Graduate School of Economics, Hitotsubashi University
Discussion Paper Series No. 2018-03

Testing for Changes in Forecasting Performance

Pierre Perron Yohei Yamamoto

May 2018

Testing for Changes in Forecasting Performance

Pierre Perron* **Yohei Yamamoto†**

Boston University **Hitotsubashi University**

May 24, 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 all possible 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) test and the Total Loss UDmax (TLUD) test. Using extensive 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 are the DSW and TLUD tests, constructed with a fixed forecasting window scheme. Some explanations are provided and two empirical applications illustrate the relevance of our findings in practice.

JEL Classification Number: C14, C22

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

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

†Hitotsubashi University, Department of Economics, 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. To this end, it is useful to clarify the purpose of forecast breakdown tests. 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-sample and out-of-sample periods simply 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 then the need for a somewhat artificial separation between the in and out-of-sample periods at some date labelled m, say. This separation date should be such that the model in the in-sample period is stable in some sense, e.g., yielding stable forecasts. This can, however, create problems; e.g., one needs a truncation point m to assess forecast failures but the choice of this value is itself predicated on some knowledge of stability.

The forecast failure test of Giacomini and Rossi (2009), GR hereafter, is a global and retrospective test which compares the in-sample average with the out-of-sample average of the sequence of forecast losses. We shall discuss this test in some details, including extensions and potential problems. Casini (2017) extends the analysis by considering a continuous-time asymptotic framework and partitioning the out-of-sample into $m_T = \lfloor T_n/n_T \rfloor$ blocks each containing n_T observations. See Casini and Perron (2018) for a review of recent developments and Perron (2006) for review of most issues discussed in this paper.

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, which ensures the maximum difference between the fitted in and out-of sample means of the losses. There are contamination issues under the rolling and recursive schemes that induce power losses. With such a fixed scheme, GR's 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 m. To alleviate this problem, which leads to important losses in power when the break in forecasting performance is not exactly at m, one can follow Inoue and Rossi (2012) and

consider maximizing the GR test over all possible values of m within a pre-specified range. This then corresponds to a sup-Wald test for a single change at some date constrained to be the separation point between 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 of such tests 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 loss series for some forecast horizon τ, $L_t^o(\hat{\beta})$ for $t = m + \tau, \ldots, T$, defined by

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

where $n = T-m-\tau+1$, $SSR_{L^o(m)}$ is the unrestricted sum of squared residuals, $SSR\left(T_b\left(m\right)\right)_{L^o(m)}$ is the sum of squared residuals assuming a one-time change at time $T_b\left(m\right)$, and $\hat{V}_{L^o(m)}$ is the long-run variance estimate of the out-of-sample loss series. In addition, we propose to work directly with the total loss series $L\left(m\right)$ to define the Total Loss sup-Wald test ($TLSW$) and the Total Loss UDmax test ($TLUD$). Using extensive simulations, based on the original design of GR, which involves single and multiple changes in the regression parameters and/or the variance of the errors, we show that with forecasting models potentially involving lagged dependent variables, 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 paper is structured as follows. Section 2 provides the statistical framework adopted and the tests considered. Section 2.1 reviews the case of a single break occurring at a known date to which the GR test applies. Section 2.2 provides a discussion of the choice of the forecasting scheme and why using a fixed scheme is preferable. Section 2.3 considers the case with unknown break dates and describes new tests to be considered. The limit distribution of some proposed tests are contained in Section 3. Section 4 offers a comprehensive simulation analysis of all tests under the three forecasting schemes. Section 4.1 pertains to the finite sample size of the proposed tests. Section 4.2 describes the setup considered to evaluate the power functions. Section 4.3 provides a summary of the main results, while Section 4.4 expands on the sources of various non-monotonic power functions. Section 5 provides two empirical applications to illustrate the relevance of our findings in practice. One relates to forecasting oil prices and the other to U.S. inflation using a Phillips curve model. Section 6 provides brief concluding remarks. An appendix contains some technical derivations.

2 The framework and the tests

Suppose that we have data (y_t, x_t) where y_t is a scalar variable to be forecasted and x_t is 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

$$\hat{y}_{t+\tau} = f_t(\hat{\beta}_m; x_t)$$

where $\hat{\beta}_m$ is the estimate of the parameter vector β $(q \times 1)$ obtained using an in-sample window of size $m \geq q$. The out-of-sample forecast procedure basically divides the sample from $t = 1, ..., T$ 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 i) fixed window (with the in-sample consisting of observations 1 to m), ii) rolling window (with the in-sample consisting of observations $t - m + 1$ to t), and iii) recursive window (with the in-sample consisting of observations 1 to t). We define the sequence of in-sample losses as $L_t^i(\hat{\beta}_m)$, defined by the in-sample fitted values of $\hat{y}_t = f_t(\hat{\beta}_m; x_t)$, and the sequence of out-of-sample losses $L_t^o(\hat{\beta}_m)$, defined by the forecast values $\hat{y}_{t+\tau}$. A popular loss function is the squared error loss, in which case $L_t^i(\hat{\beta}_m) = (y_t - f_t(\hat{\beta}_m; x_t))^2$ and $L_{t+\tau}^o(\hat{\beta}_m) = (y_{t+\tau} - \hat{y}_{t+\tau})^2$. It will also be convenient to define i) the in-sample loss sequence: $L^i(m) = (L_1^i(\hat{\beta}_m), ..., L_m^i(\hat{\beta}_m))$; ii) the out-of-sample loss sequence: $L^o(m) = (L_{m+\tau}^o(\hat{\beta}_m), ..., L_T^o(\hat{\beta}_m))$ and iii) the total loss series as the stacked vector of the in-sample and out-of-sample losses, i.e., $L(m) = (L_1^i(\hat{\beta}_m), ..., L_m^i(\hat{\beta}_m),\ L_{m+\tau}^o(\hat{\beta}_m), ..., L_T^o(\hat{\beta}_m))$, a $(T - \tau + 1) \times 1$ vector. When we time-index the total loss series, we denote it by $\{L_{t+\tau}\}_{t=1}^{T-\tau}$.

The goal is to assess whether there are instabilities in the forecast accuracy. An example is a deterioration in forecast accuracy, 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 from the numerator of the test statistic. In the case of a linear forecasting model and a quadratic loss function, GR showed that the appropriate correction factor is $\phi = 2\kappa \times tr\left[T^{-1}\sum_{t=1}^{T}(x_t x_t') \times AVar(\hat{\beta})\right]$,

where $\kappa = n^{1/2}/m$ for the fixed and rolling schemes and $\kappa = n^{-1/2}\log(1+n/m)$ for the recursive scheme.

