Forecasting trending autoregressive time series under changing persistence[*]

Robinson Kruse

CREATES, Aarhus University

Preliminary version – please do not cite!

April 2009

Abstract

Changing persistence in time series models means that a structural change from nonstationarity to stationarity or vice versa occurs over time. Such a change may have important implications for forecasting. This paper studies a variety of forecasting strategies which allow for changes in persistence by model and data selection based on pre-testing. In a large-scale Monte Carlo study we investigate the performance of several forecasting devices under many different data generating processes. These include other types of structural breaks like changes in the unconditional volatility, level shifts and broken trends. We propose a forecasting strategy which is able to outperform traditional approaches for trending macroeconomic and financial data significantly for a wide selection of breaks.

Key Words: Forecasting, changing persistence, trend, structural break, pre-testing, breakpoint estimation.

---

[*] I would like to thank Isabel Casas, David Hendry and the participants of the CREATES members meeting 2008 in Sophiendal Manor, Denmark for insightful discussions. Moreover, I am indebted to Philipp Sibbertsen for inspiring suggestions.

1 Introduction

Recent research in time series econometrics has paid a lot of attention to structural breaks in autoregressive time series models. Perron (2006) provides an excellent survey of several relevant issues in this field. One of these topics is changing persistence, which is defined as a change in the degree of integration of a time series. A leading example for changing persistence is a first-order autoregressive (AR) model where the AR parameter equals one during the pre-break period and less than unity in absolute value during the post-break period, or vice versa. Such a break in persistence may have immediate consequences for forecasting. It is well known that ignoring structural breaks in the parameters of the forecasting model may result in poor forecasting performance. This paper investigates the behaviour of standard forecasting models under changing persistence and proposes a new forecasting strategy based on pretesting. It is shown that this strategy shows a good performance even if other types of structural breaks occur. Among these we consider empirically relevant level shifts, broken trends and changing volatility specifications.

As Perron (2006) points out, changing persistence has been an issue of substantial empirical interest, especially concerning inflation rate series (e.g., Barsky, 1987, Burdekin and Siklos, 1999), short-term interest rates (e.g., Mankiw et., 1987), government budget deficits (e.g., Hakkio and Rush, 1991) and real output (e.g., Delong and Summers, 1988). An emerging strand of literature discusses changing persistence with a special emphasis on US and European inflation rates, see O'Reilly and Whelan (2005), Kang et al. (2006), Kumar and Okimoto (2007) and Pivetta and Reis (2007). Moreover, Halunga et al. (2008) consider the possibility of structural changes in the mean of the inflation data generating process next to changing persistence.

The forecasting strategy we propose in this paper is based a unit root test against changing persistence. This test has been recently suggested by Leybourne et al.

(2007) and its properties are useful in the context of forecasting. A key property of this test is its conservative behaviour under constantly stationary data generating processes. This results in a test which is able to discriminate consistently between processes with constant and changing persistence. It should be noted that this property is usually not shared by other tests against changing persistence, see Leybourne et al. (2007). Among these are the tests proposed by Kim (2000), Kim et al. (2002), Leybourne et al. (2003) and Busetti and Taylor (2004). As a by-product, Leybourne et al. (2007) suggest consistent breakpoint estimators which allows us to discard unnecessary data points which may negatively influence the forecasting performance. As analyzed in Pesaran and Timmermann (2005), it is beneficial to use only post-break data for forecast model estimation. On the contrary, it is not sensible to always assume that a change in persistence occurs. In fact, if no break occurs, useful observations for forecast model estimation are discarded. Hence, it might be beneficial to discriminate between constant and changing persistence as a first step and to estimate the breakpoint as a second step in case of a rejection.

In order to consider empirically relevant models for macroecnomic and financial data, we allow for linear deterministic trends in stationary autoregressive models. This, however, implies an estimation bias when ordinary least squares (OLS) is applied. Fortunately, several bias-correction methods have been proposed in the related literature. As shown by Andrews and Chen (1994), the estimates of the parameters of a linear trend model with serial correlation can be heavily biased, especially in small samples. Therefore, we follow Kim (2003) and apply some bias-correction methods for the estimator of an autoregressive time series model with a linear trend. Among these are the approximately median-unbiased estimator by Roy and Fuller (2001) and a bootstrap-based technique by Kim (2003). We expect that bias-corrected estimation has a positive impact on the forecasting performance but it can be expected that the question whether to conduct a pre-test or not is of greater practical relevance.

Our last contribution is to study the behaviour of several forecasting strategies under unit root processes that are subject to other types of breaks. We consider (*i*) level shifts, (*ii*) broken trends and (*iii*) changes in the unconditional variance of the error process. Moreover, we also consider the possibilities of level or trend shifts and a simultaneous change in variance. In a large-scale Monte Carlo study we investigate the empirical performance of several forecasting strategies. We consider small and moderate sample sizes and forecast horizons ranging from one to fifty steps ahead. Our results indicate clearly the usefulness of the newly proposed forecasting strategy under a variety of empirically relevant conditions.

The paper is organized as follows: section 2 discusses the autoregressive time series model under changing persistence, related unit root tests and breakpoint estimation. Section 3 covers a detailed description of all forecasting strategies, while bias-corrected estimation of the trending AR model is briefly reviewed in section 4. The Monte Carlo setup and discussions of numerical results are given in section 5. Main conclusions are drawn in section 6.

2 Changing Persistence and related statistics

2.1 AR model with changing persistence

We consider the following first-order autoregressive model that is subject to a change in persistence at some breakpoint $[\tau T]$ with $\tau \in (0,1)$:

$$y_t = \beta + \gamma t + \alpha_1 y_{t-1} + \sigma\varepsilon_t, \text{for } t = 1, 2, ..., [\tau T] \quad (1)$$

$$y_t = \beta + \gamma t + \alpha_2 y_{t-1} + \sigma\varepsilon_t, \text{for } t = [\tau T] + 1, ..., T \ . \quad (2)$$

The innovation process ε_t is assumed to be stationary, short memory and linear innovation process: $\varepsilon_t = C(L)u_t = \sum_{i=1}^{\infty} C_j u_{t-j}, 0 < C(1) < \infty, \sum_{j=0}^{\infty} j^a |C_j| < \infty$ for some $a > 1$. Furthermore, u_t is i.i.d. with mean zero, constant variance $\sigma^2 > 0$ and finite fourth-order moments.

In this model, persistence is determined through autoregressive parameters $0 \leq \alpha_1 \leq 1$ and $0 \leq \alpha_2 \leq 1$. As long as $\alpha_1 \neq \alpha_2$, a structural change occurs at time $[\tau T]$. The special case of this particular break in which $|\alpha_1| = 1$ and $|\alpha_2| < 1$ hold, is called a decline in persistence because the trending AR model in equation (1) is non-stationary, that is $I(1)$, during time $t = 1, 2, \ldots, [\tau T]$ and stationary, that is $I(0)$, afterwards. In particular, as long as $\gamma \neq 0$ holds, the AR model (1) is trend-stationary in the post-break period $t = [\tau T] + 1, [\tau T] + 2, \ldots, T$. Analogously, an increase in persistence takes place if $|\alpha_1| < 1$ and $|\alpha_2| = 1$, i.e. the process switches from (trend-)stationary to a unit root process. It should be noted that stable shifts, i.e. $\alpha_1 \neq \alpha_2$ and $\alpha_1, \alpha_2 \in (0, 1)$, do not constitute a change in persistence as the process is stationary over the whole sample period.

Changing persistence may have consequences for the validity of economic theories like the Purchasing Power Parity, for economic decision making as impulse responses change over time and for forecasting. Ignoring structural changes in the autoregressive parameter may lead to relatively poor forecasting performance, see Pesaran and Timmermann (2006). These authors consider a small number of experiments where persistence is changing over time. A main conclusion is that ignoring pre-break data can improve the forecast accuracy significantly. The breakpoint is, however, generally unknown and has to be estimated from the data. Furthermore, it does not make much sense to assume that a change in persistence occurs for all considered time series to be forecasted. If no actually no break occurs, then useful information for forecasting is wasted. Hence, it might be beneficial to test for constant persistence against a change in it as a first step and to estimate the breakpoint as a second step in case of a rejection. After breakpoint estimation, an AR model may be fitted to the post-break data. If no rejection occurs, the whole sample should be used for model estimation for reasons of estimation accuracy. This paper proposes a pre-testing strategy which is described in the next subsection.

