", indicated with a "D". The simulation results pertaining to the window and dynamic effects follow from the theoretical ones in Section 4.3. The "robust effect" is due to a failure to properly account for serial correlations in the loss sequence. As is well known, when neglected breaks are present in the losses when constructing the HAC variance estimate, they inflate the sample autocovariances and the value of the bandwidth, thereby increasing HAC variance estimates and reducing power. This is a standard problem that has been discussed at length (e.g., Vogelsang , 1999, Crainiceanu and Vogelsang, 2007, Deng and Perron, 2008, Kim and Perron, 2009, Perron and Yamamoto, 2016, Martins and Perron, 2016, Chang and Perron, 2018). The "window effect" refers to the change in the loss sequence induced by using some window that separates the in and out-of-sample data and causes a loss in power. This applies, e.g., when breaks occur in the in-sample partition so that the model is not consistently estimated. The "dynamic effect" is the most pronounced and caused by in-sample contaminations when using a dynamic model. It is well known that if a dynamic model is estimated in the presence of mean breaks the coefficient estimate for the lagged dependent variable is biased toward one as the break magnitude becomes larger (Perron, 1989, 1990, 2019). This results in forecast errors being roughly the first-differences of those from a static model. Hence, the mean breaks become outliers in the loss sequence and the tests have no power. Note that the "dynamic effect" will not occur with the fixed scheme if m is sufficiently small and there is a chance that the model can be consistently estimated in a stable in-sample window. See the theoretical results in Section 4.3.

We now explain some power functions reported in Tables 4-3 to 4-5. First, as shown in panel (a), with the static model and the fixed scheme, all non-robust tests have a monotonic power. However, the "robust effect" applies to tests constructed with a HAC variance estimate in panel (b), which is pronounced for $GR1$ and $GR2$ and also applies to $SGR1$, $SGR2$ and $TLSW$ when multiple breaks are present. With the fixed scheme, the "robust effect" applies to $GR1$ under DGPs 3 and 5, and to $GR1$, $GR2$, $SGR1$, $SGR2$ and $TLSW$ under DGP4. With either the rolling or recursive scheme, it applies to $GR2$ and $DSW1$ under DGP3, to $GR1$, $GR2$, $SGR1$, $SGR2$, $DSW1$ and $DSW2$ under DGP4 and to $GR1$, $GR2$ and $SGR2$ under DGP5. The results suggest not using GR and SGR with a HAC variance estimate under any forecasting scheme nor $TLSW$ with multiple breaks. Note that even in panel (a) the power of $GR1$ and $GR2$ is sometimes very low. $GR1$ has non-monotonic power with the rolling scheme under DGP3, with both the fixed and the rolling schemes under DGP4 and with the recursive scheme under DGP5, while $GR2$ does so with the recursive scheme under DGP5. This is because they are affected by the "window effect" when a break occurs in the in-sample window. With a fixed scheme, the "window effect" applies to $GR2$ under DGP4. With the rolling scheme, it applies to $GR2$ under DGPs 3-4 and with the recursive scheme to $GR1$ under DGP5. Also, DSW may lose power with the rolling or recursive scheme because the loss sequence takes a triangular shape (Figure 1). This applies to $DSW1$ and $DSW2$ under DGP4. For DSW, the "window effect" can be exacerbated by the "robust effect" (labelled "R, W" in panel (b)); cf. $DSW1$ with the rolling scheme under DGP3 and $DSW1$ and $DSW2$ with the rolling and recursive schemes under DGP4. The source of the "window effect" can be explained by the results in Figure 1. With the fixed scheme, the loss sequence takes a step-wise pattern for all three DGPs and tests. With the rolling or recursive scheme, it shows an abrupt increase followed by a gradual decline. For DGP-3, the increase occurs when the in-sample window covers a stable period and the initial date of the out-of-sample period coincides with the true break date. After, the window increasingly contains post-break data, which gradually causes a bias in the estimated forecast model and thus a decline of the loss sequence. More importantly, the shape is robust with either the static or dynamic model, except when the first break is included in the in-sample window (e.g., $SGR2$ under DGP5 with the dynamic model).

The non-monotonic power functions are more pronounced with a dynamic model. Panel (c) of Tables 4-3 to 4-5 (DGPs 3-5) report power functions with non-robust standard errors. With the rolling or recursive schemes, this "dynamic effect" applies to almost all tests, even when the fixed scheme is used if the in-sample window is large and includes the break date; e.g., for $GR2$ under DGPs 3-4 and for $GR1$, $GR2$ and $SGR2$ under DGP5. This suggests not to use any tests with a rolling or recursive scheme when the forecast model has lagged dependent variables. The results for the dynamic model with the robust tests are presented in panel (d) of Tables 4-3 to 4-5, which highlight all tests with a non-monotonic power function. The results are consistent with the limit of the power losses in Figure 2, which showed that coefficient breaks are transformed into spikes in the loss sequence, getting narrower as the magnitude of the change increases when using the rolling or recursive scheme.

5 Empirical applications

We now provide applications to illustrate the ability of the proposed and existing tests to detect changes in forecast accuracy; one related to the equity premium and the other to forecasting inflation. The results show the relevance of the theoretical and simulation results.

5.1 Equity premium forecasts

The equity premium of the S&P 500 returns is constructed as the difference between the stock yield and the risk free rate, following Jagannathan et al. (2000). For the stock yield, we use the dividend price ratio plus the expected future growth rate of dividends. We use the historical average of the annual growth rates of dividends since 1928:01 to proxy for the latter. For the risk free rate, we use the 10-years treasury constant maturity rate. The data were obtained from Robert Shiller's web site (http://www.econ.yale.edu/~shiller/data.htm). Figure 3 plots the resulting equity premium from 1980:1 to 2017:12. Jagannathan et al. (2000) and Lettau et al. (2008) noted that it was in an unprecedented low level in the 1990s and the early 2000s. The plot shows a sharp increase around 2007 because of the financial crisis. Afterwards, the equity premium stayed high likely because of the zero interest rate monetary policy. Lettau and Nieuwerburgh (2008) and Xu and Perron (2018) documented the presence of level shifts and structural breaks in excess return prediction regressions.

We use a static, regressing $y_{t+\tau}$ on x_t and a dynamic model, regressing $y_{t+\tau}$ on y_t and x_t (both with a constant) to produce τ-period ahead forecasts. We use the dividend price ratio as the predictor x_t, as is commonly done; e.g., Campbell and Shiller (1988) and Fama and French (1988). We consider forecast horizons $\tau = 1$, 3, 6 and 12, and apply the following tests with the quadratic loss function: the DSW tests ($DSW1$ and $DSW2$ for $\bar{\mu} = 0.25$ and 0.5), the $TLSW$ and $TLUD$ tests. The truncation $\varepsilon = 0.1$ and the maximum number of breaks is 5 for the $TLUD$ test. We also apply the GR_m tests with $m = 331$ and 347, which correspond to the outsets of the global financial crisis (2007:7) and the month prior to the initiation of the zero interest rate monetary policy (2008:11). We consider the SGR tests for $0.2T \leq m \leq 0.8T$ ($SGR1$) and for $0.3T \leq m \leq 0.7T$ ($SGR2$). The results are presented in Table 5. We first test for the presence of serial correlation in the loss sequence using the LM test of Godfrey (1978) whose results are presented in the columns labelled "$LM1$" for $m = 331$ and "$LM2$" for $m = 347$. They strongly indicate the presence of serial correlation in the loss sequences. Hence, the tests account for serial correlations using a HAC variance estimate. With the static model, there is evidence for a forecast breakdown with any of the proposed tests and the p-values when using the fixed scheme are an order of magnitude smaller compared to using the rolling or recursive scheme. Some GR-based tests also reject in all cases when the rolling window scheme is used but not with the fixed or recursive one. This is likely due to high power induced by large size distortions as documented above. The GR-based tests also show less rejections when long horizons are considered under the fixed and recursive schemes. With the dynamic model, we obtain a much clearer contrast.

The GR_m tests now show very few rejections and the tests proposed strongly reject the null hypothesis for all horizons with the fixed window. As expected from the theoretical results, the DSW test fails to reject with the rolling or recursive scheme, while the $TLSW$ and $TLUD$ tests still show rejections, although much weaker in terms of p-values.

5.2 Inflation forecasts using the Phillips curve

Forecasting inflation using the Phillips Curve was advanced by, e.g., Stock and Watson (1999) and Atkeson and Ohanian (2001), among many others. We use the model:

$$\pi_{t+\tau} = \theta_0 + \theta_1(L)u_t + \theta_2(L)\pi_t + error,$$

where π_t is a measure of inflation, u_t is the unemployment gap (the unemployment rate minus and a measure of the NAIRU). The order of the lag polynomial $\theta_1(L)$ was set to $q_u = 1$ or $q_u = 3$. Since the results are similar, we only report the case with $q_u = 3$. For the order of $\theta_2(L)$ we report results for both $q_\pi = 1$ and 3, labelled 'Dynamic 1' and 'Dynamic 3', respectively. We consider two window sizes for the GR_m tests: 1) $GR1$ with $m = 241$, as in GR, so that the period before 1979 is within the in-sample and the high-inflation period of Volker's Fed Chairmanship is in the out-of-sample; 2) $GR2$ with $m = 301$ (1984:1), in which case the out-of-sample window covers the Great Moderation. Orphanides and van Norden (2005) find that Phillips curve-based forecasts outperform an autoregressive benchmark prior to 1983 but without improvement for the period after 1984, while Dotsey et al. (2018) find that while the Phillips curve forecasts improve when the economy is weak, the improvement vanishes in the post-1984 period. In either case, the presumption is that a change in forecasting performance occurred. We consider the forecast horizons $\tau = 1, 3, 6$ and 12 months. We use the monthly real-time CPI (consumer price index) and unemployment gap data for the period 1959:1 to 2018:7. The data were obtained from the Federal Reserve Bank of Philadelphia Real-Time Data Set in which the CPI is available only after 1998:11 and the unemployment rate after 1965:11. The earlier data were obtained from the Swanson, van Dijk and Callan dataset (http://econweb.rutgers.edu/nswanson/realtime.htm). Since the data from two sources are very similar in the overlapping period, this merging should not affect the results. Also, as in GR we assume that a time-invariant NAIRU is embodied in the intercept. We construct the annual rate of inflation as $\pi_t = (1200)\ln(P_t/P_{t-1})$, where P_t is the CPI at month t, whose graph is in Figure 4.

The results are presented in Table 6 for the same tests as in Section 5.1. The LM tests for serial correlation in the loss function (Table 6; last two columns) indicate the presence of serial correlation, so the tests are constructed with a HAC correction. For "Dynamic 1", all proposed tests show strong rejections at any horizon when the fixed window scheme is used. With the rolling or recursive scheme, the DSW tests fail to reject in most cases, consistent with our theoretical results about the non-monotonic power with a rolling or recursive scheme. The results suggest multiple breaks since the value of $TLUD$ is much larger than $TLSW$, although both tests reject the null hypothesis at the 1% significance level. More interestingly, all the GR-based tests, except for the $SGR1$ with the fixed and the rolling schemes, have no power. The results are qualitatively the same for "Dynamic 3". Again, all the GR-based tests have no power, except for $SGR1$ with the fixed and the rolling schemes. In summary, the GR-based tests find no or very weak evidence of a change in forecast accuracy in the high inflation period of the Volker's Chairmanship period. They are also not able to detect a change due to the Great Moderation that occurred in the mid-1980s. In contrast, the proposed DSW, $TLSW$ and $TLUD$ tests clearly show evidence of changes in forecast accuracy when the fixed scheme is used.

6 Conclusion

We considered the issue of forecast failure (or breakdown) and proposed methods to detect changes in the forecasting performance over time. The aim is to assess retrospectively whether a given forecasting model provides forecasts which show evidence of changes (improvements or deterioration) with respect to some loss function. We adapted the classical structural change tests to the forecast failure context. First, we recommend that all tests should be carried with a fixed scheme to have best power. We considered a variety of tests: the GR test (a t-test for a change at some pre-specified date m); maximizing the GR_m test over all values of m within a pre-specified range; a Double sup-Wald test which for each m performs a sup-Wald test for a change in the mean of the out-of-sample losses and takes the maximum of such tests over some range; we also proposed to work directly with the total loss series to define the $TLSW$ and the $TLUD$ tests. The only tests having a monotonic power function for all data-generating processes are the DSW and $TLUD$ tests, constructed with a fixed forecasting window scheme. The power of the $TLUD$ test is usually higher than that of the DSW test, hence it is recommended for practical applications.

References

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

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

Atkeson, A. and Ohanian, L.E. (2001), "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review* 25(1), 2-11.

Bai, J. and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

Bai, J. and Perron, P. (2006), "Multiple Structural Change Models: A Simulation Analysis," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, eds. D. Corbae, S. Durlauf and B. E. Hansen, Cambridge: Cambridge University Press, pp. 212-237.

Campbell, J.Y. and Shiller, R.J. (1988), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195-228.

Casini, A. (2017), "Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework," Unpublished Manuscript, Department of Economics, Boston University.

Casini, A. and Perron, P. (2018), "Structural Breaks in Time Series," forthcoming in the *Oxford Research Encyclopedia of Economics and Finance*.

Chang, S.Y. and Perron, P. (2018), "A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models," *Econometric Reviews*, 37, 577-601.

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

Clark, T.E. and McCracken, M.W. (2013), "Advances in Forecast Evaluation". In *Handbook of Economic Forecasting,* Vol. 2, eds. G. Elliott and A. Timmermann. Amsterdam: Elsevier, pp. 1107-1201.

Crainiceanu, C. M. and Vogelsang, T.J. (2007), "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77, 457-476.

Deng, A. and Perron, P. (2008), "A Non-local Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change," *Journal of Econometrics*, 142, 212-240.

Dotsey, M., Fujita, S. and Stark, T. (2018), "Do Phillips Curves Conditionally Help to Forecast Inflation?" *International Journal of Central Banking*, 14, 43-92.

Fama, E.F. and French, K.R. (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3-25.

Giacomini, R. and Rossi, B. (2009), "Detecting and Predicting Forecast Breakdowns," *The Review of Economic Studies*, 76, 669-705.

Godfrey, L.G. (1978), "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables." *Econometrica* 46, 1303-1310.

Inoue, A. and Rossi, B. (2012), "Out-of-sample Forecast Test Robust to the Window Size Choice," *Journal of Business and Economic Statistics*, 30, 432-453.

Jagannathan, R., McGrattan, E.R., and Scherbina, A. (2000), "The Declining U.S. Equity Premium," *Federal Reserve Bank of Minneapolis Quarterly Review*, 24(4), 3-19.

Kim, D. and Perron, P. (2009), "Assessing the Relative Power of Structural Break Tests using a Framework Based on the Approximate Bahadur Slope," *Journal of Econometrics*, 149, 26-51.

Lettau, M., Ludvigson, S.C., and Wachter, J.A. (2008), "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?" *Review of Financial Studies*, 21, 1653-1687.

Lettau, M. and Van Nieuwerburgh, S., (2008) "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607-1652.

Martins, L.F. and Perron, P. (2016), "Improved Tests for Forecast Comparisons in the Presence of Instabilities," *Journal of Time Series Analysis*, 37, 650-659.

Orphanides, A., and van Norden, S. (2005), "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time," *Journal of Money, Credit and Banking*, 37, 583-600.

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

Perron, P. (1990), "Testing for a Unit Root in a Time Series Regression with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153-162.

Perron, P. (2006), "Dealing with Structural Breaks". In *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, eds. K. Patterson and T.C. Mills, New York: Palgrave Macmillan, pp. 278-352.

Perron, P. (2019), "A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series," forthcoming in *Time Series Econometrics: Volume 2: Structural Change*, ed. P. Perron, World Scientific.

Perron, P. and Yamamoto, Y. (2016), "On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests," *Econometric Reviews*, 35, 782-844.

Qu, Z. (2007), "Searching for Cointegration in a Dynamic System," *The Econometrics Journal*, 10, 580-604.

Stock, J.H., and Watson, M.W. (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

Vogelsang. T.J. (1999), "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88, 283-299.

Xu, J. and Perron, P. (2018), "Forecasting in the Presence of in and Out of Sample Breaks," Unpublished Manuscript, Department of Economics, Boston University.