The null hypothesis that we consider is

$$H_0 : E[L_t(\beta^*)] = \mu_0 \text{ for all } t = 1, ..., T$$

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:

$$H_1 : E[L_t(\beta^*)] \neq E[L_{t+1}(\beta^*)] \text{ for at least one } t = 1, ..., T-1.$$

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

It is useful to consider first the case with a single break in forecast accuracy occurring at some known date T_b, say, so that the alternative is

$$E\left[L_t(\beta^*)\right] = \mu_1 \text{ for } t \leq T_b, \ \mu_2 \text{ for } t > T_b. \tag{1}$$

In this case, the obvious thing to do is to apply a test for a change in the mean of the total loss series occurring at date T_b. This is achieved by setting $m = T_b$ and assessing whether the averages of the in-sample and out-of-sample loss functions are different. This is exactly 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

$$GR_m = \left(n^{-1/2}\sum_{t=m+1}^{T-\tau} SL_{t+\tau}\right)/\hat{\Omega}^{1/2},$$

where $\hat{\Omega}$ 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 test is obviously problematic since one does not know the true break date in practice even if only one break is present. This makes the test very sensitive to the choice of m. As will be shown via simulations, the test can have non-monotonic power (power that decreases as the magnitude of the change in the mean of the losses increases) for a range of choices for m.

2.2 The choice of the forecasting scheme

Before considering tests that are not based on the assumption of a known break date, it is useful to deal with this special case to analyze the relative merits and drawbacks of the

different forecasting schemes: fixed, rolling or recursive. It is well known that to get better forecasts it is, in general, better to adopt a recursive forecasting scheme, even in the presence of instabilities. This is so because the parameter estimates adapt to the underlying data-generating process to fit the data better and thereby provide more accurate forecasts. A fixed forecasting scheme fails to provide such adjustments. A rolling forecasting scheme can provide adjustments but at the expense of increased variability due to a smaller in-sample window, compared to the recursive scheme.

However, when trying to detect 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. Using 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 value of the fitted mean of the out-of-sample losses as the forecast period increases, thereby inducing a loss of power. With a rolling window scheme, the same phenomenon occurs but in a more pronounced fashion since the in-sample fitted mean can eventually reach the post-break mean if the window is small enough.

Hence, when constructing tests for changes in forecast accuracy, it is preferable to use a fixed window. This will remain true even in the case of one or multiple breaks occurring at unknown dates. We will illustrate such issues via simulations later.

2.3 The case with unknown break dates

A simple method to alleviate the dependence of the test GR_m on m is to take the supremum over a range of m, say $[m_0, m_1]$, a version denoted by SGR, viz.

$$SGR = \max_{m \in [m_0, m_1]} GR_m. \tag{2}$$

This test is tailored to the alternative hypothesis (1) with T_b unknown. Since the GR_m test follows a standard normal distribution under the null hypothesis, the limit distribution and critical values of the SGR test can be found in Inoue and Rossi (2012) for some 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 in the sense that 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.

One way to avoid this problem is to adopt the following approach. For each possible value of m, perform a sup-Wald test for a change in the mean of the out-of-sample losses and take the maximum of such tests over a range $m \in [m_0, m_1]$. We call this test the Double sup-Wald (DSW) test. More precisely, it is 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

$$SW_{L^o(m)} = \max_{T_b(m) \in [m+\epsilon n, m+(1-\epsilon)n]} \frac{SSR_{L^o(m)} - SSR(T_b(m))_{L^o(m)}}{\hat{V}_{L^o(m)}}, \tag{3}$$

where $SSR_{L^o(m)}$ is the restricted sum of squared residuals, $SSR(T_b(m))_{L^o(m)}$ is the sum of squared residuals assuming a one-time change at time $T_b(m)$, and $\hat{V}_{L^o(m)}$ is the long-run variance estimate of the out-of-sample loss series constructed using the residuals obtained from the demeaned out-of-sample loss series with the mean changing at date $T_b(m)$. The long-run variance estimate can be obtained using the method of Andrews (1991). Here, ϵ is a small trimming value which we set at 0.1. The limit distribution of the DSW test is derived in the next section. A step-by-step construction of the DSW test is given by: 1) Start with an out-of-sample method with a value of the in-sample length $m = m_0$. Note that m_0 must be small but large enough for one to estimate the model in-sample. Let $n \equiv T - m - \tau + 1$. 2) Compute the out-of-sample loss series $\left\{L_{t+\tau}^o\right\}_{t=m+1}^{T-\tau}$. 3) Consider a regression with only a constant: $L_{t+\tau}^o = \gamma + e_{t+\tau}$. Apply a sup-Wald test for the constancy of γ, usually with a HAC variance estimate if there is evidence of serial correlation in the loss sequence. Store the value as $SW_{L^o(m)}$. 4) Use $m = m + 1$ and repeat Steps 2 and 3 to compute $SW_{L^o(m)}$. Continue up to $m = m_1$. We recommend that the range of m, $m_1 - m_0$, to be some fraction of n. Denote the fraction by $\bar{\mu}$. 5) Take the maximum of the sequence of $\left\{SW_{L^o(m)}\right\}_{m=m_0}^{m_1}$.

The reason why the DSW test improves upon the SGR test is because the procedure acts in such a way as to produce three segments instead of only two with SGR. One segment is defined by m and the other by the date at which the Wald test is maximized in the range $[m + \epsilon n, m + (1 - \epsilon)n]$. This ensures that a segment with the largest possible average loss can be selected, thereby increasing power. The idea is akin to that of Qu (2007) who showed that when searching whether any part of a sample is stationary all that is needed is a search with two breaks defining three segments.

One can also consider a similar test with the UDmax test for multiple structural changes of Bai and Perron (1998). However, the size distortions were rather high. Hence, we shall

not consider it further. An alternative approach is to work directly with the total loss series $L(m)$ instead of restricting the search for a break to the out-of-sample sequence. This can have some power advantages 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 ($TLUD$). They are based on the maximal value of the sup-Wald or UDmax test over the size m of the in-sample window. More precisely,

$$TLSW = \max_{m\in[m_0,m_1]} SW_{L(m)},$$

where $SW_{L(m)}$ is again the sup-Wald test applied for a change in the mean of the total loss series $L(m)$:

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

where $SSR_{L(m)}$ is the restricted sum of squared residuals, $SSR(T_b(m))_{L(m)}$ is the sum of squared residuals assuming one-time change at time $T_b(m)$ and $\hat{V}_{L(m)}$ is the long-run variance estimate of the total loss series constructed using the residuals obtained from the demeaned total loss series with the mean changing at date $T_b(m)$. Also,

$$TLUD(k) = \max_{m\in[m_0,m_1]} UD^k_{L(m)},$$

where

$$UD^k_{L(m)} = \max_{i=1,...,k} \max_{\left\{T_b^i(m)\right\}_{i=1}^k\in\Lambda_\epsilon} \frac{SSR_{L(m)} - SSR(\{T_b^i(m)\}_{i=1}^k)_{L(m)}}{\hat{V}_{L(m)}}$$