2.2 Unit root test against a change in persistence

The literature on testing constant against changing persistence considers the null hypothesis of a constant $I(1)$ or constant $I(0)$ process while the alternative is given by a deterministic change from $I(1)$ to $I(0)$ over time, or vice versa. In principle, these tests can be carried out as one-sided tests with a known direction of change or as two-sided tests with unknown direction. In this paper, one-sided tests are carried out as the direction is usually clearly given by historical situation. One of the most recent contributions to this literature is the CUSUM of squares-based test by Leybourne et al. (2007). The authors solve an important problem that is inherent in other tests for changing persistence: The asymptotic size equals one if both, the null and the alternative hypotheses are wrong. This situation means for a unit root test against changing persistence that the process is constantly $I(0)$ and for a stationarity test that it is constantly $I(1)$. The construction of the CUSUM of squares-based test by Leybourne et al. (2007) results in a conservative test, i.e. the asymptotic size is equal to zero. This property can be of use in a forecasting context. The reason is that a spurious rejection of the null hypothesis would imply an unnecessary waste of important data points for estimation of the forecast model. Such spurious rejection do not occur when applying the Leybourne et al. (2007) test. Therefore, we focus in this work on this test. We therefore provide a brief discussion of it and the related breakpoint estimator for the true and unknown parameter τ.

The test against a change in the persistence as proposed by Leybourne et al. (2007) builds upon the test statistic R which is given by

$$R = \frac{\inf_{\tau \in \Lambda} K^f(\tau)}{\inf_{\tau \in \Lambda} K^r(\tau)},$$

where $K^f(\tau)$ and $K^r(\tau)$ are CUSUM of squares-based statistics based on the forward and reversed residuals of the data generating process as given below. The relative breakpoint $\tau \in \Lambda = [\underline{\tau}, \overline{\tau}]$ is assumed to be unknown and an estimator for τ is given below. In more detail, $K^f(\tau)$ and $K^r(\tau)$ are given by

$$K^f(\tau) = \frac{1}{[\tau T]^2 \hat{\gamma}_0^f(\tau)} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2$$

and

$$K^r(\tau) = \frac{1}{(T - [\tau T])^2 \hat{\gamma}_0^r(\tau)} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2.$$

Here, $\hat{v}_{t,\tau}$ are the residuals from the OLS regression of y_t on a constant based on the observations up to $[\tau T]$. This is

$$\hat{v}_{t,\tau} = y_t - \bar{y}(\tau)$$

with $\bar{y}(\tau) = [\tau T]^{-1} \sum_{t=1}^{[\tau T]} y_t$. Similarly $\tilde{v}_{t,\tau}$ is defined for the reversed series $z_t \equiv y_{T-t+1}$. In addition, $\hat{\gamma}_0^f(\tau)$ and $\hat{\gamma}_0^r(\tau)$ are OLS variance estimators for $\Delta\hat{v}_{t,\tau}$ and $\Delta\tilde{v}_{t,\tau}$, respectively. Analogous expressions for the case of de-trending can be found in Leybourne et al. (2007). The null hypothesis of a constant unit root process is rejected for large values of R in favor of the alternative. Regarding the unknown breakpoint, Leybourne et al. (2007) prove consistency of two breakpoint estimators which are given by

$$\hat{\tau}^f = \arg\inf_{\tau \in \Lambda} \frac{1}{[\tau T]^2} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2 \qquad \text{if } y_t \sim I(0) \quad \rightarrow \quad I(1)$$

and

$$\hat{\tau}^r = \arg\inf_{\tau \in \Lambda} \frac{1}{(T - [\tau T])^2} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2 \qquad \text{if } y_t \sim I(1) \quad \rightarrow \quad I(0).$$

Note, that $\frac{1}{[\tau T]^2}\sum_{t=1}^{[\tau T]}\hat{v}_{t,\tau}^2$ and $\frac{1}{(T-[\tau T])^2}\sum_{t=1}^{T-[\tau T]}\tilde{v}_{t,\tau}^2$ are equal to the unstandardized (without the long-run variance estimator) forward and backward statistics $K^f(\tau)$ and $K^r(\tau)$, respectively.

It is worthwhile to note that standard unit root and stationarity tests are ill-behaved if a change in persistence occurs. As a change in persistence implies that there is a fraction of observations where the process is stationary and another fraction of non-stationary observations, the behaviour of standard tests depend dramatically on the breakpoint τ. If the fraction of observations which belong to the stationary part of the sample is small, rejections are not likely and vice versa. Hence, standard unit root tests are not able to discriminate between constant $I(1)$ and changing persistence processes. However, for reasons of comparison we shall consider the standard unit root test proposed by Dickey and Fuller (1979) in this study as well.

2.3 A more general model with breaks

The changing persistence model (1) restricts other parameters than autoregressive ones to be constant. Given the paramount importance of level shifts, broken trends and changes in the unconditional volatility for economic time series and the corresponding great deal of literature covering these issues, we shall consider a much more general AR model. It is given by

$$
\begin{aligned}
y_t &= \beta_1 + \gamma_1 t + \alpha_1 y_{t-1} + \sigma_1 \varepsilon_t, \text{for } t = 1, 2, ..., [\tau T] && (3)\\
y_t &= \beta_2 + \gamma_2 t + \alpha_2 y_{t-1} + \sigma_2 \varepsilon_t, \text{for } t = [\tau T] + 1, ..., T\ , && (4)
\end{aligned}
$$

with either decreasing persistence, $|\alpha_1| = 1, |\alpha_2| < 1$, or increasing persistence, $|\alpha_1| < 1, |\alpha_2| = 1$. Even though this model allows for four different types of breaks it restricts the number of breaks to be equal to one. Another limitation is the assumption that all breaks occurs at the same time, i.e. $\tau_{\beta,\gamma,\alpha,\sigma} = \tau$. Only a few articles discuss the effects

Table 1: Forecasting strategies

	No change in persistence	
S1	$I(0)$ + trend	(black line)
S2	$I(1)$ + drift	(1.00 line)
S3	Pre-testing for unit root (choose between S1 and S2)	(red line)
	Change in persistence	
S4	Post-break data ($\widehat{\tau}$) for $I(0)$ + trend model	(green line)
S5	Pre-testing for changing persistence (choose between S3 and S4)	(blue line)

of other types of structural breaks on test for constant against changing persistence. Among these are Belaire-Franch and Cavaliere and Taylor (2006) who consider mean shifts and volatility breaks, respectively. In principle, tests are not robust and may reject the null hypothesis of constant persistence far too often. Hence, changing persistence may be confused with other types of structural breaks. Our numerical results in section 5 will shed some light on these issues and the consequences for forecasting.

3 Forecasting strategies

We consider a variety of forecasting strategies which may be divided into two groups: those accounting for structural changes in persistence and ones permitting no breaks. Among the strategies assuming no breaks are a trend-stationary AR(1) model (Strategy 1, S1 for short), the random walk forecast (Strategy 2) and a pre-testing procedure (Strategy 3) analyzed by Diebold and Kilian (2000). On the contrary, we consider two different strategies regarding changes in persistence. First, it assumed that the alternative hypothesis of the CUSUM of squares-based test is always true (Strategy 4). Second, we consider a pre-testing procedure accounting for changes in persistence with a ceratin significance level α (Strategy 5). Table 1 summarizes different forecasting strategies. The next subsections are dedicated to a more careful explanation of details regarding Strategies 1 – 5.