Table 1: Critical values of the DSW test ($\epsilon = 0.1$)

$\bar{\mu}$	10%	5%	2.5%	1%
0.20	10.609	12.217	13.779	15.620
0.25	10.928	12.782	14.018	16.310
0.30	11.264	13.065	15.087	17.688
0.35	11.648	13.529	15.247	17.660
0.40	11.761	13.770	15.537	17.777
0.45	12.134	14.027	15.768	17.968
0.50	12.469	14.279	16.031	17.961
0.55	12.932	14.565	16.184	18.455
0.60	13.103	14.850	16.512	18.562
0.65	13.367	15.003	16.654	19.027
0.70	13.596	15.181	16.622	19.103
0.75	13.769	15.418	17.075	19.130
0.80	14.108	15.870	17.736	19.968

Table 2: Size of the $TLSW$ and $TLUD$ tests ($T = 300$)

a) static, non-robust

		10%	5%	1%
TLSW	fixed	0.112	0.061	0.02
	rolling	0.117	0.063	0.019
	recursive	0.096	0.058	0.016
TLUD	fixed	0.125	0.067	0.019
	rolling	0.127	0.070	0.019
	recursive	0.107	0.061	0.018

c) dynamic, non-robust

		10%	5%	1%
TLSW	fixed	0.109	0.057	0.013
	rolling	0.120	0.057	0.013
	recursive	0.092	0.047	0.011
TLUD	fixed	0.125	0.067	0.013
	rolling	0.127	0.071	0.013
	recursive	0.105	0.055	0.010

b) static, robust

		10%	5%	1%
TLSW	fixed	0.146	0.076	0.021
	rolling	0.144	0.060	0.012
	recursive	0.096	0.058	0.016
TLUD	fixed	0.167	0.085	0.021
	rolling	0.148	0.079	0.011
	recursive	0.107	0.061	0.018

d) dynamic, robust

		10%	5%	1%
TLSW	fixed	0.142	0.072	0.019
	rolling	0.146	0.072	0.022
	recursive	0.099	0.063	0.013
TLUD	fixed	0.188	0.101	0.023
	rolling	0.186	0.102	0.014
	recursive	0.112	0.081	0.016

Table 3: Size of the DSW test ($T = 300$)

a) static, non-robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.075	0.080	0.071	0.043	0.050	0.036	0.018	0.019	0.017
	rolling	0.083	0.067	0.049	0.046	0.038	0.021	0.020	0.019	0.007
	recursive	0.075	0.082	0.085	0.039	0.047	0.045	0.015	0.017	0.014
0.5	fixed	0.104	0.097	0.093	0.064	0.056	0.058	0.019	0.023	0.025
	rolling	0.116	0.102	0.078	0.067	0.061	0.048	0.023	0.021	0.018
	recursive	0.103	0.098	0.095	0.061	0.058	0.057	0.022	0.024	0.022
0.75	fixed	0.169	0.167	0.142	0.117	0.114	0.088	0.047	0.045	0.040
	rolling	0.164	0.155	0.124	0.122	0.114	0.087	0.051	0.047	0.034
	recursive	0.172	0.173	0.147	0.113	0.108	0.087	0.050	0.047	0.042

b) static, robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.090	0.088	0.079	0.055	0.059	0.049	0.018	0.021	0.015
	rolling	0.096	0.081	0.054	0.054	0.042	0.022	0.020	0.018	0.009
	recursive	0.092	0.090	0.081	0.047	0.049	0.044	0.016	0.016	0.014
0.5	fixed	0.121	0.110	0.112	0.078	0.071	0.071	0.033	0.032	0.029
	rolling	0.135	0.114	0.096	0.089	0.075	0.054	0.035	0.028	0.025
	recursive	0.126	0.117	0.111	0.078	0.068	0.072	0.035	0.030	0.027
0.75	fixed	0.213	0.210	0.173	0.157	0.150	0.121	0.080	0.073	0.059
	rolling	0.210	0.195	0.154	0.167	0.152	0.101	0.082	0.073	0.062
	recursive	0.210	0.205	0.174	0.158	0.151	0.120	0.082	0.074	0.062

c) dynamic, non-robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.087	0.090	0.088	0.052	0.059	0.053	0.018	0.027	0.022
	rolling	0.091	0.069	0.054	0.047	0.036	0.022	0.018	0.018	0.007
	recursive	0.086	0.086	0.089	0.044	0.047	0.042	0.017	0.021	0.018
0.5	fixed	0.107	0.103	0.110	0.065	0.063	0.069	0.028	0.030	0.027
	rolling	0.116	0.111	0.081	0.071	0.065	0.049	0.029	0.024	0.019
	recursive	0.108	0.106	0.105	0.070	0.067	0.068	0.029	0.031	0.026
0.75	fixed	0.175	0.175	0.155	0.115	0.113	0.103	0.047	0.046	0.050
	rolling	0.178	0.169	0.127	0.126	0.119	0.089	0.054	0.052	0.036
	recursive	0.177	0.175	0.151	0.128	0.125	0.102	0.050	0.050	0.049

d) dynamic, robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.105	0.093	0.081	0.058	0.060	0.051	0.019	0.024	0.018
	rolling	0.103	0.077	0.054	0.052	0.039	0.022	0.019	0.020	0.008
	recursive	0.104	0.095	0.091	0.051	0.050	0.046	0.017	0.021	0.020
0.5	fixed	0.127	0.119	0.119	0.081	0.076	0.077	0.032	0.033	0.034
	rolling	0.138	0.126	0.094	0.089	0.075	0.050	0.037	0.031	0.026
	recursive	0.128	0.128	0.115	0.087	0.078	0.074	0.036	0.033	0.035
0.75	fixed	0.221	0.215	0.180	0.152	0.145	0.125	0.080	0.073	0.070
	rolling	0.222	0.204	0.154	0.171	0.160	0.099	0.083	0.075	0.062
	recursive	0.219	0.211	0.178	0.164	0.160	0.122	0.088	0.082	0.068

Table 4-1: Power comparison under DGP1 (5% level)

a) static, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.091	0.063	0.127	0.089	0.047	0.071	0.062	0.064
	0.5	0.982	0.766	0.997	0.997	0.598	0.480	0.988	0.985
	1.0	1.000	0.908	1.000	1.000	0.892	0.821	1.000	1.000
	2.5	1.000	0.963	1.000	1.000	0.974	0.942	1.000	1.000
	5.0	1.000	0.975	1.000	1.000	0.986	0.971	1.000	1.000
	7.5	1.000	0.974	1.000	1.000	0.988	0.975	1.000	1.000
	10.0	1.000	0.976	1.000	1.000	0.988	0.973	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.238	0.058	0.364	0.105	0.052	0.069	0.064	0.072
	0.5	0.991	0.598	0.988	0.993	0.637	0.512	0.991	0.991
	1.0	1.000	0.764	1.000	1.000	0.907	0.836	1.000	1.000
	2.5	1.000	0.845	1.000	1.000	0.979	0.953	1.000	1.000
	5.0	1.000	0.878	1.000	1.000	0.990	0.977	1.000	1.000
	7.5	1.000	0.870	1.000	1.000	0.991	0.977	1.000	1.000
	10.0	1.000	0.896	1.000	1.000	0.989	0.974	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.072	0.060	0.074	0.058	0.055	0.075	0.051	0.057
	0.5	0.998	0.763	0.994	0.995	0.608	0.491	0.986	0.986
	1.0	1.000	0.910	1.000	1.000	0.898	0.824	1.000	0.999
	2.5	1.000	0.965	1.000	1.000	0.974	0.944	1.000	1.000
	5.0	1.000	0.975	1.000	1.000	0.986	0.974	1.000	1.000
	7.5	1.000	0.971	1.000	1.000	0.988	0.976	1.000	1.000
	10.0	1.000	0.978	1.000	1.000	0.988	0.969	1.000	1.000

c) dynamic, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.128	0.059	0.177	0.117	0.074	0.090	0.093	0.114
	0.5	0.987	0.813	0.999	0.999	0.623	0.495	0.992	0.992
	1.0	1.000	0.943	0.999	0.999	0.900	0.824	1.000	1.000
	2.5	1.000	0.971	1.000	1.000	0.972	0.954	1.000	1.000
	5.0	1.000	0.988	1.000	1.000	0.989	0.975	1.000	1.000
	7.5	1.000	0.976	1.000	1.000	0.993	0.976	1.000	1.000
	10.0	1.000	0.976	1.000	1.000	0.986	0.973	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.511	0.047	0.734	0.240	0.064	0.080	0.082	0.096
	0.5	0.999	0.666	0.997	0.996	0.634	0.534	0.994	0.997
	1.0	1.000	0.811	0.999	0.999	0.905	0.842	1.000	1.000
	2.5	1.000	0.887	1.000	1.000	0.968	0.943	1.000	1.000
	5.0	1.000	0.914	0.999	1.000	0.976	0.964	1.000	1.000
	7.5	1.000	0.904	1.000	1.000	0.974	0.962	1.000	1.000
	10.0	1.000	0.923	1.000	1.000	0.967	0.951	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.102	0.056	0.090	0.073	0.068	0.087	0.055	0.063
	0.5	0.998	0.820	0.996	0.999	0.596	0.492	0.988	0.986
	1.0	1.000	0.942	0.999	0.999	0.896	0.833	0.999	0.999
	2.5	1.000	0.970	1.000	1.000	0.967	0.943	1.000	1.000
	5.0	1.000	0.986	1.000	1.000	0.983	0.972	1.000	1.000
	7.5	1.000	0.975	1.000	1.000	0.981	0.973	1.000	1.000
	10.0	1.000	0.974	1.000	1.000	0.982	0.968	1.000	1.000

b) static, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.090	0.061	0.111	0.096	0.083	0.109	0.069	0.081
	0.5	0.965	0.758	0.992	0.997	0.602	0.527	0.986	0.987
	1.0	1.000	0.894	1.000	1.000	0.884	0.805	1.000	1.000
	2.5	1.000	0.944	1.000	1.000	0.963	0.942	1.000	1.000
	5.0	1.000	0.965	1.000	1.000	0.977	0.956	1.000	1.000
	7.5	1.000	0.949	1.000	1.000	0.980	0.954	1.000	1.000
	10.0	1.000	0.968	1.000	1.000	0.981	0.968	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.261	0.054	0.367	0.129	0.077	0.103	0.072	0.088
	0.5	0.975	0.583	0.971	0.976	0.624	0.555	0.988	0.989
	1.0	1.000	0.735	0.997	1.000	0.884	0.816	1.000	1.000
	2.5	1.000	0.805	0.998	0.999	0.965	0.945	1.000	1.000
	5.0	1.000	0.841	0.999	0.999	0.980	0.962	1.000	1.000
	7.5	1.000	0.832	1.000	1.000	0.982	0.958	1.000	1.000
	10.0	1.000	0.860	1.000	1.000	0.982	0.966	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.085	0.060	0.069	0.064	0.079	0.115	0.061	0.071
	0.5	0.994	0.752	0.980	0.989	0.614	0.529	0.986	0.985
	1.0	1.000	0.889	1.000	1.000	0.880	0.804	0.999	0.999
	2.5	1.000	0.945	1.000	1.000	0.965	0.942	1.000	1.000
	5.0	1.000	0.965	1.000	1.000	0.980	0.959	1.000	1.000
	7.5	1.000	0.946	1.000	1.000	0.979	0.956	1.000	1.000
	10.0	1.000	0.970	1.000	1.000	0.982	0.968	1.000	1.000

d) dynamic, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.129	0.049	0.138	0.096	0.085	0.119	0.088	0.111
	0.5	0.987	0.762	0.998	0.999	0.581	0.492	0.989	0.991
	1.0	1.000	0.905	1.000	1.000	0.871	0.785	1.000	1.000
	2.5	1.000	0.954	1.000	1.000	0.955	0.926	1.000	1.000
	5.0	1.000	0.967	1.000	1.000	0.982	0.948	1.000	1.000
	7.5	1.000	0.976	1.000	1.000	0.968	0.945	1.000	1.000
	10.0	1.000	0.978	1.000	1.000	0.974	0.958	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.528	0.052	0.754	0.229	0.075	0.108	0.089	0.109
	0.5	0.998	0.630	0.992	0.990	0.604	0.528	0.991	0.993
	1.0	1.000	0.776	0.998	0.998	0.855	0.794	0.999	0.999
	2.5	1.000	0.839	1.000	1.000	0.941	0.908	1.000	1.000
	5.0	1.000	0.871	1.000	1.000	0.948	0.920	1.000	1.000
	7.5	1.000	0.886	1.000	1.000	0.937	0.917	1.000	1.000
	10.0	1.000	0.898	1.000	1.000	0.930	0.916	1.000	1.000
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.103	0.047	0.063	0.049	0.085	0.119	0.066	0.080
	0.5	0.999	0.763	0.993	0.997	0.571	0.501	0.986	0.986
	1.0	1.000	0.906	1.000	1.000	0.861	0.796	0.999	0.999
	2.5	1.000	0.960	1.000	1.000	0.952	0.923	1.000	1.000
	5.0	1.000	0.970	1.000	1.000	0.964	0.952	1.000	1.000
	7.5	1.000	0.975	1.000	1.000	0.949	0.942	1.000	1.000
	10.0	1.000	0.979	1.000	1.000	0.959	0.937	1.000	1.000

Table 4-2: Power comparison under DGP2 (5% level)

a) static, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.091	0.063	0.127	0.089	0.047	0.071	0.062	0.064
	0.5	0.135	0.375	0.229	0.263	0.780	0.675	0.512	0.626
	1.0	0.220	0.478	0.355	0.403	0.951	0.914	0.697	0.870
	2.5	0.286	0.590	0.519	0.567	0.988	0.974	0.771	0.961
	5.0	0.311	0.622	0.557	0.629	0.993	0.982	0.791	0.966
	7.5	0.324	0.623	0.572	0.647	0.993	0.982	0.838	0.983
	10.0	0.301	0.677	0.549	0.628	0.990	0.988	0.818	0.986
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.238	0.058	0.364	0.105	0.052	0.069	0.064	0.072
	0.5	0.157	0.649	0.413	0.409	0.778	0.684	0.516	0.637
	1.0	0.112	0.833	0.548	0.580	0.952	0.905	0.691	0.887
	2.5	0.083	0.886	0.638	0.686	0.988	0.976	0.781	0.968
	5.0	0.096	0.916	0.696	0.743	0.995	0.987	0.799	0.968
	7.5	0.091	0.907	0.696	0.740	0.998	0.988	0.841	0.985
	10.0	0.099	0.929	0.720	0.768	0.994	0.987	0.832	0.988
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.072	0.060	0.074	0.058	0.055	0.075	0.051	0.057
	0.5	0.063	0.383	0.146	0.193	0.786	0.675	0.480	0.602
	1.0	0.085	0.483	0.218	0.261	0.958	0.915	0.663	0.869
	2.5	0.110	0.574	0.321	0.359	0.989	0.976	0.739	0.961
	5.0	0.095	0.617	0.358	0.393	0.994	0.989	0.770	0.965
	7.5	0.117	0.620	0.356	0.419	0.998	0.987	0.814	0.982
	10.0	0.093	0.664	0.342	0.394	0.991	0.987	0.796	0.985

c) dynamic, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.128	0.059	0.177	0.117	0.074	0.090	0.093	0.114
	0.5	0.120	0.413	0.233	0.281	0.788	0.696	0.567	0.670
	1.0	0.179	0.545	0.356	0.415	0.954	0.916	0.749	0.890
	2.5	0.210	0.642	0.512	0.570	0.988	0.971	0.823	0.970
	5.0	0.254	0.669	0.569	0.639	0.994	0.984	0.830	0.972
	7.5	0.254	0.705	0.564	0.629	0.994	0.983	0.869	0.991
	10.0	0.247	0.701	0.583	0.662	0.990	0.987	0.864	0.991
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.511	0.047	0.734	0.240	0.064	0.080	0.082	0.096
	0.5	0.317	0.692	0.590	0.460	0.784	0.694	0.572	0.680
	1.0	0.298	0.847	0.689	0.646	0.943	0.905	0.751	0.906
	2.5	0.319	0.921	0.783	0.762	0.988	0.976	0.832	0.975
	5.0	0.304	0.941	0.814	0.791	0.994	0.986	0.861	0.977
	7.5	0.341	0.946	0.835	0.827	0.994	0.983	0.899	0.993
	10.0	0.331	0.952	0.830	0.816	0.992	0.987	0.887	0.993
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.102	0.056	0.090	0.073	0.068	0.087	0.055	0.063
	0.5	0.062	0.422	0.166	0.201	0.796	0.679	0.499	0.629
	1.0	0.066	0.552	0.265	0.307	0.955	0.916	0.675	0.872
	2.5	0.068	0.644	0.370	0.415	0.987	0.975	0.752	0.962
	5.0	0.070	0.673	0.376	0.443	0.995	0.984	0.788	0.968
	7.5	0.077	0.705	0.409	0.451	0.996	0.985	0.830	0.988
	10.0	0.073	0.716	0.406	0.447	0.991	0.984	0.814	0.986

b) static, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.090	0.061	0.111	0.096	0.083	0.109	0.069	0.081
	0.5	0.133	0.376	0.238	0.278	0.790	0.703	0.488	0.669
	1.0	0.192	0.468	0.313	0.376	0.940	0.881	0.639	0.888
	2.5	0.233	0.574	0.426	0.493	0.983	0.961	0.688	0.961
	5.0	0.256	0.590	0.451	0.526	0.991	0.975	0.701	0.961
	7.5	0.288	0.606	0.479	0.566	0.990	0.970	0.753	0.978
	10.0	0.278	0.620	0.497	0.572	0.986	0.977	0.723	0.979
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.261	0.054	0.367	0.129	0.077	0.103	0.072	0.088
	0.5	0.122	0.638	0.374	0.378	0.794	0.706	0.498	0.677
	1.0	0.100	0.790	0.484	0.522	0.939	0.889	0.639	0.897
	2.5	0.101	0.863	0.600	0.633	0.983	0.961	0.688	0.963
	5.0	0.081	0.897	0.630	0.671	0.992	0.978	0.707	0.966
	7.5	0.101	0.891	0.654	0.697	0.991	0.980	0.751	0.980
	10.0	0.076	0.910	0.654	0.704	0.990	0.977	0.729	0.984
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.085	0.060	0.069	0.064	0.079	0.115	0.061	0.071
	0.5	0.065	0.376	0.159	0.182	0.798	0.705	0.458	0.650
	1.0	0.072	0.471	0.186	0.239	0.944	0.895	0.609	0.880
	2.5	0.073	0.573	0.261	0.321	0.984	0.962	0.658	0.958
	5.0	0.082	0.590	0.295	0.355	0.993	0.979	0.662	0.959
	7.5	0.091	0.602	0.313	0.377	0.992	0.975	0.718	0.979
	10.0	0.088	0.618	0.321	0.367	0.987	0.979	0.693	0.980

d) dynamic, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.129	0.049	0.138	0.096	0.085	0.119	0.088	0.111
	0.5	0.098	0.385	0.218	0.255	0.788	0.695	0.541	0.705
	1.0	0.153	0.499	0.343	0.383	0.929	0.880	0.681	0.901
	2.5	0.206	0.580	0.448	0.503	0.976	0.954	0.722	0.966
	5.0	0.242	0.656	0.508	0.577	0.989	0.967	0.737	0.966
	7.5	0.234	0.640	0.500	0.562	0.987	0.972	0.796	0.985
	10.0	0.232	0.644	0.504	0.572	0.983	0.972	0.765	0.982
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.528	0.052	0.754	0.229	0.075	0.108	0.089	0.109
	0.5	0.307	0.672	0.598	0.433	0.782	0.683	0.540	0.711
	1.0	0.285	0.820	0.639	0.559	0.915	0.874	0.685	0.908
	2.5	0.283	0.892	0.710	0.672	0.967	0.943	0.740	0.974
	5.0	0.284	0.910	0.776	0.738	0.978	0.956	0.770	0.974
	7.5	0.324	0.927	0.770	0.743	0.977	0.961	0.815	0.989
	10.0	0.354	0.924	0.791	0.750	0.977	0.961	0.786	0.989
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.103	0.047	0.063	0.049	0.085	0.119	0.066	0.080
	0.5	0.047	0.383	0.160	0.182	0.786	0.692	0.485	0.673
	1.0	0.060	0.501	0.253	0.292	0.940	0.885	0.615	0.885
	2.5	0.070	0.592	0.301	0.357	0.978	0.961	0.675	0.963
	5.0	0.067	0.662	0.350	0.416	0.989	0.973	0.696	0.967
	7.5	0.065	0.637	0.346	0.392	0.988	0.978	0.739	0.982
	10.0	0.064	0.640	0.362	0.406	0.985	0.975	0.712	0.983