with $\Lambda_\epsilon = \{(T_b^1(m),...,T_b^k(m)) \ : \ \left|T_b^i(m) - T_b^{i-1}(m)\right| \geq \epsilon(T-\tau),\ T_b^0(m) = 1, T_b^k(m) \leq (1-\epsilon)(T-\tau)\}$. Also, $\hat{V}_{L(m)}$ is the long-run variance estimate of the total loss series constructed using the residuals obtained from the demeaned total loss series with the mean changing at dates $T_b^1(m),...,T_b^k(m)$. As in Bai and Perron (1998), we set $k = 5$. A step-by-step construction of the TLSW and TLUD tests is as follows: 1) Start with a value of the in-sample length $m = m_0$. A typical choice is $m_0 = \lfloor 0.15T \rfloor$. 2) Compute the total loss series $\{L_{t+\tau}\}_{t=1}^{T-\tau}$. 3) Consider a regression with only a constant: $L_{t+\tau} = \gamma + e_{t+\tau}$. Apply the sup-Wald or UDmax test for the constancy of γ usually with a HAC variance estimate if there is evidence of serial correlation in the loss sequence. Store the value as $SW_{L(m)}$ or $UD^k_{L(m)}$. 4) Use $m = m + 1$ and repeat Steps 2 and 3 to compute $SW_{L(m)}$ or $UD^k_{L(m)}$. Take a choice of $m_1 > m_0$, say $m_1 = \lfloor 0.85T \rfloor$, and continue up to $m = m_1$. Note that the choice of m_0 and m_1 does not affect the asymptotic critical values (see Theorem 2 below). 5) Take the maximum of the sequence of $\left\{SW_{L(m)}\right\}_{m=m_0}^{m_1}$ or $\{UD^k_{L(m)}\}_{m=m_0}^{m_1}$.

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 convergence in distribution. We first require the following assumption.

Assumption 1 *Under the null hypothesis of no change in forecast accuracy: i)* $\hat{\beta}_m \xrightarrow{p} \beta^*$ *for all* $m \in [m_0, m_1]$*; ii) the total loss series evaluated at* β^**,* $L_{t+\tau} = \{L_{t+\tau}(\beta^*)\}_{t=1}^{T-\tau}$*, satisfy* $E[L_{t+\tau}] = \mu$ *for all* t *and for* $r \in [0,1]$*, as* $T \to \infty$ *with* τ *fixed:*

$$T^{-1}\sum_{t=1}^{\lfloor r(T-\tau)\rfloor}\sum_{s=1}^{\lfloor r(T-\tau)\rfloor} E\left[(L_{t+\tau}-\mu)(L_{s+\tau}-\mu)\right] \xrightarrow{p} r\Omega$$

for some fixed matrix Ω*, and* $T^{-1/2}\sum_{t=1}^{\lfloor r(T-\tau)\rfloor}(L_{t+\tau}-\mu) \Rightarrow \Omega^{1/2}W(r)$*, 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. To examine their implications, it is informative to see what they imply for a linear forecasting model. Suppose one has a standard linear model $y_t = x_t'\beta + e_t$. 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 the standard functional limit theorem with long-run variance Ω. Another important feature of Assumption 1 is that the loss series do not depend on m once they are evaluated using the limit value β^* of the parameter estimates. 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 $y_t - x_t'\hat{\beta}_1$ and $y_t - x_t'\hat{\beta}_2$ 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. Therefore, Assumption 1 is fairly general when the model is stable. Note that under Assumption 1, $SW_{L(m)}$ has the same null limiting distribution as the standard sup-Wald test for a change in mean (Andrews, 1993) and the $UD_{L(m)}$ has the same null limiting distribution as the $UD\max$ test of Bai and Perron (1998). We next present the asymptotic distribution of the DSW test.

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

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

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* $\bar{\mu} = \lim_{T\to\infty}(m_1 - m_0)/n_0$.

The critical values of the DSW test were tabulated using 5,000 replications using 5,000 steps to approximate the Wiener process in terms of 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 $\epsilon = 0.1$. The results are presented in Table 1. We next consider the limit distribution of the $TLSW$ and $TLUD$ tests. Exploiting the fact that the loss series $\{L_{t+\tau}(\beta^*)\}_{t=1}^{T-\tau}$ do not asymptotically depend on the value of m under the null hypothesis, we obtain the following theorem.

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

Under Assumption 1, Theorem 2 follows trivially since the sequence of losses $\{L_{t+\tau}(\beta^*)\}_{t=1}^{T-\tau}$ do not depend on the in-sample length m under the null hypothesis. As such, the test statistics 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 Finite sample properties

This section presents simulation results to address the following issues: 1) the finite sample size of the tests proposed in this paper (Section 4.1); 2) the power function of the tests for a wide range of data-generating processes: Section 4.2 describes the experimental design to perform the finite sample analysis, Section 4.3 provides a summary of the main results, while Section 4.4 expands on the sources of various non-monotonic power functions.

4.1 Finite sample size of the proposed tests

We first examine the size of the DSW tests. The data-generating process specifies y_t to be a sequence of i.i.d. $N(0,1)$ random variables of lengths $T = 150$ and $T = 300$. We consider the squared error loss associated with 1) the static model: $y_t = c + e_t$, and 2) the dynamic model:

$y_t = c + \alpha y_{t-1} + e_t$, both estimated by OLS. We consider two versions of the tests applied to each model, namely with or without a HAC correction for potential serial correlation in the loss function. Throughout, the HAC variance estimate is constructed using Andrew's (1991) data dependent method with an AR1 approximation using the Bartlett kernel. For all four cases, we consider a fixed, rolling or recursive forecasting scheme. The exact sizes of the test for nominal 10%, 5%, and 1% levels are presented in Table 2 for $\bar{\mu} = 0.25$, 0.5, and 0.75. The number of replications is 1,000. We label the test without the HAC variance estimate by "non-robust" and the tests with it by "robust".

For $\bar{\mu} = 0.25$ and 0.5, the test shows very little size distortions, if any, for all cases. Some liberal size distortions are present when $\bar{\mu}$ is as large as 0.75. From unreported simulations, these small distortions are caused by parameter uncertainty. With the larger sample size, the distortions decrease to some extent but remain substantial for $\bar{\mu} = 0.75$. Hence, it is recommended to use $\bar{\mu} = 0.25$ or $\bar{\mu} = 0.5$.

We next examine the size of the $TLSW$ and $TLUD$ tests using the asymptotic distribution stated in Theorem 2. The data-generating process and the various specifications are as above and we again set $T = 150$, 300. We used the following values: $\epsilon = 0.1$, $m_0 = \lfloor 0.15T \rfloor$ and $m_1 = \lfloor 0.85T \rfloor$ but any reasonable variations of these choices do not change the results qualitatively. The results are presented in Table 3. The exact size is, in general, close to the nominal size, though some distortions are present when using the robust version. These decrease as T increases.

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 data-generating processes (DGP) involving single or multiple changes in level or in variance. They are specified as follows.

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

DGP2: (recurring shift in variance): $y_t = \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0, \sigma_t^2)$, $\sigma_t^2 = 1$ if $t \in \Lambda_0$, $1+\beta_A$,

otherwise.