3.1 Assuming no change in persistence

Strategy 1 Standard forecasting models for trending macroeconomic and financial data are the trend-stationary AR(1) model and the random walk. The trend-stationary AR(1) model is given by

$$y_t = \beta + \gamma t + \alpha y_{t-1} + \sigma \varepsilon_t \ ,$$

with $|\alpha| < 1$. Under the assumption of the quadratic loss function $L(y_{t+h}, \widehat{y}_{t+h|t}) = (y_{t+h} - \widehat{y}_{t+h|t})^2$, the MSFE-optimal h-step forecast of y_t generated in t from the trend-stationary AR(1) model is given by

$$\widehat{y}^{S1}_{t+h|t} = \sum_{j=0}^{h-1} \alpha^j (\beta + \gamma(t + h - j)) + \alpha^h y_t \ .$$

Ng and Vogelsang (2002) show that direct OLS estimation of the trend-stationary model dominates a two-step procedure where the data is de-trended in a first step and the persistence parameter α is estimated in a second step. Therefore, we apply the direct method in this paper. As a competitor to classic OLS estimation we consider a feasible generalized least squares (FGLS) estimator proposed by Roy and Fuller (2001) and a bootstrap mean-corrected estimator discussed in Kim (2003). These estimators are considered in section 4 in more detail.

Strategy 2 The second considered forecasting model is the unit root model with drift β, i.e.

$$y_t = \beta + y_{t-1} + \varepsilon_t \ .$$

Analogously to the trend-stationary forecast model, the h-step ahead forecast from this difference-stationary model is given by

$$\widehat{y}^{S2}_{t+h|t} = \beta h + y_t \ .$$

The drift parameter β is estimated with the OLS method. It is worthwhile to emphasize that the forecasts from both models (trend-stationary and unit root model) can be quite different from each other and this is especially true for large forecast horizons h. Furthermore, it is well known that an incorrect model may outperform the true model in terms of out-of-sample forecasting. Therefore, it is not necessarily always a good strategy to search for the true model. Pre-testing is a possibility to account for model uncertainty in a simple environment as it is given here (only two competing models are considered and one of them is nested). Instead of always using either (i) the trend-stationary model or the (ii) unit root model with drift one may gain a lot in terms of out-of-sample MSFE reduction by using a pre-test to select one of the competing models prior to forecasting.

Strategy 3 As intensively discussed in Diebold and Kilian (2000), pre-testing offers quite often a significant improvement compared to Strategy 1 or Strategy 2. The pre-test itself should contain one of the model as the null model and the other one as alternative model. Popular candidates are the KPSS test by Kwiatkowski et al. (1992) and the Dickey-Fuller (DF) test. We choose the famous and widely applied DF test as the pre-test which is also used in Diebold and Kilian (2000). In particular, we run the DF regression including a constant and a linear trend,

$$\Delta y_t = \xi_0 + \xi_1 t + \phi y_{t-1} + u_t.$$

If the null hypothesis $H_0 : \phi = 0$ is rejected at some nominal significance level in favor of $H_1 : \phi < 0$, then a trend-stationary forecast model is selected. Otherwise, the drifting unit root model is used for forecasting purposes.

3.2 Accounting for a change in persistence

Strategy 4 assumes that a change in persistence occurs. This means implicitly that the size and power of the Leybourne et al. (2007) pre-test is automatically one. Moreover, the breakpoint τ is always estimated even if no break occurs. It is expected that this strategy shows a relatively good performance if the alternative is true, i.e. a change in persistence occurs. Under validity of the null hypothesis of a constant $I(1)$ process however, the accuracy may be significantly worse than the one of the random walk model which serves as the benchmark. The effects are quantified in section 5. As we assume that the direction of change is known, we use the appropriate breakpoint estimator.

Strategy 5 A compromise between Strategies 1 to 4 is to pre-test for a unit root against a change in persistence. In case of a rejection, Strategy 4 is applied in the sense that the unknown breakpoint τ is estimated and that pre-break data is is ignored when estimating the model under the alternative. The case of a non-rejection is more involved. Due to the properties of the CUSUM of squares-based pre-test, a non-rejection can be interpreted as evidence for constant persistence. It is, however, unclear whether the non-rejection is caused by a constant $I(0)$ or a constant $I(1)$ process. Therefore, Strategy 3, that is pre-testing under constant persistence, is applied in this case. Depending on the outcome of the second pre-test a trend-stationary AR(1) model or the random walk model with drift is estimated using all observations.

An important parameter that determines the degree of compromise is the nominal significance level α at which the pre-test is carried out. As $\alpha \rightarrow 0$, both size and power of the pre-test tend to zero as well and the random walk forecast model is always chosen and estimated with all observations. If $\alpha \rightarrow 1$, Strategy 4 emerges as a special case. The optimal significance level is generally unknown and we concentrate ourselves therefore on traditional values like five and ten percent. Analogously to the Dickey-Fuller test applied in Strategy 3, a linear deterministic trend is assumed when the CUSUM of squares-based pre-test is computed.

4 Bias-corrected estimators

It is a well known fact that the classic OLS estimator for the persistence parameter of the trending AR(1) model is downward-biased in small samples. Due to the fact that we consider sample sizes of $T = 100$ and 250 it seems to be reasonable to study the performance of forecasting strategies when biased-corrected estimators are applied. Therefore, we consider the median-unbiased Roy-Fuller estimator and a bootstrap estimation procedure.

4.1 Roy–Fuller estimator

The Roy–Fuller estimator provides a simple modification to the least squares estimator for α. Let $\widetilde{\alpha} = \min(\ddot{\alpha}, 1)$, where $\ddot{\alpha} = \dot{\alpha} + (C_p(\widehat{\lambda}_1) + C_{-p}(\widehat{\lambda}_{-1}))\widehat{\sigma}_1$. Here, $\dot{\alpha}$ denotes the OLS estimator for α in $\widehat{y}_t = \alpha\widehat{y}_{t-1} + \varepsilon_t$, where $\widehat{y}_t$ is the OLS residual from the regression $y_t = \beta + \gamma t + v_t$. Hence, $\dot{\alpha}$ is the OLS estimator for previously de-trended time series y_t. Furthermore, $\widehat{\sigma}_1$ denotes the standard error of $\dot{\alpha}$. Please note, that $\widehat{\lambda}_1 = (\dot{\alpha} - 1)/\widehat{\sigma}_1$ is the Dickey-Fuller unit root test statistic, while $\widehat{\lambda}_{-1}$ is a similar statistic for the unit root hypothesis $H_0 : \alpha = -1$. The two functions $C_p(\widehat{\lambda}_1)$ and $C_{-p}(\widehat{\lambda}_{-1})$ control the way in which bias correction is conducted. They are constructed in order to make $\widetilde{\alpha}$ approximately median-unbiased at $\alpha = 1$ and $\alpha = -1$, respectively. In more detail,

$$
\begin{aligned}
C_p(\widehat{\lambda}_1) &= -\lambda_{\mathsf{MED}} + d_n(\widehat{\lambda}_1 - \lambda_{\mathsf{MED}}), \qquad \text{if } \widehat{\lambda}_1 > \widehat{\lambda}_{\mathsf{MED}} \\
&= I_p(T^{-1}\widehat{\lambda}_1) - 3[\widehat{\lambda}_1 + k(\widehat{\lambda}_1 + K)]^{-1}, \qquad \text{if } -K < \widehat{\lambda}_1 \leq \widehat{\lambda}_{\mathsf{MED}} \\
&= I_p(T^{-1}\widehat{\lambda}_1) - 3\widehat{\lambda}_1 - 1, \qquad \text{if } -\sqrt{3T} < \widehat{\lambda}_1 \leq -K \\
&= 0, \qquad \text{if } \widehat{\lambda}_1 \leq -\sqrt{3T}
\end{aligned}
$$

where I_p is the integer part of $(p+1)/2$ and $\lambda_{\mathsf{MED}} = -2.18$ denotes the median of the limiting distribution of $\widehat{\lambda}_1$ if $\alpha = 1$; $k = [3n - \lambda^2_{\mathsf{MED}}(I_p+T)][\lambda_{\mathsf{MED}}(K+\lambda_{\mathsf{MED}})(I_p+T)]^{-1}$. Further details can be found in Roy and Fuller (2001).