Table 4-3: Power comparison under DGP3 (5% level)

a) static, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.091	0.063	0.127	0.089	0.047	0.071	0.062	0.064
	0.5	0.189	0.073	0.264	0.219	0.059	0.073	0.229	0.245
	1.0	0.681	0.218	0.839	0.828	0.303	0.247	0.868	0.863
	2.5	1.000	0.963	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	-	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.238	0.058	0.364	0.105	0.052	0.069	0.064	0.072
	0.5	0.323	0.053	0.478	0.168	0.048	0.097	0.082	0.107
	1.0	0.580	0.055	0.724	0.481	0.038	0.216	0.132	0.284
	2.5	0.998	0.099	1.000	0.999	0.649	0.961	1.000	1.000
	5.0	1.000	0.179	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	0.240	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	0.246	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	W	-	-	-	-	-	-
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.072	0.060	0.074	0.058	0.055	0.075	0.051	0.057
	0.5	0.117	0.058	0.085	0.081	0.042	0.076	0.072	0.084
	1.0	0.392	0.081	0.335	0.336	0.023	0.116	0.229	0.288
	2.5	1.000	0.548	1.000	1.000	0.995	0.993	1.000	1.000
	5.0	1.000	0.987	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	-	-	-	-	-	-	-

c) dynamic, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.128	0.059	0.177	0.117	0.074	0.090	0.093	0.114
	0.5	0.286	0.084	0.348	0.293	0.087	0.093	0.333	0.358
	1.0	0.693	0.154	0.834	0.834	0.382	0.338	0.879	0.880
	2.5	1.000	0.149	1.000	1.000	1.000	0.999	1.000	1.000
	5.0	1.000	0.022	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	0.023	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	0.011	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	D	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.511	0.047	0.734	0.240	0.064	0.080	0.082	0.096
	0.5	0.601	0.066	0.800	0.364	0.056	0.104	0.106	0.139
	1.0	0.771	0.064	0.901	0.590	0.038	0.172	0.121	0.247
	2.5	0.985	0.040	1.000	0.988	0.126	0.738	0.602	0.916
	5.0	0.999	0.051	1.000	1.000	0.178	0.986	0.679	0.993
	7.5	1.000	0.040	1.000	1.000	0.046	0.999	0.174	0.934
	10.0	1.000	0.012	1.000	1.000	0.010	0.995	0.012	0.587
	cause	-	D	-	-	D	D	D	D
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.102	0.056	0.090	0.073	0.068	0.087	0.055	0.063
	0.5	0.179	0.067	0.133	0.114	0.051	0.089	0.079	0.095
	1.0	0.370	0.070	0.312	0.285	0.015	0.108	0.170	0.237
	2.5	0.919	0.054	0.920	0.928	0.136	0.571	0.836	0.940
	5.0	0.991	0.045	0.993	0.995	0.107	0.981	0.856	0.995
	7.5	0.994	0.044	0.997	0.996	0.039	0.999	0.300	0.944
	10.0	0.999	0.017	0.995	0.999	0.012	0.996	0.017	0.609
	cause	D	D	D	D	D	D	D	D

b) static, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.090	0.061	0.111	0.096	0.083	0.109	0.069	0.081
	0.5	0.209	0.067	0.268	0.236	0.079	0.106	0.250	0.269
	1.0	0.672	0.237	0.835	0.845	0.325	0.288	0.866	0.866
	2.5	1.000	0.944	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	0.944	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	0.860	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	R	-	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.261	0.054	0.367	0.129	0.077	0.103	0.072	0.088
	0.5	0.330	0.048	0.425	0.180	0.071	0.133	0.102	0.135
	1.0	0.604	0.054	0.708	0.465	0.049	0.277	0.135	0.312
	2.5	0.999	0.069	1.000	1.000	0.300	0.965	0.970	1.000
	5.0	1.000	0.012	1.000	1.000	0.047	1.000	1.000	1.000
	7.5	1.000	0.000	1.000	1.000	0.000	1.000	1.000	1.000
	10.0	1.000	0.000	1.000	1.000	0.000	1.000	1.000	1.000
	cause	-	R,W	-	-	R,W	-	-	-
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.085	0.060	0.069	0.064	0.079	0.115	0.061	0.071
	0.5	0.122	0.048	0.095	0.090	0.056	0.122	0.085	0.107
	1.0	0.426	0.074	0.326	0.336	0.034	0.157	0.224	0.303
	2.5	1.000	0.490	1.000	1.000	0.948	0.984	1.000	1.000
	5.0	1.000	0.899	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	0.915	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	0.875	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	R,W	-	-	-	-	-	-

d) dynamic, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.129	0.049	0.138	0.096	0.085	0.119	0.088	0.111
	0.5	0.276	0.072	0.320	0.294	0.086	0.107	0.304	0.336
	1.0	0.681	0.173	0.834	0.837	0.355	0.315	0.880	0.881
	2.5	1.000	0.214	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	0.988	0.025	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	0.878	0.020	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	0.944	0.005	1.000	1.000	1.000	1.000	1.000	1.000
	cause	R	R,D	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.528	0.052	0.754	0.229	0.075	0.108	0.089	0.109
	0.5	0.615	0.049	0.796	0.344	0.072	0.131	0.117	0.158
	1.0	0.772	0.058	0.904	0.595	0.050	0.230	0.123	0.271
	2.5	0.984	0.040	0.999	0.979	0.094	0.779	0.489	0.897
	5.0	1.000	0.044	1.000	1.000	0.007	0.965	0.243	0.853
	7.5	1.000	0.030	1.000	1.000	0.000	0.992	0.020	0.461
	10.0	1.000	0.004	1.000	0.999	0.000	0.991	0.000	0.119
	cause	-	R,D	-	R,D	R,D	R,D	R,D	R,D
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.103	0.047	0.063	0.049	0.085	0.119	0.066	0.080
	0.5	0.186	0.052	0.130	0.128	0.063	0.116	0.095	0.118
	1.0	0.399	0.078	0.326	0.329	0.023	0.161	0.188	0.277
	2.5	0.923	0.057	0.896	0.909	0.134	0.670	0.793	0.938
	5.0	0.971	0.043	0.866	0.911	0.005	0.953	0.427	0.885
	7.5	0.905	0.036	0.534	0.636	0.000	0.989	0.041	0.476
	10.0	0.793	0.011	0.224	0.309	0.000	0.988	0.002	0.120
	cause	R,D	R,D	R,D	R,D	R,D	R,D	R,D	R,D

Table 4-4: Power comparison under DGP4 (5% level)

a) static, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.091	0.063	0.127	0.089	0.047	0.071	0.062	0.064
	0.5	0.502	0.075	0.510	0.453	0.418	0.397	0.241	0.639
	1.0	0.989	0.071	0.989	0.989	0.999	0.999	0.993	1.000
	2.5	1.000	0.103	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	0.103	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	0.132	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	0.155	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	W	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.238	0.058	0.364	0.105	0.052	0.069	0.064	0.072
	0.5	0.821	0.044	0.898	0.723	0.084	0.115	0.099	0.229
	1.0	0.997	0.031	1.000	0.999	0.024	0.048	0.731	0.937
	2.5	1.000	0.025	1.000	1.000	0.000	0.009	1.000	1.000
	5.0	1.000	0.045	1.000	1.000	0.000	0.061	1.000	1.000
	7.5	1.000	0.084	1.000	1.000	0.000	0.187	1.000	1.000
	10.0	1.000	0.158	1.000	1.000	0.000	0.359	1.000	1.000
	cause	-	W	-	-	W	W	-	-
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.072	0.060	0.074	0.058	0.055	0.075	0.051	0.057
	0.5	0.256	0.082	0.204	0.163	0.159	0.218	0.062	0.193
	1.0	0.791	0.160	0.722	0.698	0.274	0.316	0.709	0.921
	2.5	1.000	0.706	1.000	1.000	0.998	0.997	1.000	1.000
	5.0	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	-	-	-	-	-	-	-

c) dynamic, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.128	0.059	0.177	0.117	0.074	0.090	0.093	0.114
	0.5	0.562	0.054	0.560	0.482	0.459	0.421	0.336	0.680
	1.0	0.990	0.047	0.989	0.987	0.993	0.993	0.988	1.000
	2.5	1.000	0.035	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	D	-	-	-	-	-	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.511	0.047	0.734	0.240	0.064	0.080	0.082	0.096
	0.5	0.915	0.046	0.958	0.787	0.064	0.089	0.091	0.216
	1.0	0.999	0.041	1.000	0.996	0.017	0.060	0.285	0.606
	2.5	1.000	0.038	1.000	1.000	0.004	0.032	0.224	0.851
	5.0	1.000	0.000	1.000	1.000	0.000	0.000	0.000	0.239
	7.5	1.000	0.000	1.000	1.000	0.000	0.000	0.000	0.021
	10.0	1.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000
	cause	-	D	-	-	D	D	D	D
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.102	0.056	0.090	0.073	0.068	0.087	0.055	0.063
	0.5	0.248	0.052	0.171	0.141	0.117	0.157	0.066	0.162
	1.0	0.569	0.059	0.453	0.400	0.070	0.137	0.227	0.542
	2.5	0.930	0.047	0.849	0.844	0.024	0.063	0.150	0.804
	5.0	0.943	0.000	0.631	0.673	0.000	0.000	0.000	0.267
	7.5	0.851	0.000	0.470	0.526	0.000	0.000	0.000	0.014
	10.0	0.725	0.000	0.460	0.493	0.000	0.000	0.000	0.000
	cause	D	D	D	D	D	D	D	D

b) static, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.090	0.061	0.111	0.096	0.083	0.109	0.069	0.081
	0.5	0.476	0.067	0.438	0.394	0.439	0.445	0.193	0.672
	1.0	0.987	0.081	0.955	0.967	0.999	0.999	0.663	1.000
	2.5	0.376	0.096	0.282	0.288	1.000	1.000	0.721	1.000
	5.0	0.000	0.112	0.441	0.171	1.000	1.000	0.762	1.000
	7.5	0.000	0.131	0.562	0.199	1.000	1.000	0.784	1.000
	10.0	0.000	0.166	0.645	0.295	1.000	1.000	0.781	1.000
	cause	R	R,W	R	R	-	-	R	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.261	0.054	0.367	0.129	0.077	0.103	0.072	0.088
	0.5	0.822	0.041	0.850	0.675	0.088	0.131	0.100	0.269
	1.0	1.000	0.033	1.000	0.999	0.017	0.059	0.585	0.904
	2.5	1.000	0.031	1.000	1.000	0.000	0.013	1.000	1.000
	5.0	1.000	0.066	1.000	1.000	0.000	0.010	1.000	1.000
	7.5	1.000	0.095	1.000	1.000	0.000	0.018	1.000	1.000
	10.0	1.000	0.150	1.000	1.000	0.000	0.014	1.000	1.000
	cause	-	R,W	-	-	R,W	R,W	-	-
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.085	0.060	0.069	0.064	0.079	0.115	0.061	0.071
	0.5	0.218	0.060	0.155	0.139	0.179	0.250	0.081	0.240
	1.0	0.727	0.141	0.593	0.602	0.275	0.353	0.604	0.930
	2.5	0.736	0.652	0.934	0.949	0.958	0.949	0.999	1.000
	5.0	0.001	0.992	0.977	0.974	0.990	0.999	1.000	1.000
	7.5	0.000	1.000	0.997	0.997	0.995	1.000	1.000	1.000
	10.0	0.000	1.000	0.990	0.996	0.991	0.999	1.000	1.000
	cause	R,W	-	R,W	R,W	R,W	R,W	-	-

d) dynamic, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.129	0.049	0.138	0.096	0.085	0.119	0.088	0.111
	0.5	0.526	0.063	0.481	0.434	0.453	0.456	0.222	0.705
	1.0	0.984	0.047	0.945	0.963	0.995	0.995	0.762	1.000
	2.5	0.376	0.018	0.723	0.827	1.000	1.000	0.977	1.000
	5.0	0.017	0.000	0.172	0.237	1.000	1.000	0.664	1.000
	7.5	0.001	0.000	0.105	0.131	1.000	1.000	0.142	1.000
	10.0	0.000	0.000	0.109	0.136	1.000	1.000	0.050	1.000
	cause	R	R,D	R	R	-	-	R	-
rolling	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.528	0.052	0.754	0.229	0.075	0.108	0.089	0.109
	0.5	0.893	0.045	0.957	0.763	0.070	0.117	0.098	0.255
	1.0	0.999	0.050	1.000	0.994	0.022	0.081	0.256	0.598
	2.5	1.000	0.024	1.000	1.000	0.004	0.035	0.043	0.441
	5.0	1.000	0.000	1.000	0.994	0.000	0.001	0.000	0.032
	7.5	0.996	0.000	0.991	0.900	0.000	0.000	0.000	0.000
	10.0	0.967	0.000	0.958	0.779	0.000	0.000	0.000	0.000
	cause	-	R,D	R,D	R,D	R,D	R,D	R,D	R,D
recursive	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
	0.0	0.103	0.047	0.063	0.049	0.085	0.119	0.066	0.080
	0.5	0.257	0.058	0.178	0.163	0.137	0.206	0.090	0.221
	1.0	0.542	0.056	0.367	0.358	0.087	0.175	0.224	0.591
	2.5	0.827	0.025	0.563	0.603	0.020	0.057	0.043	0.575
	5.0	0.914	0.000	0.426	0.528	0.000	0.001	0.000	0.026
	7.5	0.793	0.000	0.385	0.444	0.000	0.000	0.000	0.000
	10.0	0.690	0.000	0.396	0.433	0.000	0.000	0.000	0.000
	cause	R,D	R,D	R,D	R,D	R,D	R,D	R,D	R,D

Table 4-5: Power comparison under DGP5 (5% level)

a) static, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.091	0.063	0.127	0.089	0.047	0.071	0.062	0.064
	0.5	0.598	0.528	0.956	0.606	0.997	0.996	0.968	0.997
	1.0	0.960	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	-	-	-	-	-	-	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.238	0.058	0.364	0.105	0.052	0.069	0.064	0.072
	0.5	0.893	0.303	0.995	0.601	0.835	0.791	0.632	0.820
	1.0	0.996	0.957	1.000	0.993	1.000	1.000	1.000	1.000
	2.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	-	-	-	-	-	-	-	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.072	0.060	0.074	0.058	0.055	0.075	0.051	0.057
	0.5	0.031	0.592	0.603	0.618	0.944	0.927	0.646	0.940
	1.0	0.007	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	W	-	-	-	-	-	-	-

b) static, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.090	0.061	0.111	0.096	0.083	0.109	0.069	0.081
	0.5	0.458	0.452	0.776	0.519	0.997	0.996	0.831	0.997
	1.0	0.073	0.992	0.997	0.998	1.000	1.000	1.000	1.000
	2.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	R	-	-	-	-	-	-	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.261	0.054	0.367	0.129	0.077	0.103	0.072	0.088
	0.5	0.848	0.240	0.999	0.528	0.836	0.804	0.607	0.815
	1.0	0.968	0.899	1.000	0.899	1.000	1.000	1.000	1.000
	2.5	1.000	0.998	1.000	0.998	1.000	1.000	1.000	1.000
	5.0	1.000	0.713	1.000	0.993	1.000	1.000	1.000	1.000
	7.5	1.000	0.712	1.000	0.974	1.000	1.000	1.000	1.000
	10.0	1.000	0.996	1.000	0.993	1.000	1.000	1.000	1.000
	cause	-	R	-	R	-	-	-	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.085	0.060	0.069	0.064	0.079	0.115	0.061	0.071
	0.5	0.017	0.534	0.563	0.580	0.953	0.939	0.633	0.938
	1.0	0.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	2.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7.5	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10.0	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	cause	W,R	-	-	-	-	-	-	-