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

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

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

We set $T = 150$ and for the cases with multiple breaks as in DGPs 2 and 4, the breaks happen at every 50 periods, i.e. $\Lambda_0 \in \{(1,50),(101,150)\}$. The following tests are considered: the GR_m test with $m = 40$ ($GR1$) and $m = 100$ ($GR2$); the supremum of the GR tests as defined by (2) over $0.2T \leq m \leq 0.8T$ ($SGR1$) and $0.3T \leq m \leq 0.7T$ ($SGR2$), the DSW test with $\bar{\mu} = 0.25$ ($DSW1$) and $\bar{\mu} = 0.50$ ($DSW2$) with $m_0 = 0.2T$ for both cases, the $TLSW$ and $TLUD$ tests with $0.15T \leq m \leq 0.85T$. We again consider two forecasting models: 1) the static model: $y_t = c + e_t$, and 2) the dynamic model: $y_t = c + \alpha y_{t-1} + e_t$, both estimated by OLS. We consider two versions of the tests applied to each model, namely with or without a HAC correction for potential serial correlation in the loss function. Note that in the dynamic model an irrelevant lagged dependent variable is included, in the sense that the true value of α is 0. This is completely inconsequential. We could extend the DGPs to include genuine dynamics with a lagged dependent having a non-zero coefficient. The qualitative features would remain the same. Indeed the non-monotonicity to be reported would simply be more severe. See the discussion in Section 4.4. The static model with no HAC correction is labelled "static, non-robust" and with a HAC correction "static, robust". Similarly, the dynamic model with no HAC correction is labelled "dynamic, non-robust" and the one with a HAC correction "dynamic, robust". We again 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 we adopt to compare the various tests is whether or not 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 a test with good finite sample properties. For tests with monotonically increasing power, we then compare the relative power functions. We start with a summary of the main findings in Section 4.3 and then expand on the explanations for the non-monotonic power functions present for various tests and forecasting schemes.

4.3 Summary of the main power results

The main findings of interest can be illustrated by the results for DGPs 4 and 5 for the case of a dynamic forecasting model with a correction for potential serial correlation in the loss sequence. It transpires that only three tests have a monotonically increasing power function: the two versions of the DSW and $TLUD$ tests, both when constructed using a fixed forecasting window. All the other tests have a power function that initially increases and then 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. The power of this test appears high simply because of huge size-distortions. This is a finite sample feature possibly due to the fact that the weighting of the in-sample and the out-of-sample forecasting errors does not account for the parameter estimation errors. However, the distortions are reduced as m increases given the documented power functions of $GR2$ and $SGR2$. As a result, the power of $SGR1$ looks monotonic, however, the fact that the power of $SGR2$, which has less size-distortions, does decrease to zero with DGP5 shows that $SGR1$ can exhibit the same phenomenon if the size is controlled. 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 (except DGP1) 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, it is instructive to look at the power for a change of magnitude $\beta_A = 0.5$, otherwise the power is most often one for larger values when dealing with tests having a monotonically increasing power function. We also avoid looking at rolling scheme for which the SGR tests have huge size-distortions. Doing so, it transpires that for a single break case (DGPs 1 and 3) the $TLUD$ and SGR tests have equally the highest power. However, if multiple breaks are involved (DGPs 2, 4 and 5), then the $TLUD$ and DSW tests outweigh the SGR tests in all cases. In summary, the test with highest power, across all tests, is the Total Loss UDmax (the only exception is for DGP1 in which case the $TLUD$ and DSW tests have nearly the same power). Hence, the $TLUD$ test not only has a monotonically increasing power, it also has the highest power for small values of the alternative. The DSW test, while having a monotonic power, can have lower power than the $TLUD$ test in most cases. Accordingly, our recommendation is to use the $TLUD$ test followed by the DSW test, both using 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 under the alternative, 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 the $TLUD$ test.

4.4 Some explanations for the power properties

DGPs 1 and 2 are cases with single and multiple variance changes, respectively (Tables 4-1 and 4-2). An important feature of this type of instability is that the forecast model is unaffected by the choice of the T_m, the forecasting scheme or the nature of the changes. More precisely, 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. More precisely under DGP1, DSW has a slightly lower power than $TLSW$ and $TLUD$. Under DGP2, DSW and $TLUD$ have a higher power than $TLSW$, because the latter only accounts for a single break. Also, under DGP2, the power of these tests does not reach one because of the specific nature of the breaks; i.e., two breaks with the first and last regimes being the same. This is the most difficult case to detect with a single break test (e.g., Bai and Perron, 2006). The problem is alleviated with a UDmax type test and our results indeed show that $TLUD$ and DSW are the most powerful in this setting as they are designed to account for multiple breaks.

We now turn to models with mean breaks so that the conditional mean of the variable to be forecasted changes, i.e. DGPs 3, 4 and 5 whose results are reported in Tables 4-3 to 4-5. They reveal that the power functions exhibit non-monotonicity because of three potential sources of power decline. The first is labeled as the "robust effect" indicated with an "R" in the last row of each panel for cases with a non-monotonic power function. The second is labelled as the "window effect" indicated with a "W". The third cause is labelled as the "dynamic effect" and indicated with a "D".

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 estimate 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 choice of window that separates the in and out-of-sample data and causes a loss in power. This applies, for example, when some breaks occur in the in-sample partition so that the model is not consistently estimated. The "dynamic effect" is the most pronounced. It is basically 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, 1991). This results in forecast errors being roughly the first-differences of the forecast errors from a model without lagged dependent variables. Hence, the mean breaks are transformed into spikes in the loss sequence so that 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.

In order to illustrate the effect of the different sources of power decline, we present in Figures 1-2, 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). This is done for the static model (Figure 1) and the dynamic model (Figure 2). 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.

Let us now explain some of the power functions reported in Tables 4-3 to 4-5. First, as shown in panel (a) of Tables 4-3 to 4-5, when the static model and the fixed scheme are used, all the non-robust tests have a monotonic power. However, the "robust effect" applies when considering tests constructed with HAC variance estimate in panel (b). The "robust effect" is pronounced for $GR1$ and $GR2$ and it also applies to a lesser extent to $SGR1$, $SGR2$ and $TLSW$ when multiple breaks are present. In particular, when the fixed scheme is used, the "robust effect" applies to $GR1$ under DGPs 3 and 5, and to $GR1$, $GR2$, $SGR1$, $SGR2$ and $TLSW$ under DGP4. When using either the rolling or recursive scheme, the "robust effect" 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. These results strongly suggest not using GR and SGR with HAC variance estimate under any forecasting scheme. It also shows that $TLSW$ is unreliable when multiple breaks are suspected.