4.2 Bootstrap bias-corrected estimator

The bootstrap procedure applied to the autoregressive model involves the generation of a large number of pseudo-data sets using the estimated coefficients and re-sampled residuals. Therefore, the pseudo-data sets resemble the dependence structure that is present in the data generating process. The biases of estimators for autoregressive parameters can be estimated as follows: generate a pseudo-data set $\{y_t^*\}_{t=1}^T$ as

$$y_t^* = \widehat{\beta} + \widehat{\gamma} t + \widehat{\alpha} y_{t-1}^* + u_t^*$$

where u_t^* is a random draw with replacement from $\{\widehat{u}_t\}_{t=2}^T$. The B sets of pseudo-data are generated, each of which gives a bootstrap parameter estimate for $\theta = (\beta, \gamma, \alpha)'$, so that we obtain $\{\theta^*(i)\}_{i=1}^B$. We obtain $\theta^* = (\beta^*, \gamma^*, \alpha^*)'$ by estimating the regression model $y_t^* = \beta + \gamma t + \alpha y_{t-1}^* + u_t^*$ and the bias of θ^* is simply estimated as $\bar{\theta}^* - \widehat{\theta}$, where $\bar{\theta}^*$ is the sample average of $\{\theta^*(i)\}_{i=1}^B$. Using this bootstrap bias-estimator, the bias-corrected estimator for θ is $\widehat{\theta}^B = \widehat{\theta} - (\bar{\theta}^* - \widehat{\theta}) = 2\widehat{\theta} - \bar{\theta}^*$. For further details regarding this estimator, see Kim (2003).

5 Monte Carlo study

This section deals with the Monte Carlo simulation setup, presentation and discussion of numerical results. The performance of forecasting strategies 1–5, see Table 1, is evaluated for a collection of different data generating processes. Before we turn to the details, some specific remarks are in order. The forecast horizon h takes values from 1 to 50. This choice corresponds to short- and medium-term forecasting when having monthly and quarterly data in mind. The sample sizes we consider are $T = 100$ and $T = 250$. If the data generating process contains a structural break, then the breakpoint is located in the middle of the in-sample period, i.e. $\tau = 0.5$. The specified loss function is the mean squared forecast error (MSFE) and the benchmark is the random walk forecast. Relative MSFEs are reported graphically. The number of Monte Carlo repetitions is set equal to 1,000. The nominal significance levels for both unit root pre-tests (DF and CUSUM of squares) are set equal to ten percent.

We consider 13 different data generating processes, see Table 2. Among these are a constant $I(1)$ process (DGP I), processes with a change in persistence (DGPs II–IV), constant $I(1)$ processes with volatility change (DGPs V–VI) or level shift (DGPs VII–VIII) or both (DGPs IX–X) and finally constant $I(1)$ processes with a broken trend (DGP XI) and an additional volatility change (DGPs XII–XIII). The exact parametrization can be found in Table 2. Typical realizations of the DGPs II–XIII are visualized in Figure 1.

5.1 Constant $I(1)$

Results for a unit root process without any structural change, i.e. constant $I(1)$ (DGP I), are depicted in Figure 2. Regardless of the applied estimator for the trending AR(1) model, we conclude that Strategy 4 (green line) which assumes a persistence change is extremely bad performing. The larger the forecast horizon, the worse is its performance. The reported numbers for $\widehat{\tau}$ in Table 3 show that approximately half of the sample is wasted. If the trending AR(1) model is estimated with all available observations (black line), the forecast accuracy is increased but still poor. Moreover, the rejection frequencies for the DF and CUSUM pre-test indicate that both tests are correctly sized which is not surprising as small-sample adjusted critical values are used.

Both pre-testing strategies (red and blue line) show satisfying performance since relative MSFEs are close to one. It is not surprising that Strategy 3 performs a bit better than Strategy 5 because Strategy 5 is based on two pre-tests with individual test sizes of ten percent while Strategy 3 requires the application of only one pre-test. It is should be noted that the performance of both strategies can be improved

Table 2: Data generating processes

Parameters	Level		Trend		Persistence		Volatility	
	β_1	β_2	γ_1	γ_2	α_1	α_2	σ_1	σ_2
Constant $I(1)$								
DGP I	0	0	0	0	1.00	1.00	1	1
Change in persistence from $I(1)$ to $I(0)$								
DGP II	0	0	0	0	1.00	0.90	1	1
DGP III	0	0	0	0	1.00	0.50	1	1
DGP IV	0	0	0	0	1.00	0.10	1	1
Constant $I(1)$, volatility change								
DGP V	0	0	0	0	1.00	1.00	5	1
DGP VI	0	0	0	0	1.00	1.00	10	1
Constant $I(1)$, level shift								
DGP VII	0	5	0	0	1.00	1.00	1	1
DGP VIII	0	10	0	0	1.00	1.00	1	1
Constant $I(1)$, level and volatility shift								
DGP IX	0	25	0	0	1.00	1.00	5	1
DGP X	0	50	0	0	1.00	1.00	10	1
Constant $I(1)$, broken trend								
DGP XI	0	0	0	1	1.00	1.00	1	1
Constant $I(1)$, broken trend and volatility shift								
DGP XII	0	0	0	1	1.00	1.00	5	1
DGP XIII	0	0	0	1	1.00	1.00	10	1

Notes: General data generating process is given by $y_t = \beta_1 + (\beta_2 - \beta_1)1(t > [\tau T]) + (\gamma_1 + (\gamma_2 - \gamma_1))t1(t > [\tau T]) + (\alpha_1 + (\alpha_2 - \alpha_1)1(t > [\tau T]))y_{t-1} + (\sigma_1 + (\sigma_2 - \sigma_1)1(t > [\tau T]))\varepsilon_t$. Innovations are standard normally distributed, $\varepsilon_t \sim N(0,1)$.

Table 3: Summary statistics for unit root pre-tests and a breakpoint estimator, $T = 100$

DGP	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
DF	0.097	0.391	0.520	0.674	0.230	0.228	0.097	0.066	0.121	0.138	0.003	0.132	0.186
CUSUM	0.096	0.307	0.531	0.806	0.324	0.487	0.100	0.085	0.399	0.465	0.123	0.258	0.443
$\widehat{\tau}$	0.533	0.412	0.397	0.401	0.644	0.639	0.577	0.628	0.639	0.641	0.626	0.643	0.640

by decreasing the nominal significance levels for the pre-tests. This will be, however, costly in terms of power and will result in less convincing performance when structural breaks occur. Moreover, the relative loss in forecast precision by applying the double pre-testing strategy 5 is limited and does not exceed twelve percent at all.

5.2 Change in persistence

Under changing persistence we observe a clear advantage for Strategies 4 and 5 which are explicitly designed for this type of DGP, see Figures 3–5. Therefore, it is not surprising that all other strategies are relatively imprecise. As the power of the CUSUM of squares-based pre-test ranges between thirty and eighty percent, Strategy 4 (with artificial power of hundred percent) performs of course best in terms of relative MSFE. When applying the OLS estimator, we observe a clear difference between the trend-stationary forecast model (S1) and the pre-testing strategy S3. S1 leads to much higher relative MSFE values than S3 which can be explained by the fact that the OLS estimator is seriously downward-biased. In case of bias-corrected estimation this difference is much less pronounced. We can see from Table 3 that the power of the Dickey-Fuller pre-test is relatively large if the change in persistence is large and vice versa. Unsurprisingly, the same conclusion holds for the CUSUM of squares-based pre-test. Moreover, it should be noted that the breakpoint estimator for τ is somewhat downward biased so that more than potentially necessary information is included in the estimation of the trending AR(1) model.