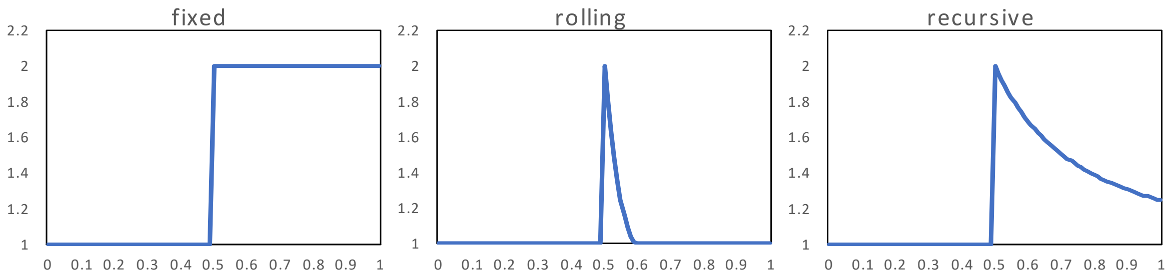
c) dynamic, non-robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.128	0.059	0.177	0.117	0.074	0.090	0.093	0.114
	0.5	0.512	0.145	0.948	0.180	0.993	0.989	0.955	0.996
	1.0	0.408	0.198	1.000	0.201	1.000	1.000	1.000	1.000
	2.5	0.000	0.085	1.000	0.012	1.000	1.000	1.000	1.000
	5.0	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000
	7.5	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000
	10.0	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000
	cause	D	D	-	D	-	-	-	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.511	0.047	0.734	0.240	0.064	0.080	0.082	0.096
	0.5	0.883	0.101	0.997	0.483	0.462	0.431	0.202	0.412
	1.0	0.686	0.149	1.000	0.248	0.786	0.759	0.349	0.762
	2.5	0.008	0.063	1.000	0.009	0.556	0.529	0.008	0.062
	5.0	0.000	0.000	1.000	0.000	0.308	0.237	0.000	0.000
	7.5	0.000	0.000	1.000	0.000	0.265	0.196	0.000	0.000
	10.0	0.000	0.000	1.000	0.000	0.228	0.167	0.000	0.000
	cause	D	D	-	D	D	D	D	D
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.102	0.056	0.090	0.073	0.068	0.087	0.055	0.063
	0.5	0.068	0.155	0.221	0.145	0.548	0.529	0.161	0.450
	1.0	0.023	0.212	0.339	0.181	0.781	0.776	0.173	0.640
	2.5	0.000	0.086	0.032	0.013	0.772	0.732	0.000	0.020
	5.0	0.000	0.000	0.000	0.000	0.627	0.507	0.000	0.000
	7.5	0.000	0.000	0.000	0.000	0.526	0.444	0.000	0.000
	10.0	0.000	0.000	0.016	0.000	0.477	0.366	0.000	0.000
	cause	D	D	D	D	D	D	D	D

d) dynamic, robust

	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
fixed	0.0	0.129	0.049	0.138	0.096	0.095	0.110	0.095	0.130
	0.5	0.411	0.152	0.612	0.170	0.995	1.000	0.745	1.000
	1.0	0.228	0.167	0.246	0.156	1.000	1.000	0.980	1.000
	2.5	0.000	0.090	0.015	0.014	1.000	1.000	0.620	1.000
	5.0	0.000	0.000	0.001	0.000	1.000	1.000	0.145	1.000
	7.5	0.000	0.000	0.000	0.000	1.000	1.000	0.165	1.000
	10.0	0.000	0.000	0.001	0.000	1.000	1.000	0.115	1.000
	cause	R,D	R,D	R	R,D	-	-	R	-
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
rolling	0.0	0.528	0.052	0.754	0.229	0.110	0.110	0.095	0.135
	0.5	0.835	0.095	0.998	0.470	0.430	0.405	0.210	0.425
	1.0	0.603	0.126	1.000	0.253	0.595	0.615	0.315	0.560
	2.5	0.005	0.066	1.000	0.010	0.260	0.315	0.000	0.005
	5.0	0.000	0.000	0.964	0.000	0.075	0.105	0.000	0.000
	7.5	0.000	0.000	0.963	0.000	0.030	0.060	0.000	0.000
	10.0	0.000	0.000	0.990	0.000	0.030	0.070	0.000	0.000
	cause	R,D	R,D	R,D	R,D	R,D	R,D	R,D	R,D
	bA	GR1	GR2	SGR1	SGR2	DSW1	DSW2	TLSW	TLUD
recursive	0.0	0.103	0.047	0.063	0.049	0.105	0.100	0.065	0.100
	0.5	0.052	0.165	0.179	0.145	0.555	0.510	0.150	0.445
	1.0	0.013	0.187	0.159	0.149	0.650	0.665	0.115	0.410
	2.5	0.000	0.095	0.012	0.013	0.535	0.490	0.000	0.000
	5.0	0.000	0.000	0.000	0.000	0.230	0.125	0.000	0.000
	7.5	0.000	0.000	0.000	0.000	0.125	0.075	0.000	0.000
	10.0	0.000	0.000	0.000	0.000	0.100	0.080	0.000	0.000
	cause	R,D	R,D	R,D	R,D	R,D	R,D	R,D	R,D

Figure 1. Shapes of $E[L^*(r)]$ for a coefficient break in a static regression
a) small break ($\Delta_\beta = 1$):



b) large break ($\Delta_\beta = 5$)

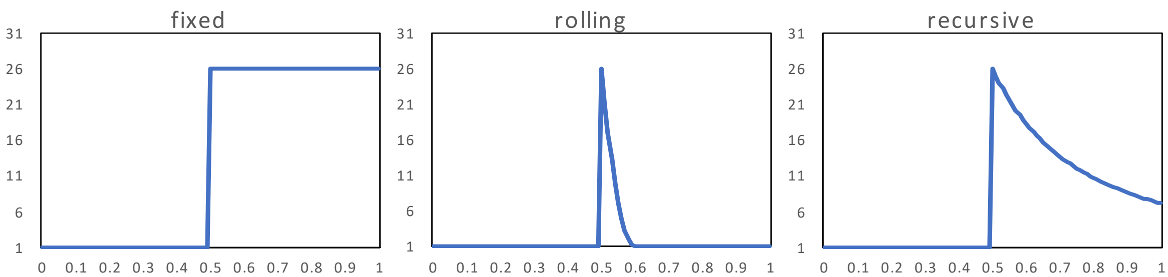
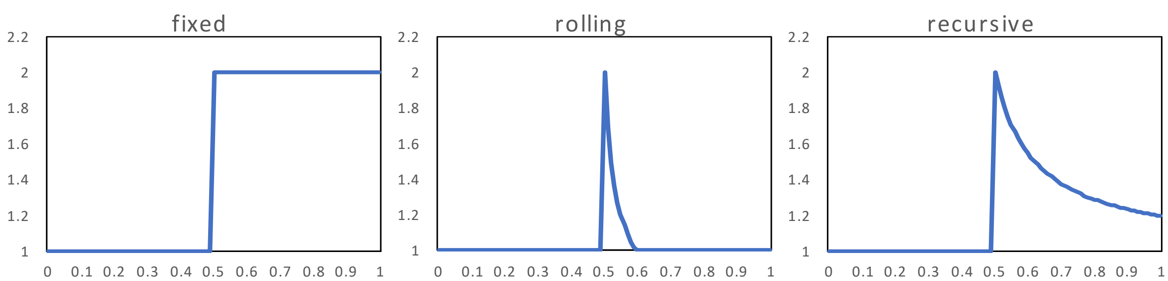


Figure 2. Shapes of $E[L^*(r)]$ for a coefficient break in a dynamic regression
a) small break ($\Delta_\beta = 1$):



b) large break ($\Delta_\beta = 5$):

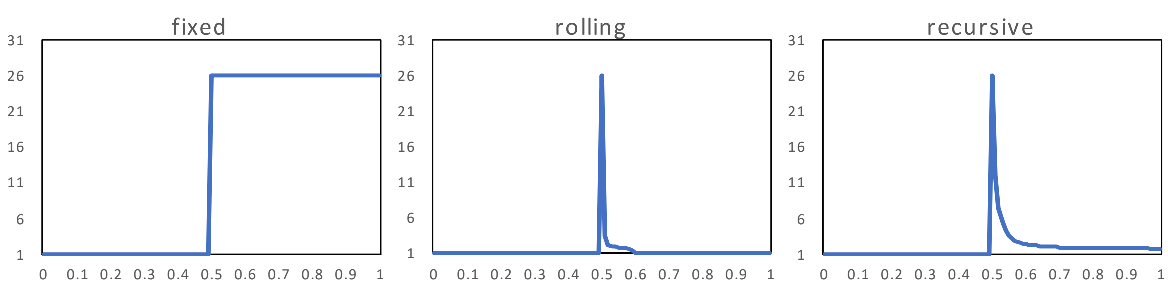


Figure 3. Equity premium for the S&P 500

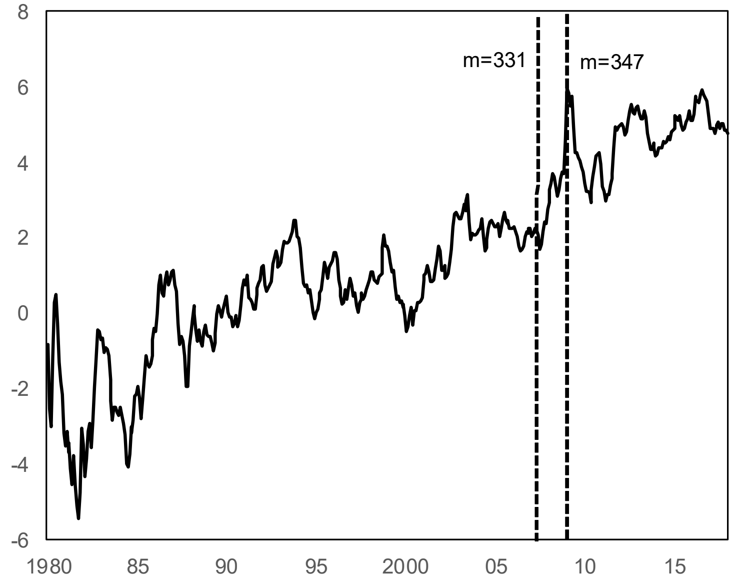


Table 5. Tests for changes in forecast performance: equity premium forecasts

Static

	tau	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2	LM1	LM2
fixed	1	214.89 ***	214.89 ***	266.23 ***	534.37 ***	2.29 **	2.55 **	2.81 **	2.33 *	432.37 ***	432.78 ***
	3	212.85 ***	212.85 ***	270.02 ***	551.42 ***	2.10 **	2.17 **	2.38	2.21	435.27 ***	408.26 ***
	6	196.44 ***	196.44 ***	291.79 ***	622.41 ***	1.98 **	1.88 *	2.20	2.20	436.88 ***	408.91 ***
	12	223.76 ***	223.76 ***	282.76 ***	522.12 ***	1.96 **	1.67 *	2.08	2.08	434.40 ***	416.75 ***
rolling	1	15.09 **	44.78 ***	168.45 ***	168.45 ***	2.78 ***	3.19 ***	6.28 ***	6.28 ***	416.82 ***	422.25 ***
	3	15.73 **	52.28 ***	151.58 ***	151.58 ***	2.53 **	3.10 ***	6.00 ***	6.00 ***	421.08 ***	423.34 ***
	6	18.36 ***	55.21 ***	185.80 ***	185.80 ***	2.35 **	2.67 ***	8.02 ***	6.87 ***	423.03 ***	424.54 ***
	12	16.35 ***	72.05 ***	186.13 **	186.13 ***	2.39 ***	2.01 **	7.75 ***	6.54 ***	418.71 ***	419.76 ***
recursive	1	179.75 ***	179.75 ***	190.88 ***	190.88 ***	2.57 **	2.93 ***	3.31 ***	2.61 **	419.00 ***	422.98 ***
	3	239.51 ***	239.51 ***	234.21 ***	234.21 ***	2.35 **	2.83 ***	3.15 **	2.41 *	423.03 ***	423.95 ***
	6	322.67 ***	322.67 ***	319.95 ***	319.95 ***	2.20 **	2.47 **	2.60 *	2.30 *	424.89 ***	425.10 ***
	12	205.12 ***	205.12 ***	208.90 ***	208.90 ***	2.29 **	1.91 *	2.62 *	2.43 *	419.54 ***	419.95 ***

Dynamic

	tau	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2	LM1	LM2
fixed	1	23.94 ***	23.94 ***	164.48 ***	164.48 ***	1.57	2.94 ***	11.95 ***	11.95 ***	72.71 ***	51.37 ***
	3	90.60 ***	90.60 ***	126.27 ***	191.85 ***	1.09	0.79	3.95 ***	3.95 ***	329.85 ***	241.43 ***
	6	198.75 ***	198.75 ***	307.34 ***	395.40 ***	2.27 **	0.88	2.29	2.27 *	376.24 ***	369.13 ***
	12	223.18 ***	223.18 ***	298.22 ***	619.16 ***	2.11 **	1.43	2.29	2.23	423.79 ***	413.79 ***
rolling	1	4.42	4.80	166.02 ***	166.02 ***	0.35	0.93	5.14 ***	3.03 **	68.68 ***	51.17 ***
	3	3.14	3.47	43.14 ***	102.92 ***	0.53	1.63	3.43 ***	1.22	304.46 ***	237.76 ***
	6	2.42	2.42	27.63 ***	54.11 ***	1.13	0.57	8.12 ***	2.90 **	346.24 ***	312.03 ***
	12	2.34	2.81	17.96 ***	34.56 ***	1.50	0.19	7.50 ***	4.33 ***	386.88 ***	378.56 ***
recursive	1	4.52	4.52	162.42 ***	162.42 ***	2.89 ***	3.72 ***	13.84 ***	11.91 ***	67.25 ***	51.16 ***
	3	2.62	2.85	51.26 ***	97.82 ***	0.63	3.98 ***	6.68 ***	5.00 ***	313.95 ***	238.64 ***
	6	3.78	3.78	33.78 ***	51.63 ***	0.51	1.32	2.88 **	1.92	357.41 ***	313.73 ***
	12	4.67	4.67	27.42 ***	35.34 ***	1.28	0.13	1.43	1.29	391.78 ***	384.18 ***

Figure 4. U.S. real-time inflation rate

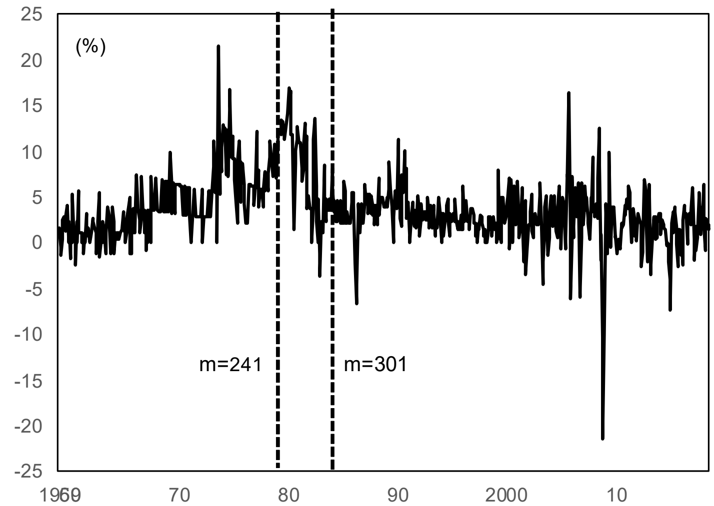


Table 6. Tests for change in forecasting performance: inflation forecasts

Dynamic 1

	tau	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2	LM1	LM2
fixed	1	84.69 ***	84.69 ***	25.65 ***	97.96 ***	0.29	0.08	4.48 ***	1.09	262.17 ***	377.83 ***
	3	86.86 ***	86.86 ***	18.10 ***	100.57 ***	0.91	0.45	4.32 ***	1.68	215.57 ***	324.12 ***
	6	34.00 ***	34.00 ***	13.93 ***	90.18 ***	0.99	0.34	4.38 ***	1.44	206.69 ***	304.09 ***
	12	66.44 ***	66.44 ***	15.53 ***	90.46 ***	0.93	0.39	4.53 ***	1.43	173.27 ***	270.93 ***
rolling	1	11.51 *	11.51	20.68 ***	97.59 ***	0.88	0.30	5.47 ***	0.89	251.68 ***	365.56 ***
	3	7.64	7.64	15.95 ***	73.09 ***	1.21	0.39	7.13 ***	1.34	214.88 ***	320.59 ***
	6	10.22	10.22	16.68 ***	73.20 ***	1.27	0.35	8.06 ***	1.44	211.15 ***	322.59 ***
	12	10.66	10.66	16.30 ***	74.73 ***	1.63	0.47	10.02 ***	1.91	210.89 ***	322.20 ***
recursive	1	15.47 **	15.49 **	20.58 ***	99.77 ***	0.54	0.18	2.22	1.02	267.17 ***	380.19 ***
	3	9.10	9.10	12.26 **	73.10 ***	1.06	0.49	2.52 *	1.50	230.30 ***	330.83 ***
	6	7.06	7.06	13.64 ***	73.30 ***	1.02	0.44	2.42 *	1.35	211.93 ***	308.10 ***
	12	6.93	6.93	15.00 ***	74.90 ***	1.06	0.41	2.86 **	1.30	192.71 ***	286.08 ***

Dynamic 3

	tau	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2	LM1	LM2
fixed	1	65.94 ***	65.94 ***	20.08 ***	102.20 ***	0.77	0.34	4.40 ***	1.50	244.07 ***	359.22 ***
	3	32.88 ***	32.88 ***	13.26 ***	88.70 ***	1.18	0.52	4.26 ***	1.52	206.98 ***	309.69 ***
	6	28.18 ***	28.18 ***	14.07 ***	88.94 ***	1.02	0.35	4.37 ***	1.56	212.29 ***	310.01 ***
	12	70.15 ***	70.15 ***	15.58 ***	90.15 ***	1.07	0.35	4.57 ***	1.39	191.00 ***	291.01 ***
rolling	1	10.06	11.39	20.71 ***	95.36 ***	1.03	0.35	5.51 ***	1.14	242.31 ***	365.42 ***
	3	8.76	8.76	15.81 ***	73.32 ***	1.45	0.44	7.21 ***	1.49	213.46 ***	319.28 ***
	6	9.36	9.36	16.46 ***	73.23 ***	1.29	0.40	7.54 ***	1.42	216.42 ***	327.91 ***
	12	11.56 *	11.56	17.52 ***	75.18 ***	1.77 *	0.47	10.97 ***	2.00	204.48 ***	317.7 ***
recursive	1	12.85 **	13.41 *	18.78 ***	89.55 ***	0.90	0.41	2.57 *	1.32	256.18 ***	371.77 ***
	3	7.12	7.12	11.88 **	73.34 ***	1.28	0.63	2.36	1.47	218.54 ***	310.21 ***
	6	9.24	9.24	14.05 ***	73.31 ***	1.16	0.55	2.64 *	1.54	213.83 ***	312.89 ***
	12	8.09	8.09	16.14 ***	75.28 ***	1.20	0.42	3.01 **	1.33	198.17 ***	295.06 ***

"Testing for Changes in Forecasting Performance" by Pierre Perron and Yohei Yamamoto Supplementary Material

a) Comments about the complementarity of forecast breakdown and structural change tests.