It is interesting to see that even in panel (a) the power of $GR1$ and $GR2$ sometimes remains very low. $GR1$ shows a 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. When the fixed scheme is used, the "window effect" applies to $GR2$ under DGP4. When the rolling scheme is used, it applies to $GR2$ under DGPs 3 and 4 and when the recursive scheme is used to $GR1$ under DGP5. It is also worth noting that DSW may lose power when the rolling scheme or the recursive scheme is used. This is because the loss sequence takes a triangular shape as shown in Figure 1. In particular, this applies to $DSW1$ and $DSW2$ under DGP4. For DSW, the "window effect" can be exacerbated by the "robust effect", indicted by the label "R, W" in panel (b); cf. $DSW1$ with the rolling scheme under DGP3 and $DSW1$ and $DSW2$ with the rolling and the recursive schemes under DGP4. The source of the "window effect" can be explained by looking at the results in Figure 1. When the fixed scheme is used, the loss sequence takes a step-wise pattern, abrupt change(s) followed by a flat region. This is qualitatively the same for all three DGPs and test statistics. Importantly, the shape of the loss sequence changes when the rolling or recursive scheme is used. The typical loss sequence then shows an abrupt increase followed by a gradual decline. In the simplest case of DGP3, the increase occurs when the in-sample window covers a stable period and the initial date of the out-of-sample period ($m + \tau$) coincides with the true break date. After this point, the window increasingly contains the post-break data, which gradually causes a bias in the forecast model estimation and thus a decline of the loss sequence[1].

The non-monotonic power functions are more pronounced when a dynamic model is used, i.e., the "dynamic effect" is especially important. Panel (c) of Tables 4-3 to 4-5 report power functions using a dynamic model with non-robust standard errors. When the rolling or the recursive schemes are used, this "dynamic effect" applies to all tests under DGPs 3-5. In addition, the "dynamic effect" applies even when the fixed scheme is used if the in-sample window is relatively large and includes the break, for instance with $GR2$ under DGPs 3 and 4 and with $GR1$, $GR2$ and $SGR2$ under DGP5[2]. This evidence suggests not to use any tests with the 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 the tests with a non-monotonic power function. To gain some

[1]More importantly, the shape is robust regardless of whether the static or the dynamic model is used, except for the case in which the first break is always included in the in-sample window ($SGR2$ under DGP5 with the dynamic model since $T/4 < 0.3T$).

[2]More precisely, the minimum in-sample lengths are 40 for $GR1$, 100 for $GR2$, 30 for $SGR1$ and 45 for $SGR2$. The first break location is 75 for DGP3, 50 for DGP4 and 38 for DGP5. Hence, with the fixed scheme, $GR2$ suffers from the "dynamic effect" under DGPs 3-4 and $GR1$, $GR2$ and $SGR2$ under DGP5.

insight about the cause of the power losses when using the rolling or recursive scheme, Figure 2 shows that the mean breaks are transformed into spikes in the loss sequence [3].

5 Empirical applications

This section provides empirical examples to illustrate how the proposed tests and the existing methods are able to detect changes in forecast accuracy. One is related to oil price forecasts and the other to forecasting inflation. The results will clearly illustrate the empirical relevance of the simulation results.

5.1 Oil price forecasts

We consider first forecasting oil prices using a simple linear model. The series used is WTI monthly crude oil prices (U.S. dollars per barrel in logarithm form) from 1986:1 to 2011:11 plotted in Figure 3. We consider the following simple forecasting model

$$p_{t+\tau} = c + e_{t+\tau},$$

(labelled 'static model'), and the model with a lagged dependent variable (labelled 'dynamic model')

$$p_{t+\tau} = c + \alpha p_t + e_{t+\tau}.$$

These are not intended to be good forecasting models. Indeed, a quick look at the graph of the series reveals large instabilities in level after 2000 and an upward trend in the mid-sample. Hence, it is expected that these simple models produce forecasts that easily break down or change over time. Hence, we would expect any good test to indicate a clear rejection of the null hypothesis of no change in forecast accuracy.

We applied the following testing procedures: the DSW tests ($DSW1$ and $DSW2$ for $\bar{\mu} = 0.25$ and 0.5), the $TLSW$ and $TLUD$ tests. For these tests, we used the truncation $\varepsilon = 0.1$ and the maximum number of breaks is set at five for the UDmax tests. We also applied the GR_m tests ($GR1$ and $GR2$ with in-sample length $m = 55$ and 259) and the SGR tests ($SGR1$ and $SGR2$ for $0.2T \leq m \leq 0.8T$ and $0.3T \leq m \leq 0.7T$). The specific choices of the in-sample window size $m = 55$ (1990:7) is chosen based on the fact that it correspond to the period before the oil price hike that occurred due to the Gulf War. The second choice $m = 259$ (2007:7) corresponds to an in-sample window before the recent financial crisis. We consider forecasts with $\tau = 1$ and the fixed, rolling and recursive window schemes are used.

[3] A few exceptions are $GR2$ under DGPs 3 and 4 and $GR1$, $GR2$ and $SGR2$ under DGP5.

To proceed, we test for serial correlation in the squared loss using the LM test to ascertain whether a HAC correction is needed in implementing the tests. Table 5-1 indicates that the null of no serial correlation is strongly rejected. Given this result, Table 5-2 presents the tests constructed by accounting for serial correlations in the loss series using a HAC variance estimate. First, if we use the static model, all of the newly proposed tests ($DSW1$, $DSW2$, $TLSW$, and $TLUD$) reject the null hypothesis at the 1% level, suggesting strong evidence of forecast breakdowns. On the other hand, the evidence in favor of a rejection is very weak using any of the existing tests. The GR_m tests show almost no rejections and the SGR tests reject only when the rolling scheme is used (likely due to the very large size distortions reported earlier). Suppose now that the dynamic model is used. The new tests using a fixed window all reject even at the 1% level, as in the case of the static model. As expected from the simulations, these tests do not reject when using a rolling or recursive window since the changes in levels then translate into outliers. This is consistent with our recommendation of using a fixed window. On the other hand, the existing methods again provide very weak evidence of forecast breakdowns. The GR_m tests yield almost no rejection and the SGR tests barely reject with a fixed window.

5.2 US inflation forecasts using the Phillips curve

The second example follows the application in Giacomini and Rossi (2009). We consider inflation forecasts using the Phillips Curve:

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

where π_t is an inflation measure and u_t is the unemployment gap (the difference between the unemployment rate and a measure of the NAIRU). The difference between this forecasting model and that used in Giacomini and Rossi (2009) is that they use the "first difference" of the inflation rate and we use the inflation rate in level. We use the same monthly CPI (consumer price index; revised version) and the unemployment gap data[4] which spans the period 1959:1 to 2004:6. The same transformation ($\pi_t = (1200/\tau)\ln(P_t/P_{t-\tau})$), where P_t is the CPI at month t, was applied to construct the inflation rate. The order of the lag polynomial $\theta_1(L)$ is either $q_u = 1$ or $q_u = 3$. Since the results are similar, we only report the case with $q_u = 3$. We also consider the order of $\theta_2(L)$ to be $q_\pi = 0, 1$, and 3. The case with $q_\pi = 0$ does not use lagged inflation as a predictor and, hence, we label this case as the 'static' one, while the cases with $q_\pi = 1$ and $q_\pi = 3$ are labelled 'dynamic 1' and

[4]They were downloaded from Barbara Rossi's web site.