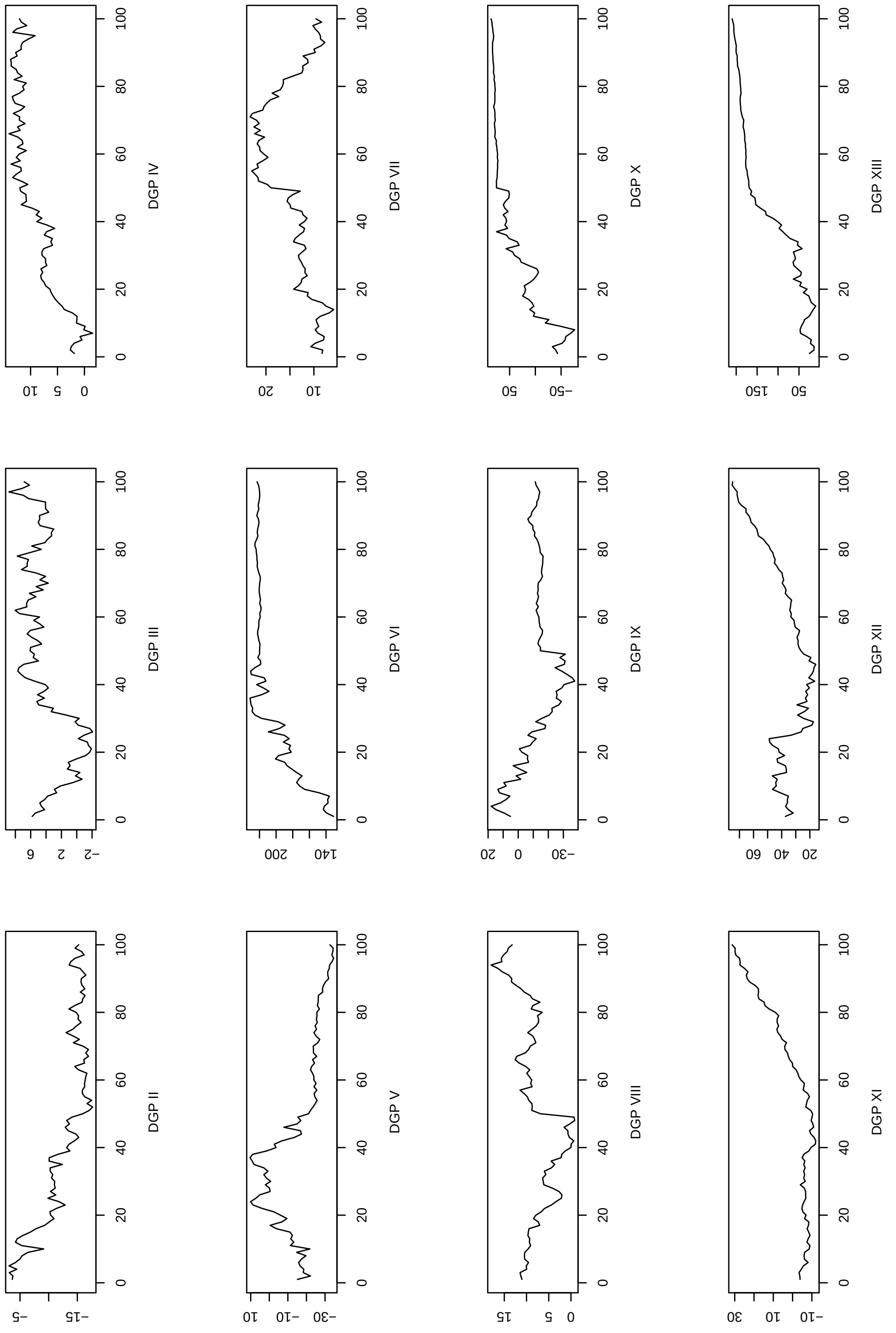

Figure 1: Typical realizations of data generating processes II - XIII, $T = 100$

5.3 Constant $I(1)$ with other types of structural changes

5.3.1 Changes in the unconditional variance

Figures 6 and 7 depict the results for unit root processes which exhibit a structural change in the unconditional variance of the error term. It can be seen that strategies S4 and S5 are able to beat the benchmark. If the volatility break is large, then gains can be achieved by applying S4 or S5 for nearly all forecast horizons. For moderate breaks, the evidence is mixed. For h greater than 15, the double pre-testing strategy S5 is better performing than the benchmark strategy S2. Strategies 1 and 3 suffer from the fact that (i) the trend-stationary model is misspecified and that (ii) the Dickey-Fuller test has non-trivial power when y_t is generated according to DGPs V and VI, see Table 3. Furthermore, the CUSUM of squares-based pre-test has even more power and is therefore able to detect the structural break although it is not a change in persistence. The breakpoint estimator is upward-biased in this case implying that some data points are wasted. It is worthwhile noting that strategy S1 shows a relatively bad performance when the biased OLS estimator is applied instead of a bias-corrected estimator. The differences between the the bootstrap approach and Roy-Fuller technique are very little for the DGPs that has been considered so far.

5.3.2 Level shifts with and without changes in the unconditional variance

We now turn to the results for unit root processes with level shifts which may be subject to changes changes in the unconditional variance of the error term. The DGPs VII and VIII cover only level shifts, while DGPs IX and X include both types of breaks at the same point in time, i.e. $\tau_{\beta,\sigma} = 0.5$. Results are shown graphically in Figures 8–11. If only level shifts occur, the picture is nearly identical to the case DGP I, where a unit root process with out any type of breaks is specified, see Figures 8 and 9. The reason is that both pre-tests are more or less correctly sized, see Table 3. The breakpoint estimator is a bit upward-biased which results in a poor performance of S4. It is interesting to note that the $I(0)$ plus trend model performs best, although relatively weak when compared to the random walk forecast, if the Roy-Fuller estimator is applied. If additionally to the level shift a change in volatility occurs at the same time, results are a mixture of the results obtained for individual types of breaks, see Figures 10 and 11. The shape of MSFE ratios look very much the same like in the case of DGPs V and VI (only volatility change), although they are upward shifted which is due the structural break on the mean of the process.

6 Conclusions

This paper considers forecasting trending autoregressive time series which may be subject to several types of structural breaks. Special attention is paid to the concept of changing persistence which constitutes a new type of structural break which has been analyzed recently. A change in persistence means that the time series process switches from stationarity to non-stationarity over time, or vice versa. Such a change may have important implications for forecasting. Several forecasting strategies are analyzed in a large-scale Monte Carlo study. A newly proposed strategy which builds upon pre-tests for changing persistence and unit roots shows relatively good overall performance. Its usefulness is underlined by the fact that it is more accurate than other approaches even if other types of structural breaks occur like level shifts, broken trends or changes in the unconditional error variance. The performance of this forecasting strategy can be improved by using a bias-corrected estimator for the trending autoregressive forecasting model rather than the classic OLS method.

References

Andrews, D.W.K., and Chen,H.Y. (1994): "Approximate median-unbiased estimation of autoregressive models." *Journal of Business & Economic Statistics* 12, 187—204.

Barsky, R.B. (1987): ”The Fisher hypothesis and the forecastibility and persistence of inflation.” *Journal of Monetary Economics* 19, 3–24.

Belaire-Franch, J. (2005): ”A Proof of the power of Kim’s test against stationary processes with structural breaks.” *Econometric Theory* 21, 1172–1176.

Busetti, F. and A.M.R. Taylor (2004): ”Tests of Stationarity against a Change in Persistence.” *Journal of Econometrics* 123, 33–66.

Burdekin, R.C.K. and P.L. Siklos (1999): ”Exchange rate regimes and shifts in inflation persistence: does nothing else matter?” *Journal of Money, Credit and Banking* 31, 235–247.

Cavaliere, G. and A.M.R. Taylor (2006): ”Testing for a Change in Persistence in the Presence of a Volatility Shift.” *Oxford Bulletin of Economics and Statistics* 68, 761–781.

DeLong, J.B. and L.H. Summers (1988): ”How does macroeconomic policy affect output?” *Brookings Papers on Economic Activity* 2, 433–494.