We here expand on a comment made in the introduction about the complementarity of forecast breakdown and structural change tests. Since forecast breakdowns are often associated with changes in the parameters of the forecasting models, one may question the value-added of forecast breakdown tests over simply using a structural change test directly on the forecasting model. We expand upon an interesting case alluded to in the introduction, whereby it is shown that structural change and forecast breakdown tests complement each other; when one test has poor power for some parameter configurations, the other has high power. We consider data generated such that the relevant predictor changes from some x_t to some w_t at date $t = T_b$, i.e.,

$$y_{t+\tau} = \begin{cases} \alpha + x_t\beta + e_{t+\tau} & \text{for } t = 1, ..., T_b \\ \alpha + w_t\beta + e_{t+\tau} & \text{for } t = T_b + 1, ..., T - \tau \end{cases}. \tag{A.1}$$

However, w_t is not accessible to the researcher, hence one uses the following misspecified model over the entire sample

$$y_{t+\tau} = \alpha + x_t\beta + e_{t+\tau} \quad \text{for } t = 1, ..., T - \tau. \tag{A.2}$$

We first theoretically illustrate how the forecast breakdown and structural change tests behave under this setting. For simplicity, let x_t, w_t and e_t be independent scalar random variables with $E(x_t) = E(w_t) = \mu$ and $E(e_t) = 0$. We also let $Var(x_t) = \sigma_x^2$, $Var(w_t) = \sigma_w^2$ and $Var(e_t) = \sigma^2$. For the forecasting tests, assume that the fixed window scheme is applied (shown to be the preferred one in this paper) with an in-sample length smaller than or equal to T_b so that the coefficient β is consistently estimated. It is easy to show that the expected value of the limit of the quadratic forecasting loss changes from σ^2 to $\sigma^2 + \sigma_w^2\beta^2 + \sigma_x^2\beta^2$ before and after $t = T_b$. Hence, the magnitude of the change is

$$\sigma_w^2\beta^2 + \sigma_x^2\beta^2. \tag{A.3}$$

For the structural change tests, we consider the Wald statistic (divided by T) assuming, for simplicity, the break date to be known:

$$T^{-1}W = [(SSR_r - SSR_u)/T]/[SSR_u/T].$$

The restricted ($\sim$) and unrestricted residuals ($\hat{}$) are then, respectively,

$$\widetilde{e}_{t+\tau} = \begin{cases} e_{t+\tau} - x_t(\widetilde{\beta} - \beta) - (\widetilde{\alpha} - \alpha) & \text{for } t = 1, ..., T_b \\ e_{t+\tau} + w_t\beta - x_t\widetilde{\beta} - (\widetilde{\alpha} - \alpha) & \text{for } t = T_b + 1, ..., T - \tau \end{cases},$$

$$\hat{e}_{t+\tau} = \begin{cases} e_{t+\tau} - x_t(\hat{\beta}_{(1)} - \beta) - (\hat{\alpha}_{(1)} - \alpha) & \text{for } t = 1, ..., T_b \\ e_{t+\tau} + w_t\beta - x_t\hat{\beta}_{(2)} - (\hat{\alpha}_{(2)} - \alpha) & \text{for } t = T_b + 1, ..., T - \tau \end{cases}.$$

After some algebra, one can show that $\widetilde{\alpha} \xrightarrow{p} \alpha$ when $E(x_t) = E(w_t)$ (otherwise, $\widetilde{\alpha} \xrightarrow{p} \alpha + [\lambda E(x_t) + (1-\lambda)E(w_t)]$). Also $\hat{\alpha}_{(1)}, \hat{\alpha}_{(2)} \xrightarrow{p} \alpha$, $\widetilde{\beta} \xrightarrow{p} \lambda\beta$, $\hat{\beta}_{(1)} \xrightarrow{p} \beta$ and $\hat{\beta}_{(2)} \xrightarrow{p} 0$, where $\lambda = \lim_{T\to\infty} T_b/T$, so that the limit of $(SSR_r - SSR_u)/T$ is $\lambda(1-\lambda)\sigma_x^2\beta^2$ and that of SSR_u/T is $\sigma^2 + (1-\lambda)\sigma_w^2\beta^2$. Hence, the limit of the scaled Wald statistic is

$$\lambda(1-\lambda)\sigma_x^2\beta^2/[\sigma^2 + (1-\lambda)\sigma_w^2\beta^2]. \tag{A.4}$$

Comparing (A.3) and (A.4) yields the following insight. When the variance of the unavailable predictor σ_w^2 increases, the change in the forecast loss increases and the change in forecast performance is easier to detect. However, a large variance would not help to detect a change in β associated with the predictor x_t via a structural change test because an increase in σ_w^2 lowers the value of test statistic. And vice-versa, when σ_w^2 is small, the change in the loss function is reduced while the value of the Wald test for a change is increased.

To quantify the implications of this insight, we implement a simple Monte Carlo simulation. We set $\alpha = \beta = 1$ and generate $e_t \sim i.i.d.\ N(0,1)$ for $t = 1, ..., T$. The variables are generated by $x_t \sim i.i.d.N(0,1)$ and $w_t \sim i.i.d.N(0,\sigma_w^2)$, which are independent of each other. Let $T = 150$, $T_b = 75$ and $\tau = 1$, for the sake of illustration (the results remain qualitatively the same with other parameter values). For the forecasting tests, we consider the GR_m test with $m = 40$, the SGR test with $[0.2T] \leq m \leq [0.8T]$, the DSW test with $m_0 = [0.2T]$ and $\bar{\mu} = 0.5$, as well as the $TLSW$ and $TLUD$ tests described in the text. We only consider the fixed window scheme for reason discussed in the text. For the structural break tests, we consider the full-sample regression model (A.2) and test for variations in both (α, β). Specifically, we use the $\sup F$ test of Andrews (1993) and the $UD\max$ test of Bai and Perron (1998). The maximum number of breaks for the $UD\max$ tests is five and the truncation parameter $\epsilon = 0.1$ is used. The number of replications is 1,000. Table S.1 shows the rejection frequencies when we vary σ_w from 0.0 to 10.0. It illustrates a clear complementarity between the power of the forecasting and structural change tests. When σ_w is small, the rejection frequencies of the forecasting tests are small for all tests considered. However, the rejection frequencies of the structural change tests are large because the denominator in (A.4) is small, which leads a large value of test statistic. The simulation results show that this feature continues to hold with more general tests with an unknown break date or with multiple breaks.

Table S.1: Rejection frequencies

σ_w	GR	SGR	DSW	TLSW	TLUD	SupF	UD max
0.0	0.123	0.253	0.137	0.611	0.677	0.995	0.996
0.5	0.252	0.424	0.189	0.507	0.595	0.995	0.995
1.0	0.564	0.832	0.340	0.509	0.592	0.959	0.960
2.5	0.999	1.000	0.885	0.996	0.997	0.483	0.496
5.0	1.000	1.000	0.966	1.000	1.000	0.141	0.171
7.5	1.000	1.000	0.983	1.000	1.000	0.095	0.125
10.0	1.000	1.000	0.979	1.000	1.000	0.073	0.091

b) Proof of Theorem 1

For the proof, we denote $T_b(m)$ by T_b for simplicity. The Wald test for a constant mean versus one break at time $t = T_b = m_0 + \tau + \lfloor \lambda n_0 \rfloor$ for the series $\left\{L^o_{t+\tau}\right\}_{t=m}^{T-\tau}$ is

$$W^m(T_b) = \frac{SSR_{L^o(m)} - SSR(T_b)_{L^o(m)}}{\hat{V}_{L^o(m)}},$$

where $SSR_{L^o(m)}$, $SSR(T_b)_{L^o(m)}$ and $\hat{V}_{L^o(m)}$ are defined after (1). First, for a given m, the restricted SSR is:

$$\begin{aligned} SSR_{L^o(m)} &= \textstyle\sum_{t=m}^{T-\tau} L^{o2}_{t+\tau} - \dfrac{1}{T-\tau-m+1}\left(\textstyle\sum_{t=m}^{T-\tau} L^o_{t+\tau}\right)^2 \\ &= \textstyle\sum_{t=m}^{T-\tau} L^{o2}_{t+\tau} - \dfrac{n_0}{T-\tau-m+1}\left(n_0^{-1/2}\textstyle\sum_{t=m}^{T-\tau} L^o_{t+\tau}\right)^2, \end{aligned}$$

and the unrestricted SSR assuming a break at $t = T_b$ is given by

$$\begin{aligned} SSR(T_b)_{L^o(m)} &= \textstyle\sum_{t=m}^{T_b-\tau} L^{o2}_{t+\tau} - \dfrac{n_0}{T_b-\tau-m+1}\left(n_0^{-1/2}\textstyle\sum_{t=m}^{T_b-\tau} L^o_{t+\tau}\right)^2 \\ &\quad + \textstyle\sum_{t=T_b-\tau+1}^{T-\tau} L^{o2}_{t} - \dfrac{n_0}{T-T_b}\left(n_0^{-1/2}\textstyle\sum_{t=T_b-\tau+1}^{T-\tau} L^o_{t+\tau}\right)^2. \end{aligned}$$

Hence,

$$\begin{aligned} SSR_{L^o(m)} - SSR(T_b)_{L^o(m)} &= -\frac{n_0}{T-\tau-m+1}\left(n_0^{-1/2}\textstyle\sum_{t=m}^{T-\tau} L^o_{t+\tau}\right)^2 \\ &\quad + \frac{n_0}{T_b-\tau-m+1}\left(n_0^{-1/2}\textstyle\sum_{t=m}^{T_b-\tau} L^o_{t+\tau}\right)^2 \\ &\quad + \frac{n_0}{T-T_b}\left(n_0^{-1/2}\textstyle\sum_{t=T_b-\tau+1}^{T-\tau} L^o_{t+\tau}\right)^2. \end{aligned}$$

Let $\mu = \lim_{T\to\infty}(m-m_0)/n_0$. Then, using $T = n_0 + m_0 + \tau - 1$ and $T_b = m_0 + \tau + \lfloor \lambda n_0 \rfloor$, we have

$$\begin{aligned} \frac{n_0}{T-\tau-m+1} &= \frac{n_0}{n_0-(m-m_0)} \to \frac{1}{1-\mu}, \\ \frac{n_0}{T_b-\tau-m+1} &= \frac{n_0}{\lfloor \lambda n_0 \rfloor-(m-m_0)+1} \to \frac{1}{\lambda-\mu}, \\ \frac{n_0}{T-T_b} &= \frac{n_0}{n_0-\lfloor \lambda n_0 \rfloor-1} \to \frac{1}{1-\lambda}. \end{aligned}$$

Because $L^o_{t+\tau-1} = L_t$ for $t = m+1, ..., T-\tau+1$ and using Assumption 1

$$\begin{aligned} n_0^{-1/2}\textstyle\sum_{t=m}^{T-\tau} L^o_{t+\tau} &= n_0^{-1/2}\textstyle\sum_{t=m-\tau+1}^{T-2\tau+1} L_{t+\tau} \\ &= n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{T-2\tau+1} L_{t+\tau} - n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{m-\tau} L_{t+\tau} \\ &\Rightarrow \Omega^{1/2}\left[W(1)-W(\mu)\right], \end{aligned}$$

$$
\begin{aligned}
n_0^{-1/2}\textstyle\sum_{t=m}^{T_b-\tau} L_{t+\tau}^o &= n_0^{-1/2}\textstyle\sum_{t=m-\tau+1}^{T_b-2\tau+1} L_{t+\tau} \\
&= n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{T_b-2\tau+1} L_{t+\tau} - n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{m-\tau} L_{t+\tau} \\
&\Rightarrow \Omega^{1/2}\left[W(\lambda)-W(\mu)\right],
\end{aligned}
$$

$$
\begin{aligned}
n_0^{-1/2}\textstyle\sum_{t=T_b-\tau+1}^{T-\tau} L_{t+\tau}^o &= n_0^{-1/2}\textstyle\sum_{t=T_b-2\tau+2}^{T-2\tau+1} L_{t+\tau} \\
&= n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{T-2\tau+1} L_{t+\tau} - n_0^{-1/2}\textstyle\sum_{t=m_0-\tau}^{T_b-2\tau+1} L_{t+\tau} \\
&\Rightarrow \Omega^{1/2}\left[W(1)-W(\lambda)\right].
\end{aligned}
$$

Combining the above results yields

$$
\begin{gathered}
SSR_{L^o(m)} - SSR(T_b)_{L^o(m)} \\
\Rightarrow \Omega[-\frac{\left[W(1)-W(\mu)\right]^2}{1-\mu} + \frac{\left[W(\lambda)-W(\mu)\right]^2}{\lambda-\mu} + \frac{\left[W(1)-W(\lambda)\right]^2}{1-\lambda}],
\end{gathered}
$$

and, under the null hypothesis, we have $\hat{V}_{L^o(m)} \xrightarrow{p} \Omega$. Note that $\mu = \lim_{T\to\infty}(m-m_0)/n_0 \leq \lim_{T\to\infty}(m_1-m_0)/n_0 = \bar{\mu}$, so that $\mu \in [0,\bar{\mu}]$. We also have for a trimming parameter ϵ,

$$
\begin{aligned}
T_b &\in [m+\tau+\epsilon n, m+\tau+(1-\epsilon)n], \\
\frac{T_b-m_0-\tau}{n_0} &\in \left[\frac{(m-m_0)+\epsilon n}{n_0}, \frac{(m-m_0)+(1-\epsilon)n}{n_0}\right], \\
\frac{T_b-m_0-\tau}{n_0} &\in \left[\frac{(m-m_0)+\epsilon(n_0+m_0-m)}{n_0}, \frac{(m-m_0)+(1-\epsilon)(n_0+m_0-m)}{n_0}\right].
\end{aligned}
$$

Taking the limit implies $\lambda \in [\mu+\epsilon(1-\mu), 1-\epsilon(1-\mu)]$, and the result follows.