'dynamic 3', respectively. We consider two window sizes. The first choice is $m = 241$, as in Giacomini and Rossi (2009). This means that the forecaster chooses the period before 1979 as the in-sample so that the Volker's Chairmanship period of high inflation stays in the out-of-sample (see Figure 4). The second choice is $m = 301$ (1984:01), in which case the so-called Great Moderation which is deemed to have occurred in the middle of the 80's is in the out-of-sample period. In either case, the presumption is that there would be general consensus of forecast breakdowns.

We first implement the LM tests for serial correlation in the squared loss series for each model and values of m. The results in Table 6-1 indicate that in all cases the null of no serial correlation in the loss series is strongly rejected. Hence, we use the HAC correction to construct the tests, presented in Table 6-2. The tests considered are the same as in the previous application. For $GR1$ and $GR2$, we use $m = 241$ and $m = 301$, respectively. If we use the static model, the DSW tests and the $TLUD$ test show strong rejections of no change in forecast accuracy. The $TLSW$ does not reject since the loss series likely has multiple breaks and except for the largest one the breaks are not accounted for in this case. More interestingly, all the GR-based tests, except for the SGR tests with a rolling window (again likely because of its large size distortions), do not detect a change in forecast accuracy. Consider next the versions of the forecasting model that includes lagged inflation (dynamic 1 and 3). The DSW tests still strongly reject the null of no change in forecast accuracy and the $TLSW$ and $TLUD$ tests now provide even stronger rejections. More strikingly, all of the GR-based tests completely lose power once the model includes lagged dependent variables.

In conclusion, if we use the existing methods we fail to find any evidence of a change in forecast accuracy or a forecast breakdown which is deemed to have taken place in the high inflation period of the Volker's Chairmanship period. These methods are also not able to detect a change due to the Great Moderation which is deemed to have occurred in the mid-80's. In contrast, the newly proposed DSW and $TLUD$ tests clearly show evidence of a change in forecast accuracy in the form of a forecast breakdown.

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 original GR test (a t-test for a change at some pre-specified date m; maximizing the GR_m test over all possible 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. Using extensive simulations, 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.

Acknowledgment. Yamamoto acknowledges financial support for this work from the MEXT Fund for the Promotion of Joint International Research.

Appendix: Proof of Theorem 1

In this appendix, we omit the superscript of o for the out-of-sample loss sequence, i.e., we denote L_t^o by L_t for simplicity. The Wald test for a constant mean versus one break at time $t = T_b = m_0 + \lfloor \lambda n_0 \rfloor$ for the series $\{L_{t+\tau}\}_{t=m+1}^{T-\tau}$ is

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

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

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

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

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

Hence,

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

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

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

Also, using Assumption 1

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

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

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

Combining the above results yields

$$
SSR_{L(m)} - SSR(T_b)_{L(m)} \Rightarrow
$$

$$
\Omega[-\frac{1}{1-\mu} \left[W(1) - W(\mu) \right]^2 + \frac{1}{\lambda - \mu} \left[W(\lambda) - W(\mu) \right]^2 + \frac{1}{1-\lambda} \left[W(1) - W(\lambda) \right]^2]
$$

and under the null hypothesis we have $\hat{V}_{L(m)} \xrightarrow{p} \Omega$. 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,

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

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

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.

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

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

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

Chang, S.Y. and P. Perron (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.

Crainiceanu, C. M. and T.J. Vogelsang (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 P. Perron (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.

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

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

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

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

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. (1991) "A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series," mimeo, Princeton University.

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

Perron, P. and Y. Yamamoto (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.

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.

Table 1: Critical values of the DSW test

$\bar{\mu}$	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

Note: The trimming parameter $\epsilon = 0.1$ is used.

Table 2: Size of the DSW test ($T = 150$)

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

$(T = 300)$

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 3: Size of the $TLSW$ and $TLUD$ tests

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

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 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	-	-	-	-	-	-	-	-
	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	0.323	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	-	-	-	-	-	-
	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	0.117	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	-	-	-	-	-	-	-	-

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	-	-	-	-	-	-	-
	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	0.330	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	-	-	-
	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	0.122	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	-	-	-	-	-	-

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	-	-	-	-	-	-
	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	0.601	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
	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	0.179	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

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	-	-	-	-	-	-
	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	0.615	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
	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	0.186	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	-	-	-	-	-	-
	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	0.821	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	-	-
	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	0.256	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	-	-	-	-	-	-
	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	0.915	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
	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	0.248	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	-
	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	0.822	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	-	-
	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	0.218	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	-
	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	0.893	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
	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	0.257	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	-	-	-	-	-	-	-