Dickey, D.A. and W.A. Fuller (1979): ”Distribution of the Estimators for Autoregressive Time Series With a Unit Root.” *Journal of the American Statistical Association* 74, 427–431.

Diebold, F.X and L. Kilian (2000): ”Unit-Root Tests Are Useful for Selecting Forecasting Models.” *Journal of Business & Economic Statistics* 18, 265—273.

Hakkio, C.S. and M. Rush (1991): ”Is the budget deficit too large?” *Economic Inquiry* 29, 429–445.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2006): ”Modified tests for a change in persistence” *Journal of Econometrics* 134, 441–469.

Kang, K.H., C.J. Kim and J. Morley (2006): ”Regime Shifts in U.S. Inflation Persistence.” *mimeo.*

Kim, J. (2000): ”Detection of Change in Persistence of a Linear Time Series.” *Journal of Econometrics* 95, 97–116.

Kim, J., J. Belaire Franch and R. Badilli Amador (2002): ”Corrigendum to Detection of Change in Persistence of a Linear Time Series.” *Journal of Econometrics* 109, 389–392.

Kim, J.H. (2003): ”Forecasting autoregressive time series with bias-corrected parameter estimators.” *International Journal of Forecasting* 19, 493—502.

Kumar, M.S. and T. Okimoto (2007): ”Dynamics of Persistence in International Inflation Rates.” *Journal of Money, Credit and Banking* 39, 1457–1479.

Leybourne, S.J., T. Kim, V. Smith and P. Newbold (2003): ”Tests for a Change in Persistence against the Null of Difference Stationarity.” *Econometrics Journal* 6, 291–311.

Leybourne, S.J., A.M.R. Taylor and T. Kim (2007): ”CUSUM of Squares-Based Tests for a Change in Persistence.” *Journal of Time Series Analysis* 28, 408–433.

Mankiw, N.G., J.A. Miron, and D.N. Weil (1987): ”The adjustment of expectations to change in regime: a study of the founding of the federal reserve.” *American Economic Review* 77, 358–374.

O’Reilly, G. and K. Whelan (2005): ”Has Euro Area Inflation Persistence Changed Over Time?” *The Review of Economics and Statistics* 87, 709–720.

Perron, P. (2006): ”Dealing with Structural Breaks.” *in: T.C. Mills and K. Patterson (eds), Palgrave Handbook of Econometrics, vol.1: Econometric Theory*, Palgrave Macmillan Publishers, 278–352.

Pesaran, M.H. and A. Timmermann (2005): "Small sample properties of forecasts from autoregressive models under structural breaks." *Journal of Econometrics* 129, 183–217.

Pivetta, F., Reis, R. (2007): "The persistence of inflation in the United States." *Journal of Economic Dynamics and Control* 31, 1326 – 1358.

R Development Core Team (2004): "R: A language and environment for statistical computing." Available at www.r-project.org.

Roy, A. and Fuller, W.A. (2001): "Estimation for autoregressive time series with a root near one." *Journal of Business & Economic Statistics* 19, 482—493.