c) Table S.2: Additional critical values of the *DSW* test

$\epsilon = 0.05$					$\epsilon = 0.15$				
$\bar{\mu}$	10%	5%	2.5%	1%	$\bar{\mu}$	10%	5%	2.5%	1%
0.20	11.796	13.738	15.306	17.729	0.20	9.627	11.368	13.037	15.432
0.25	12.237	14.085	15.788	18.148	0.25	10.210	12.048	13.866	16.249
0.30	12.582	14.384	16.253	18.356	0.30	10.374	12.100	14.004	16.448
0.35	12.936	14.662	16.484	18.631	0.35	10.631	12.248	14.068	16.729
0.40	13.498	15.356	17.265	19.328	0.40	11.205	12.946	14.788	17.317
0.45	13.752	15.561	17.276	19.364	0.45	11.293	13.104	14.952	17.478
0.50	14.007	15.808	17.431	19.385	0.50	11.539	13.338	15.088	17.540
0.55	14.359	15.985	17.795	20.052	0.55	11.929	13.534	15.237	17.760
0.60	14.403	16.166	17.984	20.247	0.60	12.202	13.887	15.536	17.626
0.65	14.887	16.589	18.245	20.343	0.65	12.342	14.243	15.631	17.812
0.70	15.245	16.873	18.412	20.772	0.70	12.773	14.434	16.005	17.998
0.75	15.261	16.890	18.695	20.837	0.75	12.801	14.509	16.130	18.159
0.80	15.750	17.412	19.337	21.097	0.80	13.067	15.008	17.060	19.548

d) Table S.3: Additional simulation results for the size of the $TLSW$ and $TLUD$ tests for $T = 150$

a) static, non-robust

		10%	5%	1%
TLSW	fixed	0.113	0.062	0.013
	rolling	0.118	0.064	0.015
	recursive	0.096	0.051	0.01
TLUD	fixed	0.125	0.064	0.013
	rolling	0.126	0.072	0.016
	recursive	0.101	0.057	0.010

c) dynamic, non-robust

		10%	5%	1%
TLSW	fixed	0.166	0.093	0.031
	rolling	0.146	0.083	0.023
	recursive	0.097	0.055	0.015
TLUD	fixed	0.201	0.114	0.036
	rolling	0.177	0.096	0.027
	recursive	0.117	0.063	0.016

b) static, robust

		10%	5%	1%
TLSW	fixed	0.125	0.069	0.018
	rolling	0.132	0.072	0.022
	recursive	0.107	0.061	0.013
TLUD	fixed	0.150	0.081	0.020
	rolling	0.159	0.088	0.025
	recursive	0.129	0.071	0.016

d) dynamic, robust

		10%	5%	1%
TLSW	fixed	0.154	0.088	0.025
	rolling	0.146	0.089	0.028
	recursive	0.108	0.066	0.017
TLUD	fixed	0.200	0.111	0.026
	rolling	0.194	0.109	0.035
	recursive	0.149	0.080	0.018

e) Table S.4: Additional simulation results for the size of the DSW test for $T = 150$

a) static, non-robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.097	0.086	0.093	0.065	0.048	0.050	0.033	0.025	0.020
	rolling	0.101	0.075	0.051	0.072	0.049	0.034	0.032	0.018	0.012
	recursive	0.089	0.084	0.077	0.063	0.052	0.048	0.031	0.022	0.023
0.5	fixed	0.114	0.121	0.117	0.075	0.071	0.071	0.028	0.026	0.028
	rolling	0.118	0.114	0.101	0.081	0.074	0.065	0.031	0.025	0.021
	recursive	0.113	0.117	0.115	0.078	0.076	0.071	0.029	0.025	0.026
0.75	fixed	0.174	0.165	0.160	0.116	0.112	0.105	0.055	0.049	0.046
	rolling	0.174	0.160	0.151	0.125	0.112	0.104	0.056	0.052	0.049
	recursive	0.164	0.152	0.147	0.123	0.119	0.110	0.052	0.049	0.046

b) static, robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.131	0.122	0.122	0.092	0.083	0.082	0.044	0.035	0.027
	rolling	0.128	0.090	0.059	0.092	0.056	0.037	0.045	0.031	0.017
	recursive	0.124	0.112	0.106	0.089	0.076	0.072	0.044	0.032	0.030
0.5	fixed	0.161	0.160	0.151	0.112	0.107	0.100	0.050	0.041	0.043
	rolling	0.170	0.155	0.131	0.109	0.100	0.083	0.055	0.043	0.036
	recursive	0.167	0.156	0.151	0.111	0.103	0.095	0.053	0.044	0.038
0.75	fixed	0.278	0.254	0.244	0.218	0.193	0.186	0.108	0.096	0.091
	rolling	0.269	0.244	0.222	0.217	0.191	0.177	0.117	0.106	0.100
	recursive	0.270	0.247	0.236	0.218	0.191	0.178	0.109	0.101	0.094

c) dynamic, non-robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.109	0.099	0.127	0.078	0.071	0.092	0.036	0.032	0.037
	rolling	0.108	0.078	0.065	0.072	0.052	0.041	0.035	0.024	0.015
	recursive	0.103	0.097	0.103	0.070	0.062	0.071	0.034	0.026	0.031
0.5	fixed	0.124	0.137	0.137	0.086	0.086	0.089	0.033	0.033	0.034
	rolling	0.124	0.115	0.106	0.088	0.078	0.071	0.038	0.033	0.027
	recursive	0.121	0.128	0.133	0.090	0.090	0.093	0.033	0.031	0.034
0.75	fixed	0.189	0.183	0.190	0.132	0.131	0.133	0.061	0.058	0.055
	rolling	0.193	0.177	0.165	0.136	0.128	0.122	0.073	0.064	0.058
	recursive	0.192	0.177	0.176	0.131	0.129	0.129	0.061	0.060	0.058

d) dynamic, robust

		10%			5%			1%		
mu_bar	m0	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T	0.3T	0.2T	0.1T
0.25	fixed	0.135	0.124	0.134	0.095	0.081	0.093	0.044	0.037	0.039
	rolling	0.126	0.098	0.072	0.085	0.057	0.042	0.047	0.027	0.016
	recursive	0.132	0.123	0.127	0.088	0.075	0.083	0.043	0.035	0.035
0.5	fixed	0.167	0.169	0.161	0.119	0.115	0.117	0.053	0.049	0.047
	rolling	0.168	0.154	0.135	0.116	0.103	0.084	0.062	0.051	0.043
	recursive	0.176	0.173	0.171	0.119	0.114	0.113	0.055	0.051	0.049
0.75	fixed	0.286	0.261	0.253	0.226	0.203	0.194	0.119	0.108	0.102
	rolling	0.280	0.256	0.234	0.225	0.202	0.184	0.129	0.115	0.106
	recursive	0.282	0.257	0.251	0.234	0.209	0.203	0.119	0.110	0.106

f) Theoretical results about the limit of the loss function under the various DGPs with a single change

We consider the asymptotic behavior of the loss sequence when there is a single coefficient or variance break. Our results pertain to the loss sequence in large samples. These are directly used when constructing the DSW, $TLSW$ and $TLUD$ tests. The surprise losses are used when constructing the GR_m and SGR tests. However, since the surprise losses series subtract the average of the in-sample losses, the dynamics of both sequences are similar. Hence, we consider only the loss sequence, namely the expected values of $p\lim_{T\to\infty} L_t = L^*(r)$, defined on the unit interval $r \in [0,1]$ where $r = \lim_{T\to\infty}(t/T)$. The squared loss function is applied. For simplicity, and without substantive loss of generality, we consider a single break model generated by:

$$y_{t+\tau} = x_t\beta_t + e_{t+\tau}, \quad \text{for } t = 1, ..., T-\tau, \tag{A.5}$$

where $\beta_t = \beta_1$ for $t \leq [T\lambda_0]$ and $\beta_t = \beta_2$ for $t > [T\lambda_0]$. Again for simplicity, the predictor x_t is a scalar that satisfies $E(x_t^2) = \sigma_x^2$ and $E(x_t x_{t-j}) = \sigma_{xj}$. We also assume that e_t is a white noise with mean 0 and variance σ_t^2, where $\sigma_t^2 = \sigma_1^2$ for $t \leq [T\lambda_0]$ and $\sigma_t^2 = \sigma_2^2$ for $t > [T\lambda_0]$. Let the in-sample length be $m = [T\lambda]$ with m chosen so that $\lambda \leq \lambda_0$. In the following, we consider the following two cases: a coefficient change, i.e., $\beta_1 = 0$ and $\beta_2 = \Delta_\beta$ with $\sigma_1^2 = \sigma_2^2 = \sigma^2$; a variance change, i.e., $\sigma_1^2 = \sigma^2$ and $\sigma_2^2 = \sigma^2 + \Delta_{\sigma^2}$ with $\beta_1 = \beta_2 = \beta$. Suppose we use the static regression model of $y_{t+\tau}$ on x_t to produce a τ-period ahead forecast at time t. For the out-of-sample procedure, we use the estimate of the coefficient β obtained from the in-sample information given in $t \in [1,T]$, labelled as $\hat{\beta}_{[1,t]}$. We consider the three window schemes and estimate the coefficient using OLS for the sample period $[1, m-\tau]$ with the fixed scheme, $[t-m+1, t-\tau]$ with the rolling scheme, and $[1, t-\tau]$ with the recursive scheme. Hence, $\hat{\beta}_{[1,t]}$ with the same t can be different depending on the window scheme. When the static regression is used, the expected value of $L^*(r)$ is such that

$$E[L^*(r)] = E\left[p\lim_{T\to\infty}(y_{t+\tau} - x_t\hat{\beta}_{[1,t]})^2\right] = \sigma_r^2 + \sigma_x^2(\beta_r - \beta^*_{[0,r]})^2, \tag{A.6}$$

where we denote $\beta^*_{[0,r]} = p\lim_{T\to\infty}\hat{\beta}_{[1,t]}$. We also denote the true value of β_t and σ_t^2 by β_r and σ_r^2. Next, we consider a dynamic regression, i.e., the regression of $y_{t+\tau}$ on y_t and x_t. Including y_t as a predictor while the true model is (A.5) is inconsequential under the null hypothesis, because the true value for α is zero. Things are quite different when instabilities are present. When the dynamic model is used, with $\alpha^*_{[0,r]} = p\lim_{T\to\infty}\hat{\alpha}_{[1,t]}$, the expected value of $L^*(r)$ is:

$$\begin{aligned} E[L^*(r)] &= E\left[p\lim_{T\to\infty}(y_{t+\tau} - \hat{\alpha}_{[1,t]}y_t - x_t\hat{\beta}_{[1,t]})^2\right], \\ &= (1+\alpha^{*2}_{[0,r]})\sigma_r^{*2} + \sigma_x^2[(1-\alpha^*_{[0,r]})\beta_r - \beta^*_{[0,r]}]^2. \end{aligned} \tag{A.7}$$

The case of a coefficient change. Consider the case of coefficient change, i.e., $\beta_1 \neq \beta_2$ with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. For the static model, we obtain the following results for the limit of the coefficient estimates. With the fixed scheme, $\beta^*_{[0,r]} = \beta_1$, for $0 \le r \le 1$. With the rolling scheme,

$$\beta^*_{[0,r]} = \begin{cases} \beta_1 & \text{for } 0 \le r \le \lambda_0, \\ \frac{\lambda_0 - r + \lambda}{\lambda}\beta_1 + \frac{r-\lambda_0}{\lambda}\beta_2 & \text{for } \lambda_0 < r \le \lambda_0 + \lambda \ , \\ \beta_2 & \text{for } \lambda_0 + \lambda < r \le 1. \end{cases}$$

With the recursive scheme,

$$\beta^*_{[0,r]} = \begin{cases} \beta_1 & \text{for } 0 \le r \le \lambda_0, \\ \frac{\lambda_0}{r}\beta_1 + \frac{r-\lambda_0}{r}\beta_2 & \text{for } \lambda_0 < r \le 1. \end{cases}$$

Note that the coefficient estimate is contaminated when the rolling and the recursive schemes are used. When the dynamic model is used, the results are more complex, so we focus on the case with $\beta_1 = 0$ and $\beta_2 = \Delta_\beta$, without loss of generality With the fixed scheme,

$$\begin{bmatrix} \hat{\alpha}_{[1,t]} \\ \hat{\beta}_{[1,t]} \end{bmatrix} = \begin{bmatrix} m^{-1}\sum_{s=1}^{m-\tau} y_s^2 & m^{-1}\sum_{s=1}^{m-\tau} x_s y_s \\ m^{-1}\sum_{s=1}^{m-\tau} x_s y_s & m^{-1}\sum_{s=1}^{m-\tau} x_s^2 \end{bmatrix}^{-1} \begin{bmatrix} m^{-1}\sum_{s=1}^{m-\tau} y_s y_{s+\tau} \\ m^{-1}\sum_{s=1}^{m-\tau} x_s y_{s+\tau} \end{bmatrix}.$$

Using $m^{-1}\sum_{s=1}^{m-\tau} y_s^2 \xrightarrow{p} \sigma^2$, $m^{-1}\sum_{s=1}^{m-\tau} x_s y_s \xrightarrow{p} 0$, $m^{-1}\sum_{s=1}^{m-\tau} x_s^2 \xrightarrow{p} \sigma_x^2$, $m^{-1}\sum_{s=1}^{m-\tau} y_s y_{s+\tau} \xrightarrow{p} 0$, and $m^{-1}\sum_{s=1}^{m-\tau} x_s y_{s+\tau} \xrightarrow{p} 0$,

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma_x^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{for } 0 \le r \le 1.$$

For the rolling scheme, let $\phi = (r - \lambda_0)/\lambda$. After some algebra, we obtain:

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{for } 0 \le r \le \lambda_0, \\ \begin{bmatrix} \frac{\sigma_x^2 \sigma_{x\tau} \phi(1-\phi)\Delta_\beta^2}{\sigma_x^2\sigma^2 + \sigma_x^2\phi(1-\phi)\Delta_\beta^2} \\ \frac{\sigma_{x\tau}\sigma^2\phi\Delta_\beta}{\sigma_x^2\sigma^2 + \sigma_x^2\phi(1-\phi)\Delta_\beta^2} \end{bmatrix} & \text{for } \lambda_0 < r \le \lambda_0 + \lambda, \\ \begin{bmatrix} 0 \\ \Delta_\beta \end{bmatrix} & \text{for } \lambda_0 + \lambda < r \le 1. \end{cases}$$

For the recursive scheme, let $\psi = (r - \lambda_0)/r$. After some algebra, we obtain:

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{for } 0 \le r \le \lambda_0, \\ \begin{bmatrix} \frac{\sigma_x^2 \sigma_{x\tau} \psi(1-\psi)\Delta_\beta^2}{\sigma_x^2\sigma^2 + \sigma_x^2\psi(1-\psi)\Delta_\beta^2} \\ \frac{\sigma_{x\tau}\sigma^2\psi\Delta_\beta}{\sigma_x^2\sigma^2 + \sigma_x^2\psi(1-\psi)\Delta_\beta^2} \end{bmatrix} & \text{for } \lambda_0 < r \le 1. \end{cases}$$

As in (A.6), when the static model is used, $E\left[L^*(r)\right] = \sigma^2 + \sigma_x^2 \Gamma(r)$, where $\Gamma(r) = (\beta_r - \beta^*_{[0,r]})^2$. Therefore, i) with the fixed scheme,

$$\Gamma(r) = \begin{cases} 0 & \text{for } 0 < r \leq \lambda_0, \\ \Delta_\beta^2 & \text{for } \lambda_0 < r \leq 1. \end{cases}$$

ii) with the rolling scheme,

$$\Gamma(r) = \begin{cases} 0 & \text{for } 0 \leq r \leq \lambda_0, \\ \left(\frac{\lambda_0 - r + \lambda}{\lambda}\beta_1 + \frac{r - \lambda_0}{\lambda}\beta_2 - \beta_2\right)^2 = \frac{(\lambda_0 - r + \lambda)^2}{\lambda^2}\Delta_\beta^2 & \text{for } \lambda_0 < r \leq \lambda_0 + \lambda, \\ 0 & \text{for } \lambda_0 + \lambda < r \leq 1. \end{cases}$$

iii) with the recursive scheme,

$$\Gamma(r) = \begin{cases} 0 & \text{for } 0 \leq r \leq \lambda_0, \\ \left(\frac{\lambda_0}{r}\beta_1 + \frac{r - \lambda_0}{r}\beta_2 - \beta_2\right)^2 = \frac{\lambda_0^2}{r^2}\Delta_\beta^2 & \text{for } \lambda_0 < r \leq 1. \end{cases}$$

When the dynamic model is used, we have: i) with the fixed scheme,

$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \leq r \leq \lambda_0, \\ \sigma^2 + \sigma_x^2 \Delta_\beta^2 & \text{for } \lambda_0 < r \leq 1. \end{cases}$$

ii) With the rolling scheme,

$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \leq r \leq \lambda_0, \\ D_{rol}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) & \text{for } \lambda_0 < r \leq \lambda_0 + \lambda, \\ \sigma^2 & \text{for } \lambda_0 + \lambda < r \leq 1, \end{cases}$$

where $D_{rol}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) \equiv N_{rol}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) / [\sigma_x^2 \sigma^2 + \sigma_x^2 \phi(1-\phi)\Delta_\beta^2]^2$, with

$$\begin{aligned} & N_{rol}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) \\ = \; & \sigma_x^4 \sigma^6 + [2\sigma_x^2 \phi(1-\phi) + (\sigma_x^2 - \sigma_{x\tau}\phi)^2]\sigma_x^2 \sigma^4 \Delta_\beta^2 \\ & + [\sigma_{x\tau}^2 \phi(1-\phi) + 2(\sigma_x^2 - \sigma_{x\tau}\phi)(1 - \sigma_{x\tau}) + \phi(1-\phi)]\phi(1-\phi)\sigma_x^4 \sigma^2 \Delta_\beta^4 \\ & + (1 - \sigma_{x\tau})^2 \phi^2 (1-\phi)^2 \sigma_x^6 \Delta_\beta^4. \end{aligned}$$

and iii) with the recursive scheme,

$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \leq r \leq \lambda_0, \\ D_{rec}(\sigma, \psi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) & \text{for } \lambda_0 < r \leq 1, \end{cases}$$

where $D_{rec}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) \equiv N_{rec}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x\tau})/[\sigma_x^2\sigma^2 + \sigma_x^2\phi(1-\phi)\Delta_\beta^2]^2$ with

$$\begin{aligned} & N_{rec}(\sigma, \phi, \Delta_\beta, \sigma_x, \sigma_{x,\tau}) \\ = \; & \sigma_x^4\sigma^6 + [2\sigma_x^2\psi(1-\psi) + (\sigma_x^2 - \sigma_{x\tau}\psi)^2]\sigma_x^2\sigma^4\Delta_\beta^2 \\ & +[\sigma_{x\tau}^2\psi(1-\psi) + 2(\sigma_x^2 - \sigma_{x\tau}\psi)(1-\sigma_{x\tau}) + \psi(1-\psi)]\psi(1-\psi)\sigma_x^4\sigma^2\Delta_\beta^4 \\ & +(1-\sigma_{x\tau})^2\psi^2(1-\psi)^2\sigma_x^6\Delta_\beta^4. \end{aligned}$$

The case of a variance change. We now consider the case of variance change, i.e., $\sigma_1^2 = \sigma^2$ and $\sigma_2^2 = \sigma^2 + \Delta_{\sigma^2}$ with $\beta_1 = \beta_2 = \beta$. It is relatively easy to derive the limit of the coefficient estimate and of the loss sequence. Under any window schemes, we can show that: i) when the static model is used $\beta^*_{[0,r]} = \beta$, for $0 \le r \le 1$, and when the dynamic model is used,

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{bmatrix} 0 \\ \beta \end{bmatrix} \quad \text{for } 0 \le r \le 1.$$