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

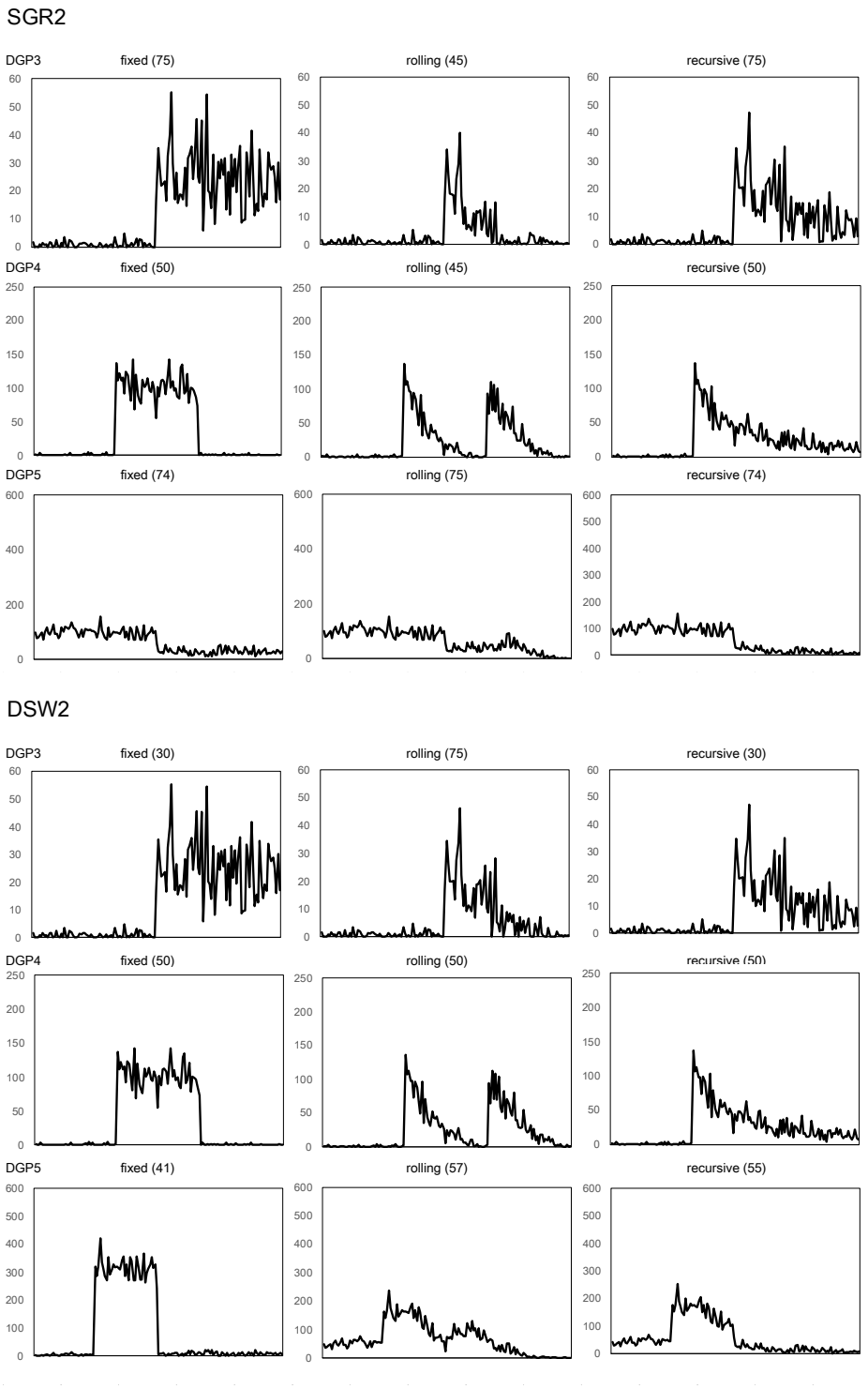
b) static, robust

	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	-	-	-	-	-	-	-

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: A realization of loss sequences: static model



TLSW

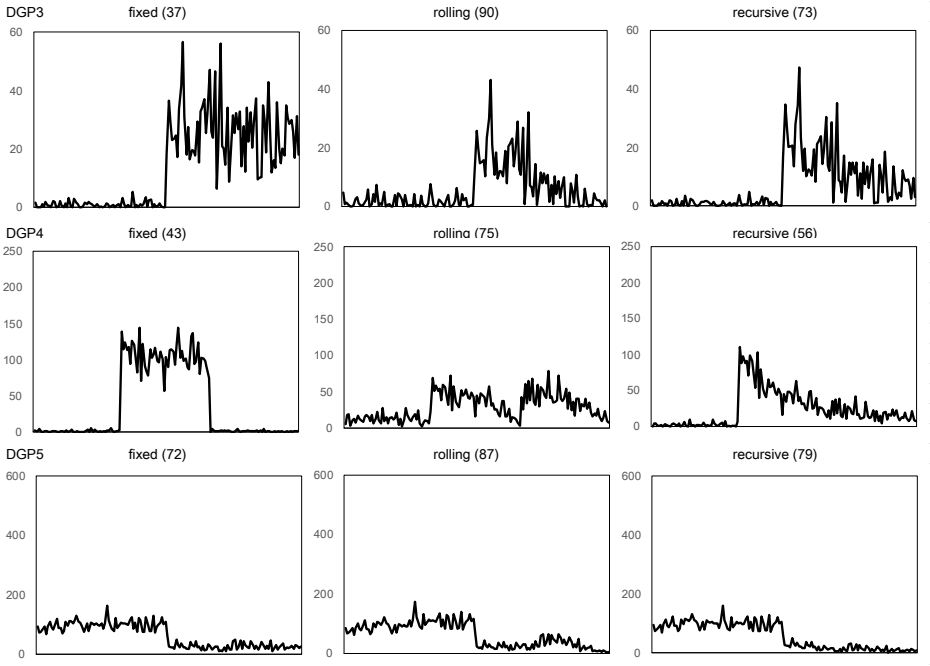
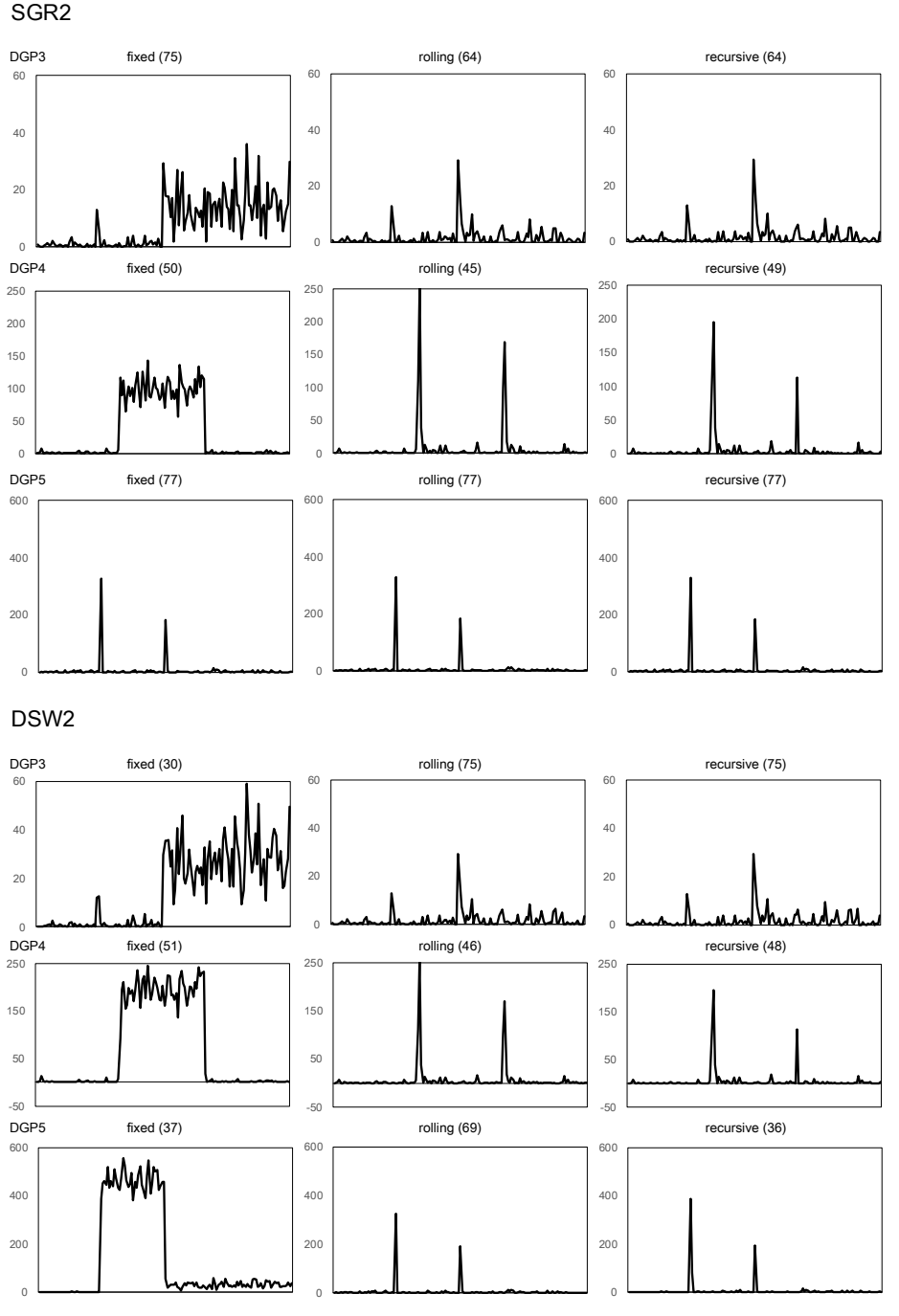