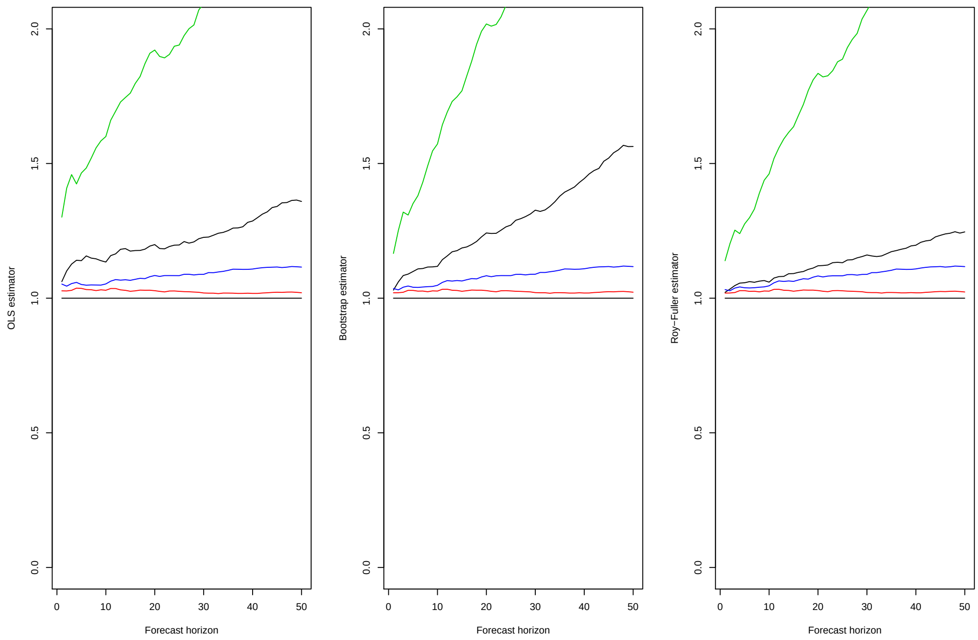

Figure 2: DGP I: Constant $I(1)$ ($\alpha_2 = 1$), $T = 100$

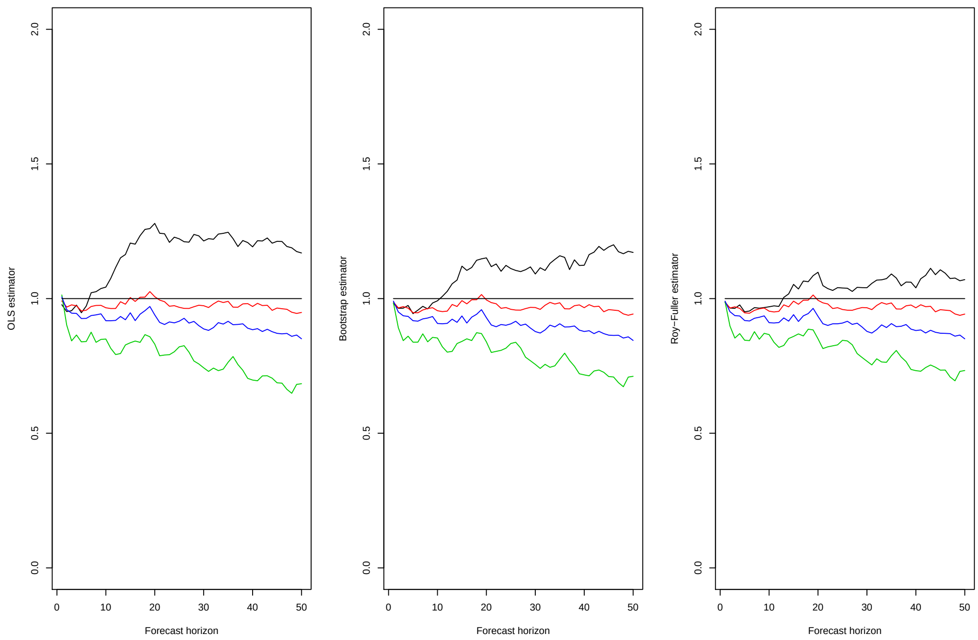

Figure 3: DGP II: Change in persistence ($\alpha_2 = 0.7$), $T = 100$

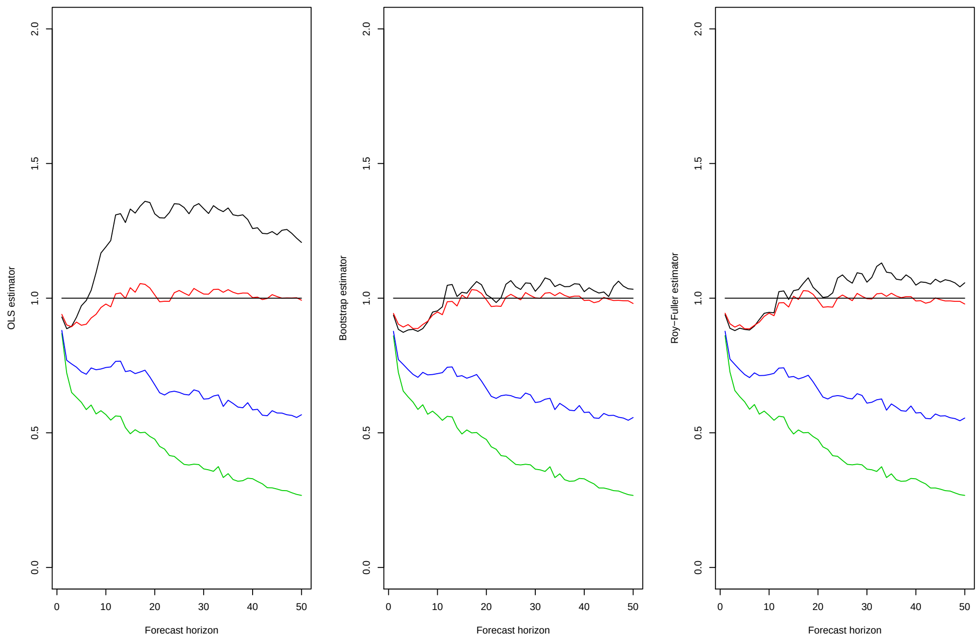

Figure 4: DGP III: Change in persistence ($\alpha_2 = 0.5$), $T = 100$

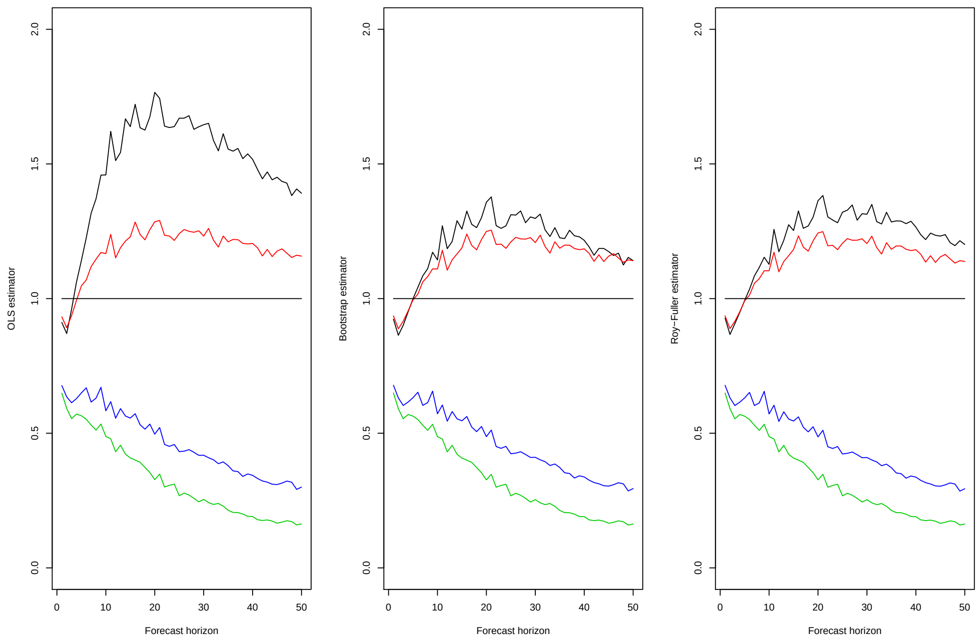

Figure 5: DGP IV: Change in persistence ($\alpha_2 = 0.1$), $T = 100$

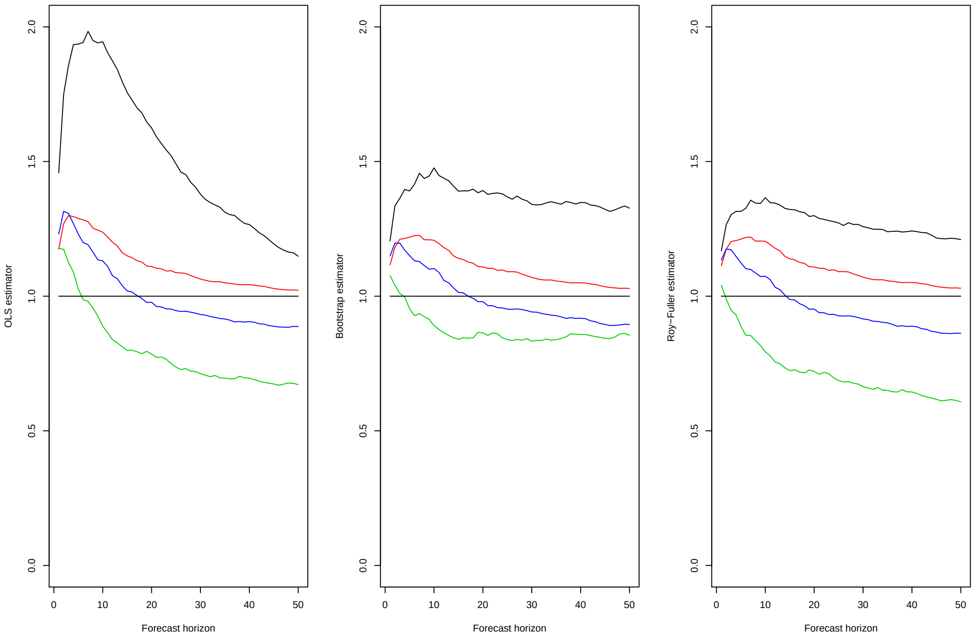

Figure 6: DGP V: Volatility change ($\sigma_1 = 5$), $T = 100$