Therefore, under any window schemes, when the static model is used, (A.6) yields

$$\begin{aligned} E[L^*(r)] &= \sigma_r^2 + \sigma_x^2(\beta_r - \beta^*_{[0,r]})^2 \\ &= \sigma_r^2 \\ &= \begin{cases} \sigma^2 & \text{for } 0 \le r \le \lambda_0, \\ \sigma^2 + \Delta_{\sigma^2} & \text{for } \lambda_0 < r \le 1, \end{cases} \end{aligned}$$

and when the dynamic model is used,

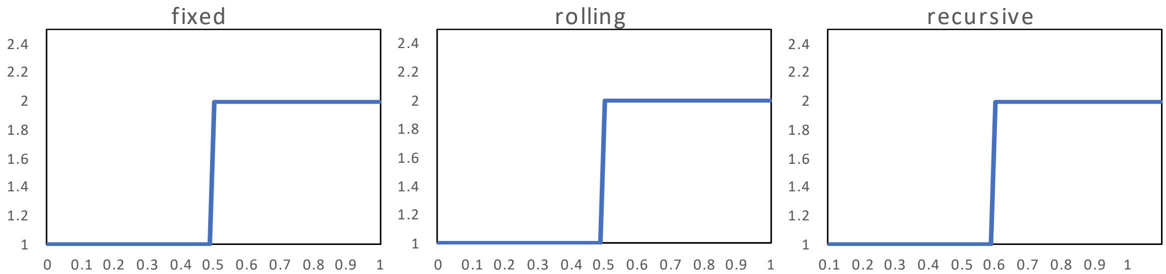
$$\begin{aligned} E[L^*(r)] &= (1 + \alpha^{*2}_{[0,r]})\sigma_r^2 + \sigma_x^2[(1-\alpha^*_{[0,r]})\beta_r - \beta^*_{[0,r]}]^2 \\ &= \sigma_r^2 \\ &= \begin{cases} \sigma^2 & \text{for } 0 \le r \le \lambda_0, \\ \sigma^2 + \Delta_{\sigma^2} & \text{for } \lambda_0 < r \le 1. \end{cases} \end{aligned}$$

Numerical illustration. We compute the numerical values of $E[L^*(r)]$ for $0 \le r \le 1$. We set $\sigma^2 = 1$, $\lambda = 0.1$ and $\lambda_0 = 0.5$ and a small break $(\Delta_\beta, \Delta_{\sigma^2} = 1)$ and a large break $(\Delta_\beta, \Delta_{\sigma^2} = 5)$ to investigate how the magnitude of the break affects the shape of the loss sequence. Figures 1 and 2 present $E[L^*(r)]$ for the coefficient change case for the static and dynamic regressions, respectively. The upper (lower) panels report the case of a small (large) break. When the static model is used (Figure 1), we see a stepwise change in the loss sequences when the fixed scheme is used. However, the same change in coefficient translates into a spiked shape with the rolling scheme and a triangular shape with the recursive scheme. This explains why the fixed scheme is to be preferred and why using the rolling or recursive scheme induces a loss of power. Note, however, that the break magnitude only changes the height of the change in the loss sequence, not the shapes of the loss sequences. Hence, increasing the magnitude of the break size should still increase power under all schemes.

Things are different when considering the dynamic model, We still have the same general shapes for the loss sequences. However, the spikes and the triangular shape for the rolling and the recursive schemes become more narrow and closer to an outlier as the break becomes larger. This explains why, when using a dynamic model, the use of a rolling or recursive scheme leads to a non-monotonic power function, i.e., the power decreases as the break magnitude increases.

Figures S.1 and S.2 present $E[L^*(r)]$ for the variance change case when the static and dynamic regressions are used, respectively. Here, the results are simple. The limit of the loss sequence always yields a stepwise change, whose magnitude depends on the change in variance. Hence, all tests should have similar power functions that are monotonically increasing in the magnitude of the variance break.

Figure S.1. Shapes of $E[L^*(r)]$ for a variance break in a static regression
a) small break ($\Delta_{\sigma^2} = 1$):



b) large break ($\Delta_{\sigma^2} = 5$):

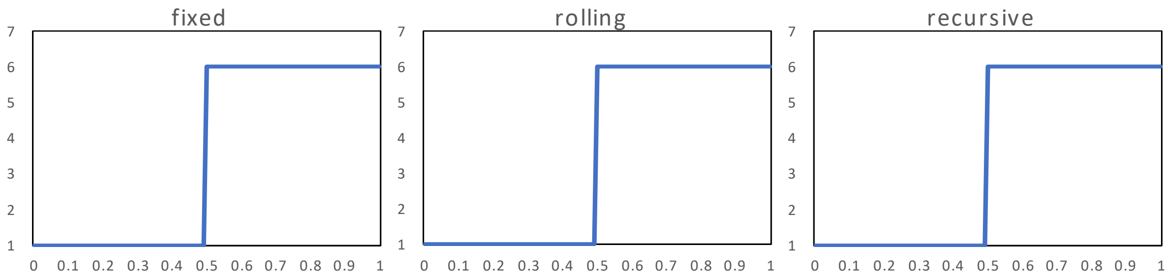
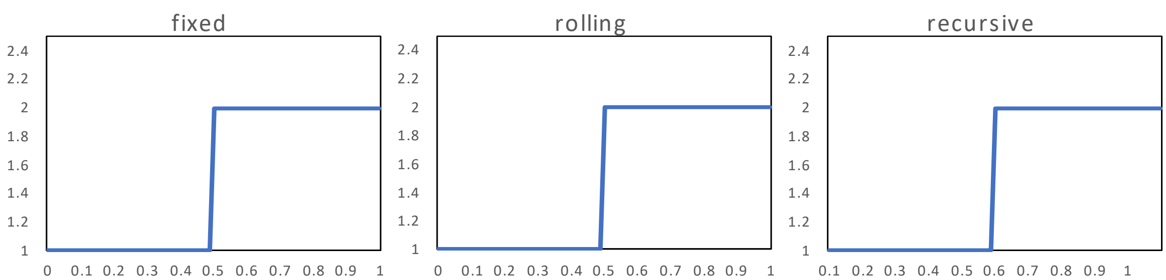
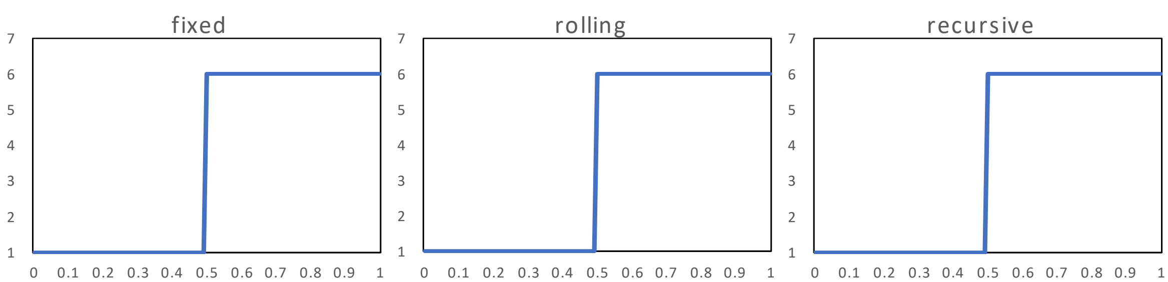


Figure S.2. Shapes of $E[L^*(r)]$ for a variance break in a dynamic regression
a) small break ($\Delta_{\sigma^2} = 1$):



b) large break ($\Delta_{\sigma^2} = 5$):



g) Theoretical results for the iterated forecast method

We assess the properties of the loss sequence when the iterated forecasting method is used. We consider the case of a coefficient change with no extra predictors x_t, i.e., an unconditional forecast, to avoid the potential complication of forecasting x_t. The data are generated by

$$y_t = \beta_t + e_t, \quad \text{for } t = 1, ..., T,$$

where

$$\beta_t = \begin{cases} 0 & \text{for } t \leq [T\lambda_0], \\ \Delta_\beta & \text{for } t > [T\lambda_0], \end{cases}$$

and e_t is a white noise with mean 0 and variance σ^2. Suppose we use an AR1 model, in which the true value of α is zero for simplicity and without loss of generality, so that the regression is:

$$y_{t+1} = \beta + \alpha y_t + error.$$

After we obtain the OLS coefficient estimate $\hat{\alpha}_{[1,t]}$ and $\hat{\beta}_{[1,t]}$, the iterative method constructs a $\tau-$period ahead forecast as follows

$$\begin{aligned} \hat{y}_{t+\tau} &= \hat{\beta}_{[1,t]} + \hat{\alpha}_{[1,t]} y_{t+\tau-1}, \\ &= \hat{\beta}_{[1,t]}(1 + \hat{\alpha}_{[1,t]}) + \hat{\alpha}^2_{[1,t]} y_{t+\tau-2}, \\ &\vdots \\ &= \hat{\beta}_{[1,t]} \overline{\hat{\alpha}}_{[1,t]} + \hat{\alpha}^\tau_{[1,t]} y_t, \end{aligned}$$

where $\overline{\hat{\alpha}}_{[1,t]} \equiv \sum_{j=0}^{\tau} \hat{\alpha}^j_{[1,t]}$. Because τ is asymptotically small, the expected value of $L^*(r)$ becomes

$$\begin{aligned} E\left[L^*(r)\right] &= E\left[p \lim_{T\to\infty} (\hat{y}_{t+\tau} - \hat{\beta}_{[1,t]} \overline{\hat{\alpha}}_{[1,t]} + \hat{\alpha}^\tau_{[1,t]} y_t)^2\right], \\ &= (1 + \alpha^{*2\tau}_{[0,r]})\sigma^{*2}_r + [(1 - \alpha^{*\tau}_{[0,r]})\beta_r - \beta^*_{[0,r]} \overline{\hat{\alpha}}^*_{[0,r]}]^2, \end{aligned} \tag{A.8}$$

where $\alpha^*_{[0,r]} = p \lim_{T\to\infty} \hat{\alpha}_{[1,t]}$, $\beta^*_{[0,r]} = p \lim_{T\to\infty} \hat{\beta}_{[1,t]}$ and $\overline{\hat{\alpha}}^*_{[0,r]} = p \lim_{T\to\infty} \sum_{j=0}^{\tau} \alpha^{*j}_{[0,r]}$. Note that we now have an expression for $E[L^*(r)]$ influenced by τ, which did not appear in the direct forecast counterparts; see (A.7). Hence, it is interesting to assess the effects of τ. The limit of the coefficient estimate is the same as that for the dynamic regression model with $x_t = 1$ for all t since τ is fixed and, hence, small in large samples. With the fixed scheme,

$$\begin{bmatrix} \hat{\alpha}_{[1,t]} \\ \hat{\beta}_{[1,t]} \end{bmatrix} = \begin{bmatrix} m^{-1} \sum_{s=1}^{m-1} y_s^2 & m^{-1} \sum_{s=1}^{m-1} y_s \\ m^{-1} \sum_{s=1}^{m-1} y_s & m^{-1} \sum_{s=1}^{m-1} 1 \end{bmatrix}^{-1} \begin{bmatrix} m^{-1} \sum_{s=1}^{m-1} y_s y_{s+\tau} \\ m^{-1} \sum_{s=1}^{m-1} y_{s+\tau} \end{bmatrix}.$$

We have $m^{-1} \sum_{s=1}^{m-1} y_s^2 \xrightarrow{p} \sigma^2$, $m^{-1} \sum_{s=1}^{m-1} y_s \xrightarrow{p} 0$, $m^{-1} \sum_{s=1}^{m-1} 1 \xrightarrow{p} 1$, $m^{-1} \sum_{s=1}^{m-1} y_s y_{s+1} \xrightarrow{p} 0$ and $m^{-1} \sum_{s=1}^{m-1} y_{s+1} \xrightarrow{p} 0$ so that

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma_x^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{for } 0 \leq r \leq 1.$$

With the rolling scheme, we let $\phi = (r - \lambda_0)/\lambda$ and obtain

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{for } 0 \le r \le \lambda_0, \\ \begin{bmatrix} \frac{\phi(1-\phi)\Delta_\beta^2}{\sigma^2+\phi(1-\phi)\Delta_\beta^2} \\ \frac{\sigma^2\phi\Delta_\beta}{\sigma^2+\phi(1-\phi)\Delta_\beta^2} \end{bmatrix} & \text{for } \lambda_0 < r \le \lambda_0 + \lambda, \\ \begin{bmatrix} 0 \\ \Delta_\beta \end{bmatrix} & \text{for } \lambda_0 + \lambda < r \le 1. \end{cases}$$

With the recursive scheme, we let $\psi = (r - \lambda_0)/r$ and obtain

$$\begin{bmatrix} \alpha^*_{[0,r]} \\ \beta^*_{[0,r]} \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{for } 0 \le r \le \lambda_0, \\ \begin{bmatrix} \frac{\psi(1-\psi)\Delta_\beta^2}{\sigma^2+\psi(1-\psi)\Delta_\beta^2} \\ \frac{\sigma^2\psi\Delta_\beta}{\sigma^2+\psi(1-\psi)\Delta_\beta^2} \end{bmatrix} & \text{for } \lambda_0 < r \le 1. \end{cases}$$

We can then obtain $E\left[L^*(r)\right]$ by plugging $\alpha^*_{[0,r]}$ and $\beta^*_{[0,r]}$ into (A.8). With the fixed scheme,

$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \le r \le \lambda_0, \\ \sigma^2 + \Delta_\beta^2 & \text{for } \lambda_0 < r \le 1. \end{cases}$$

With the rolling scheme,

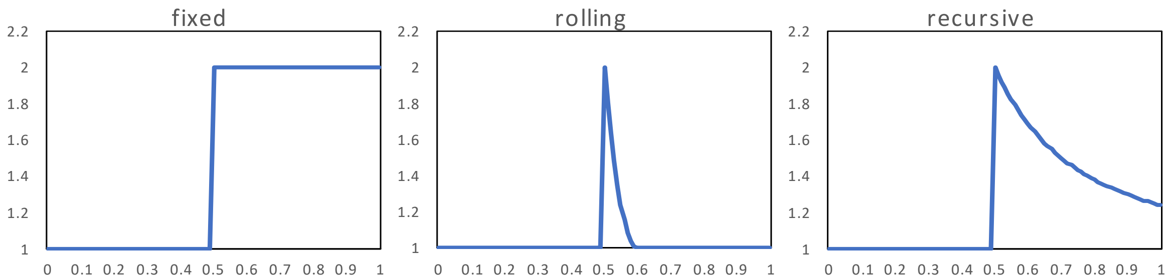
$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \le r \le \lambda_0, \\ R_{rol}(\sigma, \phi, \Delta_\beta) & \text{for } \lambda_0 < r \le \lambda_0 + \lambda, \\ \sigma^2 & \text{for } \lambda_0 + \lambda < r \le 1, \end{cases}$$

where $R_{rol}(\sigma, \phi, \Delta_\beta) \equiv (1 + \alpha^{*2\tau}_{[0,r]})\sigma^2 + [(1 - \alpha^{*\tau}_{[0,r]})\Delta_\beta - \beta^*_{[0,r]}\overline{\tilde{\alpha}}^*_{[0,r]}]^2$ with $\alpha^*_{[0,r]} = [\phi(1 - \phi)\Delta_\beta^2]/[\sigma^2 + \phi(1-\phi)\Delta_\beta^2]$ and $\beta^*_{[0,r]} = [\sigma^2\phi\Delta_\beta]/[\sigma^2 + \phi(1-\phi)\Delta_\beta^2]$. With the recursive scheme,

$$E\left[L^*(r)\right] = \begin{cases} \sigma^2 & \text{for } 0 \le r \le \lambda_0, \\ R_{rec}(\sigma, \psi, \Delta_\beta) & \text{for } \lambda_0 < r \le 1, \end{cases}$$

where $R_{rec}(\sigma, \phi, \Delta_\beta) \equiv (1 + \alpha^{*2\tau}_{[0,r]})\sigma^2 + [(1 - \alpha^{*\tau}_{[0,r]})\Delta_\beta - \beta^*_{[0,r]}\overline{\tilde{\alpha}}^*_{[0,r]}]^2$ with $\alpha^*_{[0,r]} = [\psi(1 - \psi)\Delta_\beta^2]/[\sigma^2 + \psi(1 - \psi)\Delta_\beta^2]$ and $\beta^*_{[0,r]} = [\sigma^2\psi\Delta_\beta]/[\sigma^2 + \psi(1 - \psi)\Delta_\beta^2]$. Since the expressions above are not quite intuitive, we compute the numerical values of $E[L^*(r)]$ for $0 \leq r \leq 1$. In particular, we set the parameter values $\sigma^2 = 1$, $\lambda = 0.1$ and $\lambda_0 = 0.5$. We consider the forecasting horizons $\tau = 3$ and $\tau = 12$. For each value of τ, we consider a small break ($\Delta_\beta = 1$) and a large break ($\Delta_\beta = 5$) to investigate how the magnitude of the break affects the shape of the loss sequence. Figures S.3 presents $E[L^*(r)]$ for the coefficient change case when $\tau = 3$. The upper panel reports the case of a small break and the lower panel that of a large break. The pattern of the loss sequence is very similar to what was obtained using the direct forecasting (dynamic regression) in that the change in the coefficient translates into a spike when the rolling or the recursive schemes are used. For the direct forecast method, τ does not appear in the results as it is assumed small relative to T. Figure S.4 presents $E[L^*(r)]$ for the coefficient change case when $\tau = 12$ and the results are very similar to those of $\tau = 3$, except that the effect of increasing the break magnitude on the shapes of $E[L^*(r)]$ for the rolling and the recursive schemes is somewhat weaker when τ is larger. Otherwise, all qualitative results reported in the text for the "direct forecast method" continue to hold if one uses the "indirect forecast method".