DGP3
fixed (37)
rolling (90)
recursive (73)
DGP4
fixed (43)
rolling (75)
recursive (56)
DGP5
fixed (72)
rolling (87)
recursive (79)

Figure 2: A realization of loss sequences: dynamic model



TLSW

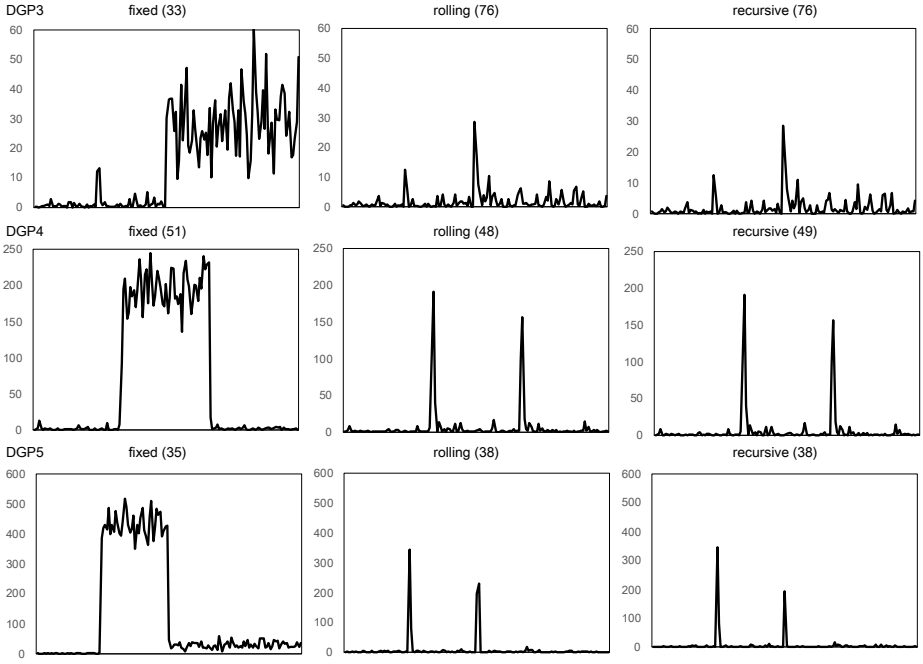

DGP3
fixed (33)
rolling (76)
recursive (76)
DGP4
fixed (51)
rolling (48)
recursive (49)
DGP5
fixed (35)
rolling (38)
recursive (38)

Figure 3: Crude oil price (WTI: US dollar / barrel, in logarithm)

Table 5-1: Serial correlation tests for the squared loss series (oil price forecasts)

	m=55		m=259	
	static	dynamic	static	dynamic
fixed	302.08***	242.52***	296.88***	44.40***
rolling	231.99***	49.66***	293.47***	44.28***
recursive	295.52***	54.97***	294.53***	44.17***

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 5-2: Tests for change in forecasting performance (oil price forecasts)

static model

	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2
fixed	315.28***	315.28***	609.89***	609.89***	1.19	1.59	2.71*	2.71*
rolling	16.09***	26.00***	417.22***	417.22***	6.03***	1.70*	5.37***	4.35***
recursive	350.80***	350.80***	728.24***	728.24***	1.84*	1.63	2.48*	2.48*

dynamic model

	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2
fixed	140.08***	140.08***	176.14***	176.14***	1.45	0.84	5.23***	5.23***
rolling	9.17	9.17	6.47	6.47	4.01***	1.01	1.77	1.77
recursive	14.00**	14.00*	6.57	6.57	0.17	0.79	0.93	0.93

Note: Same as Table 5-1.

Figure 4 : US inflation rate

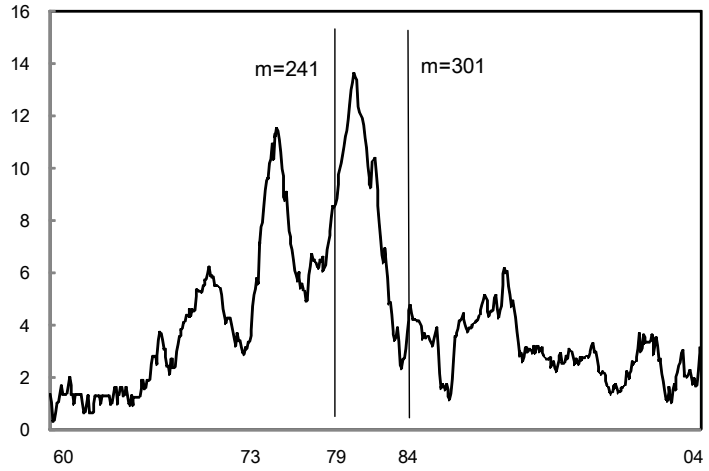


Table 6-1 : Serial correlation tests for the squared loss series (US inflation forecasts)

	m=241			m=301		
	static	dynamic 1	dynamic 3	static	dynamic 1	dynamic 3
fixed	527.60***	522.86***	522.86***	539.98***	541.71***	541.71***
rolling	521.72***	522.86***	522.86***	539.69***	541.71***	541.71***
recursive	525.47***	522.86***	522.86***	540.52***	541.71***	541.71***

Note: Same as Table 5-1.

Table 6-2 : Tests for change in forecasting performance (US inflation forecasts)

static

	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2
fixed	74.91***	337.00***	4.42	451.51***	0.12	0.30	1.53	1.53
rolling	36.36***	46.13***	8.72*	466.16***	0.52	0.24	7.40***	6.20***
recursive	55.14***	150.54***	7.66*	865.20***	0.04	0.40	1.11	0.75

dynamic 1

	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2
fixed	21.45***	21.45***	23.86***	2981.04***	0.08	0.02	0.94	0.20
rolling	15.71**	15.71**	263.20***	2981.04***	0.13	0.03	0.38	0.38
recursive	16.41**	16.41**	265.14***	2981.04***	0.12	0.02	0.26	0.26

dynamic 3

	DSW1	DSW2	TLSW	TLUD	GR1	GR2	SGR1	SGR2
fixed	22.32***	22.32***	23.48***	2993.64***	0.08	0.02	1.10	0.20
rolling	15.74**	15.74**	266.60***	2996.90***	0.13	0.03	0.38	0.38
recursive	16.24***	16.24**	267.38***	2992.51***	0.12	0.02	0.26	0.26

Note: Same as Table 5-1.