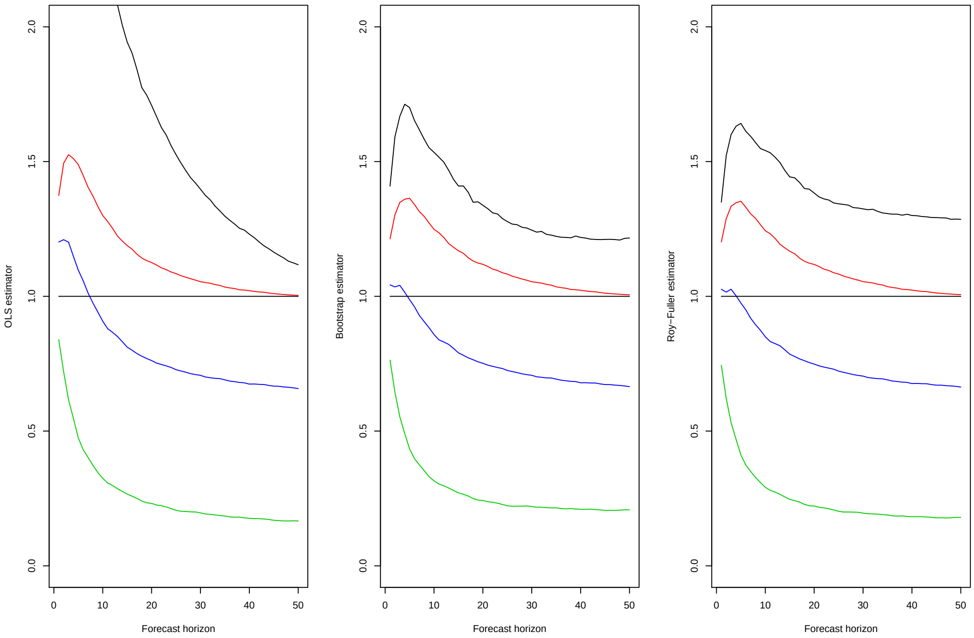

Figure 7: DGP VI: Volatility change ($\sigma_1 = 10$), $T = 100$

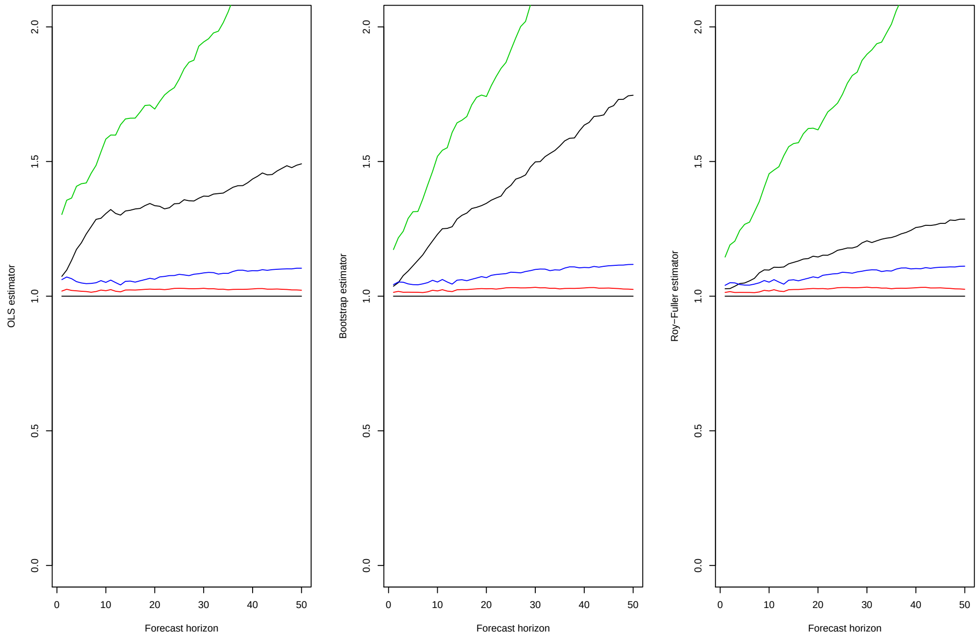

Figure 8: DGP VII: Level shift ($\beta_2 = 5$), $T = 100$

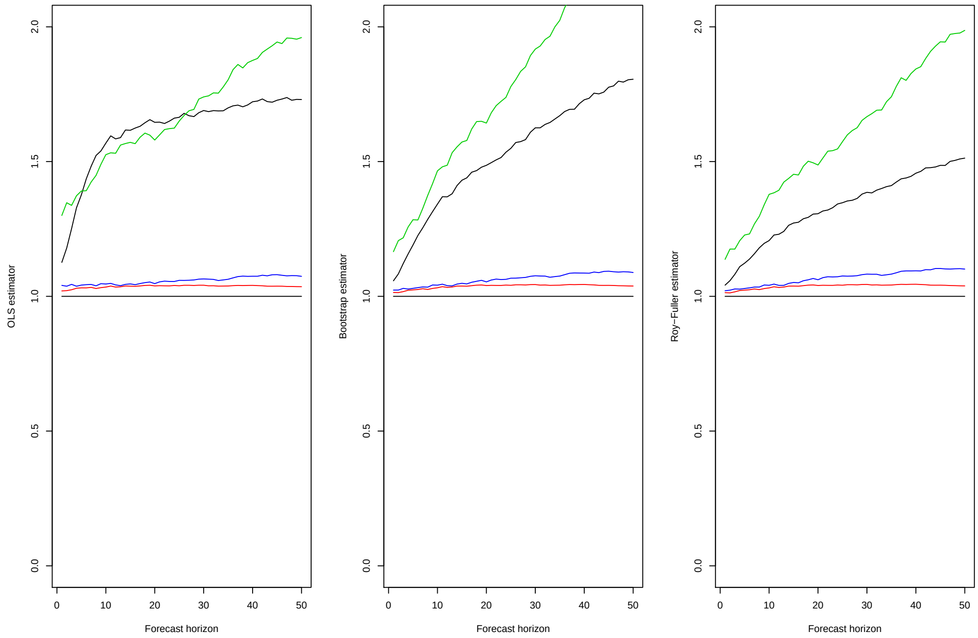

Figure 9: DGP VIII: Level shift ($\beta_2 = 10$), $T = 100$

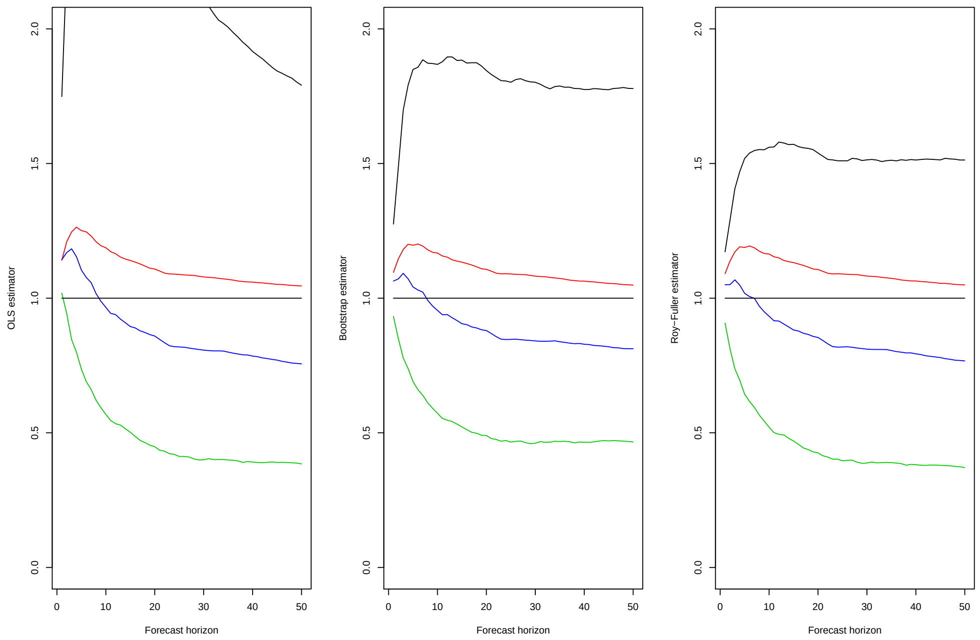

Figure 10: DGP IX: Level and volatility shift ($\beta_2 = 25, \sigma_1 = 5$), $T = 100$

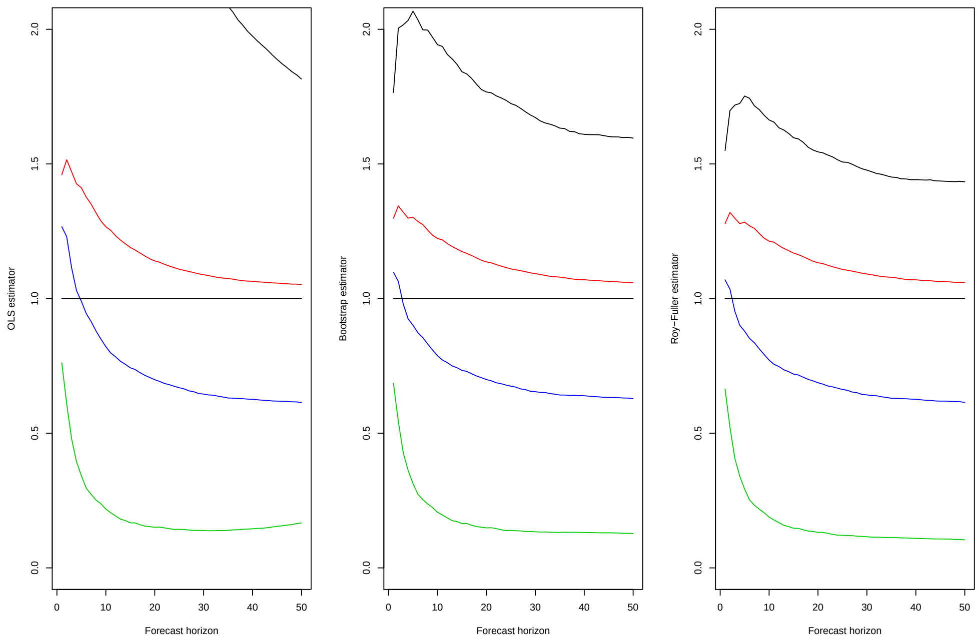

Figure 11: DGP X: Level and volatility shift ($\beta_2 = 50, \sigma_1 = 10$), $T = 100$