Figure S.3. Shapes of $E[L^*(r)]$: Coefficient break with iterated forecasts ($\tau = 3$)
a) Small break ($\Delta_\beta = 1$)



b) large break ($\Delta_\beta = 5$)

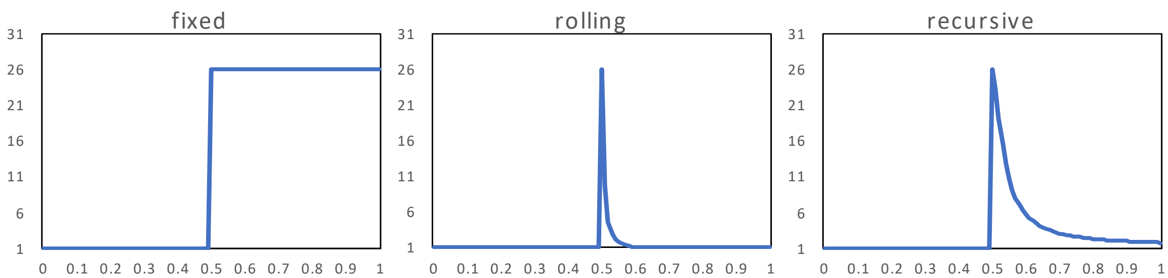
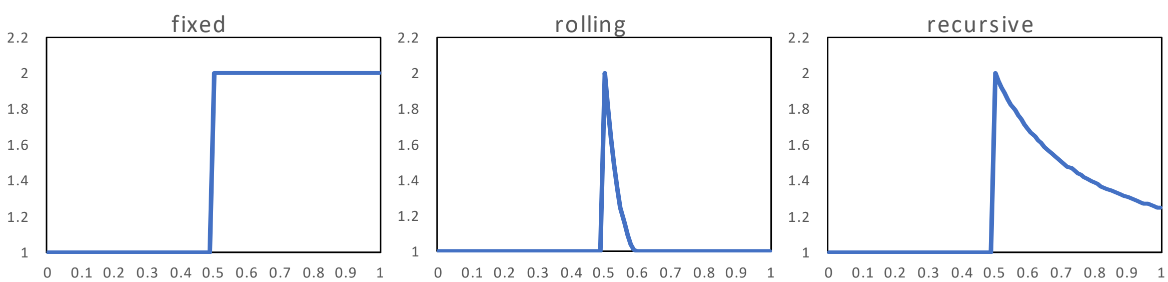
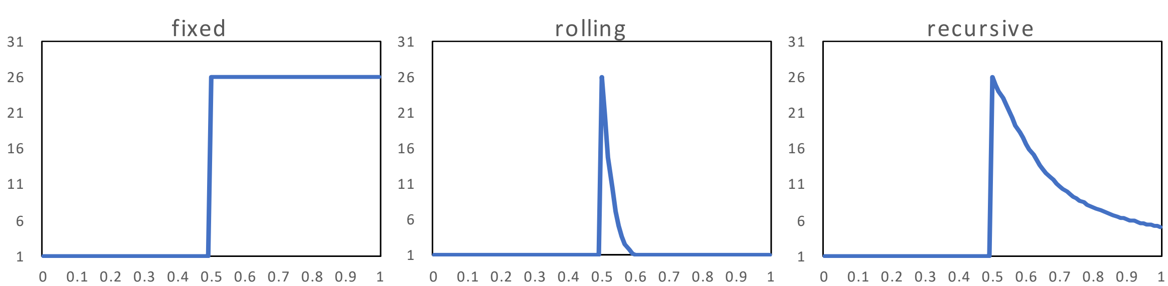


Figure S.4. Shapes of $E[L^*(r)]$: Coefficient break with iterated forecasts ($\tau = 12$)
a) Small break ($\Delta_\beta = 1$)



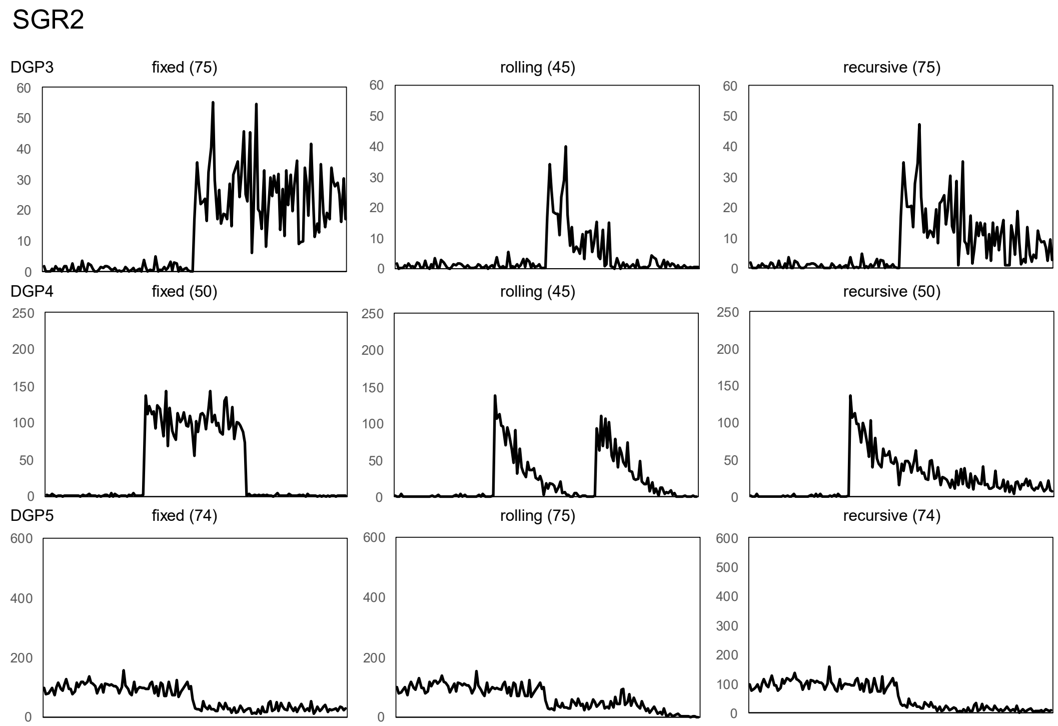
b) Large break ($\Delta_\beta = 5$)



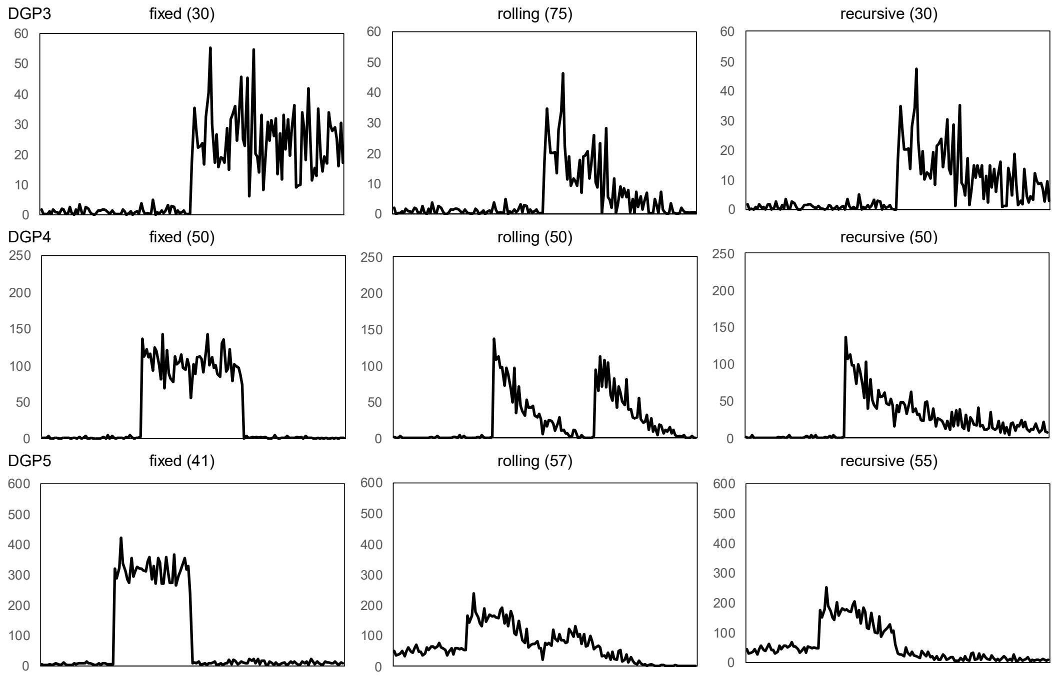
h) A "typical" realization of the loss sequences

Figures S.5 and S.6 present a "typical" realization of the loss sequences for DGPs 3-5 under the three forecasting schemes for the tests $SGR2$, $DSW2$ and $TLSW$ (the results using $SGR1$, $DSW1$ and $TLUD$ are, respectively, almost equivalent and, hence, omitted). Because the loss sequence is generated for every m, we present the one for which the test statistic is maximized, say m^*, whose value is indicated in parenthesis above each path.

Figure S.5: A realization of loss sequences: static model



DSW2


DGP3
fixed (30)
rolling (75)
recursive (30)
DGP4
fixed (50)
rolling (50)
recursive (50)
DGP5
fixed (41)
rolling (57)
recursive (55)

TLSW

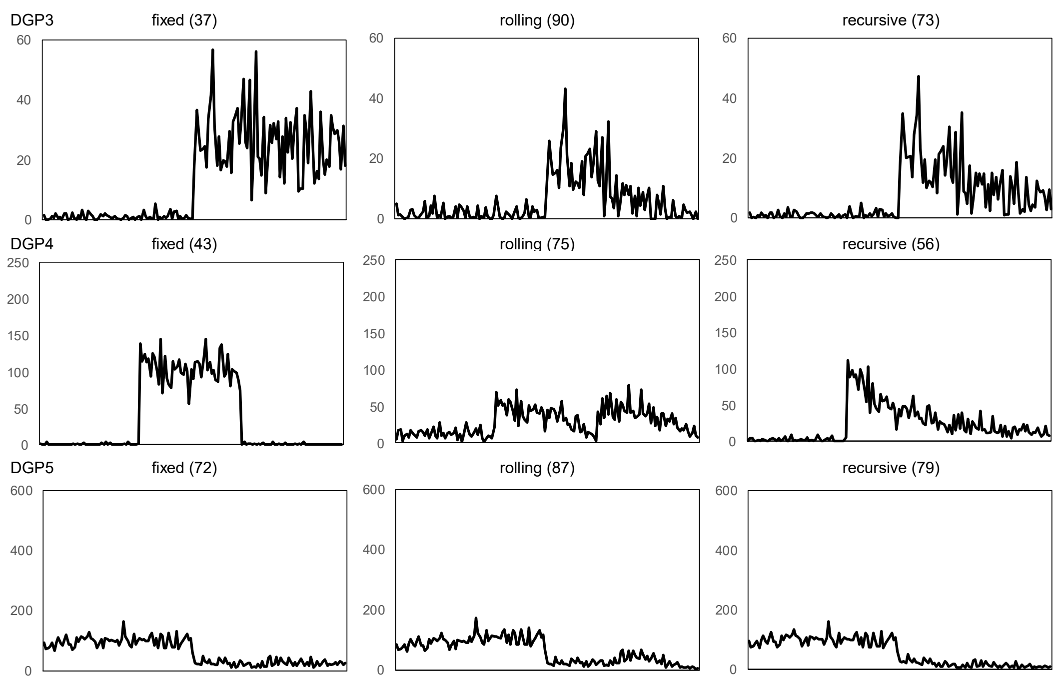
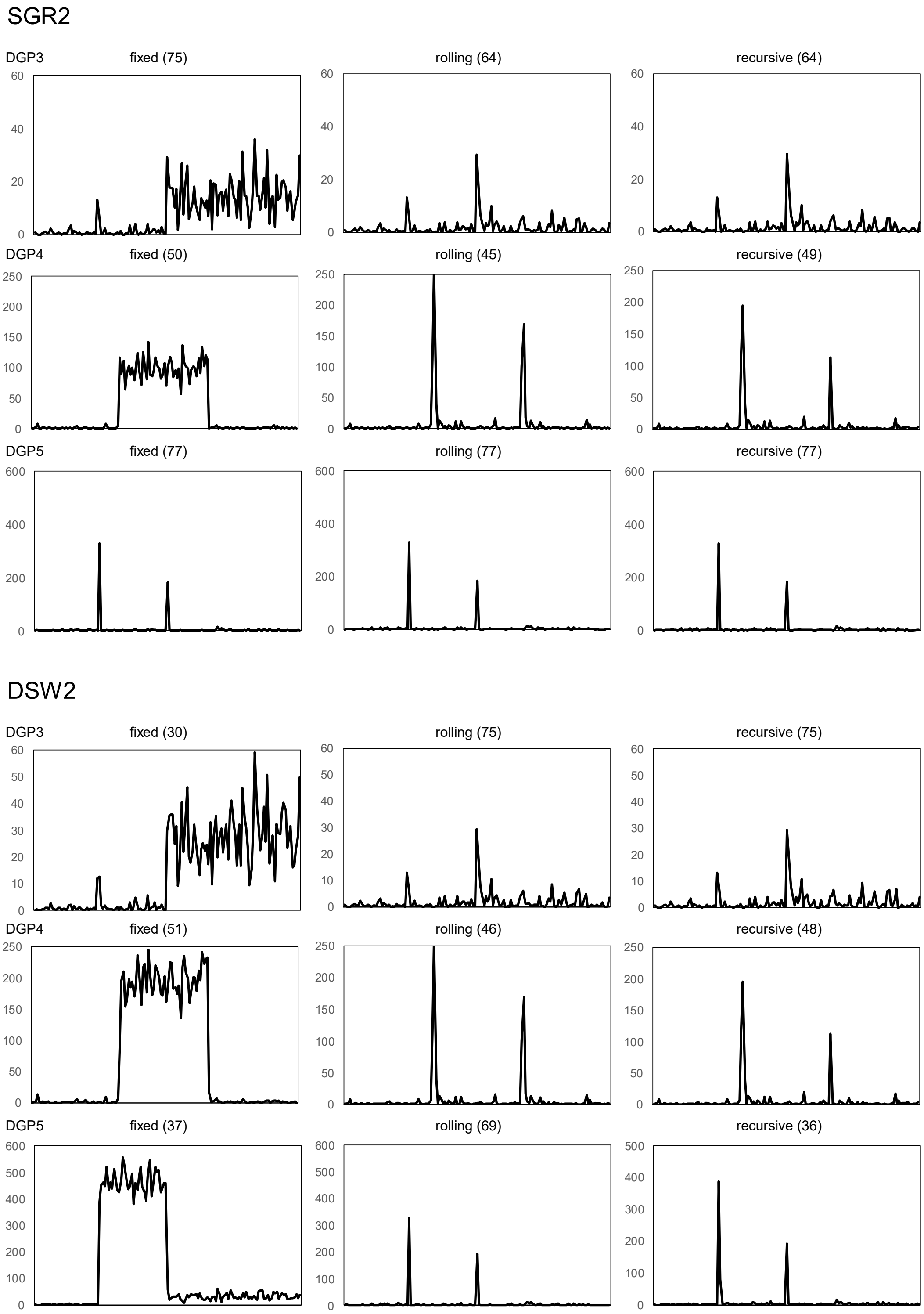
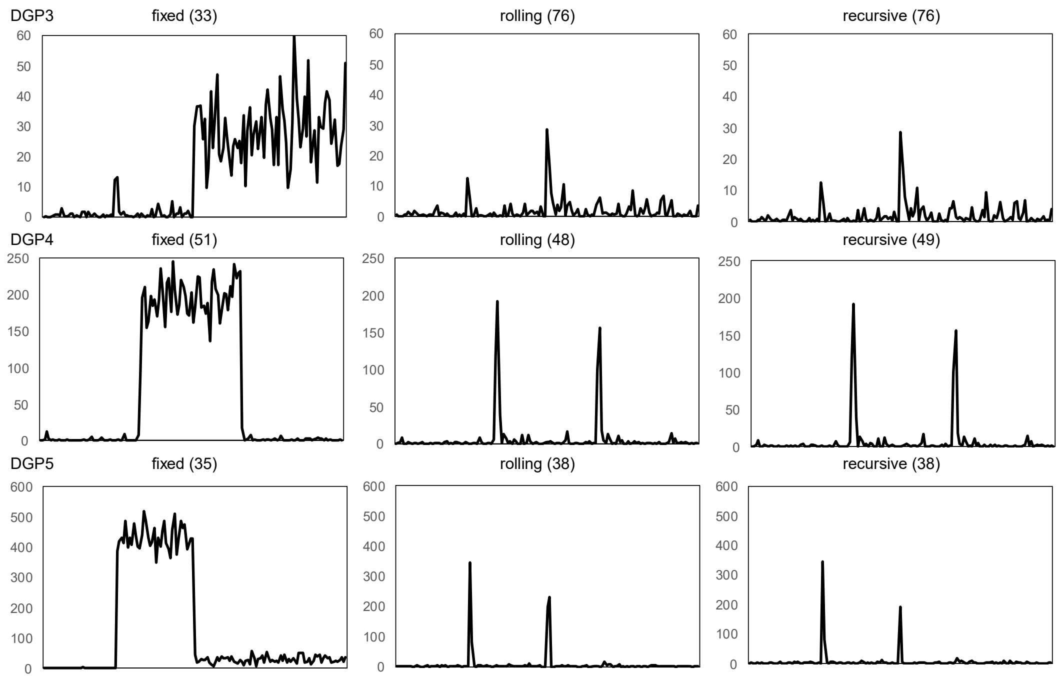

DGP3
fixed (37)
rolling (90)
recursive (73)
DGP4
fixed (43)
rolling (75)
recursive (56)
DGP5
fixed (72)
rolling (87)
recursive (79)

Figure S.6: A realization of loss sequences: dynamic model



TLSW


DGP3
fixed (33)
rolling (76)
recursive (76)
DGP4
fixed (51)
rolling (48)
recursive (49)
DGP5
fixed (35)
rolling (38)
recursive (38)