Department of Economics Discussion Paper Series

Selecting a Model for Forecasting

Jennifer L. Castle, Jurgen A. Doornik and David F. Hendry

Number 861
November, 2018

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

Selecting a Model for Forecasting

Jennifer L. Castle, Jurgen A. Doornik and David F. Hendry*
Economics Department and Institute for New Economic Thinking at the
Oxford Martin School, University of Oxford, UK

November 9, 2018

Abstract

We investigate the role of the significance level when selecting models for forecasting as it controls both the null retention frequency and the probability of retaining relevant variables when using binary decisions to retain or drop variables. Analysis identifies the best selection significance level in a bivariate model when there are location shifts at or near the forecast origin. The trade-off for selecting variables in forecasting models in a stationary world, namely that variables should be retained if their non-centralities exceed 1, applies in the wide-sense non-stationary settings with structural breaks examined here. The results confirm the optimality of the Akaike Information Criterion for forecasting in completely different settings than initially derived. An empirical illustration forecasting UK inflation demonstrates the applicability of the analytics. Simulation then explores the choice of selection significance level for 1-step ahead forecasts in larger models when there are unknown location shifts present under a range of alternative scenarios, using the multipath tree search algorithm, *Autometrics* (Doornik, 2009), varying the target significance level for the selection of regressors. The costs of model selection are shown to be small. The results provide support for model selection at looser than conventional settings, albeit with many additional features explaining the forecast performance, with the caveat that retaining irrelevant variables that are subject to location shifts can worsen forecast performance.

Keywords: Model selection; forecasting; location shifts; significance level; *Autometrics*

1 Introduction

There are many approaches to formulating models when the sole objective is forecasting, from the very parsimonious through to large systems. However, there is little agreement on which approaches perform best on a forecasting criterion: see Makridakis and Hibon (2000) and Fildes and Ord (2002) for evidence from forecast competitions. Clements and Hendry (2001) suggest this lack of agreement is the outcome of intermittent distributional shifts differentially impugning alternative formulations. We address this critique by analysing the selection of models to optimise mean square forecast error performance in wide-sense non-stationary settings with structural breaks.

The paper focuses on regression models that are linear in the parameters, and considers model selection that is controlled by the nominal significance level for statistical significance when selecting forecasting models subject to breaks. Loose significance levels (such as those implied by AIC: see Akaike, 1973) have been shown to be optimal to select regression models for stationary processes if evaluating on a 1-step ahead mean-square forecast error (MSFE) criterion; see Shibata (1980) who showed that AIC

*Financial support from the Robertson Foundation (award 9907422) and Institute for New Economic Thinking (grant 20029822) is gratefully acknowledged. We thank participants at the 2018 International Symposium of Forecasting, Michael P. Clements, Andrew Martinez, Felix Pretis, and Sophocles Mavroeidis for helpful comments and suggestions, and Michael McCracken for suggesting comparisons with bagging, which will be reported in a later paper.

is an asymptotically efficient selection method when the DGP is an infinite order process, also see Ing and Wei (2003). Many other criteria have been proposed that aim to have optimal properties in certain settings but information criteria alone are not a sufficient principle for selecting models as they do not ensure congruence, so a mis-specified model could be selected: see Bontemps and Mizon (2003). In the simulation exercise, we explore general-to-specific (*Gets*) model selection, to narrow down the class of forecasting models to undominated models, yielding the benefits of well-specified encompassing models in-sample, albeit non-stationarities may preclude those benefits continuing over the forecast horizon.

Here we investigate the significance level for selecting variables, to establish how tight a selection criterion should be when the specific purpose is forecasting facing a potentially non-stationary environment induced by location shifts of the conditioning variables' distributions. The analysis commences with a bivariate conditional model that is part of a 3-variable system in which the selection decision is whether to retain or exclude one of the regressors. Such a design is empirically relevant as demonstrated by an example forecasting UK inflation, where autoregressive models are augmented with the unemployment rate. This bivariate model is analysed both for stationary and non-stationary settings where location shifts occur at or near the forecast origin. The static setting still requires forecasts of the conditioning variables, and alternative forecasting devices are considered, including the two extremes of the class of robust forecasting devices proposed by Castle, Clements, and Hendry (2015). The results confirm that regressors should be retained for forecasting if their non-centralities exceed 1, regardless of whether or not there is a structural break, or of the forecasting device used. These analytic results map to a selection significance level of 16% in the bivariate case, much looser than conventional significance levels used. The results closely match that of AIC, which can be interpreted as a likelihood ratio χ^2 test for a pair of nested models with 1 degree of freedom and a penalty of 2, and also gives a significance level of approximately 16%: see Pötscher (1991) and Leeb and Pötscher (2009).

In their taxonomy of forecast errors in systems where some conditioning variables are forecast off-line, Hendry and Mizon (2012) show that a key source of forecast failure is any induced shift in the equilibrium mean of the variable being forecast, irrespective of whether or not those conditioning variables are included in the forecasting model. Consequently, we include a simulation exercise that evaluates a wide range of settings including larger models, break types and magnitudes at or near the forecast origin and the method of forecasting. We consider a range of significance levels from the very tight (0.001), eliminating almost all potentially irrelevant variables, to the very loose (0.50), enabling retention of relevant variables even if they are only marginally significant. The results enable evaluation of the costs when forecasting of omitting relevant variables and from incorrectly retaining irrelevant variables. Overall, the results support looser than conventional significance levels for selecting forecasting models, with a 10% target significance level often producing superior forecasts.

The paper is structured as follows. First, section 2 motivates the paper and then section 3 formulates the analysis and section 4 considers the choice of selection significance level for forecasting in a stationary DGP. Then section 5 analyses selection in a non-stationary DGP where a location shift occurs out-of-sample in one of the regressors, and investigates the consequences of that variable's inclusion or exclusion in the forecasting model. Section 6 considers the impacts on selection of in-sample shifts using different forecasting devices and section 7 summarises the analytic results. Section 8 presents the simulation evidence on the performance of the various approaches, examining the preferred significance level to minimize MSFE across experiment designs, and section 9 concludes. Appendix A includes the analytic derivations and appendix B provides supplementary tables.

2 Motivation

Two popular models within the large literature on inflation forecasting include single-equation forecasting models based on past inflation (univariate models such as ARIMA) and what are often termed 'Phillips curve forecasts', augmenting the univariate model with an activity variable such as the unemployment rate or output gap, see Stock and Watson (2009). The framework considered below, although

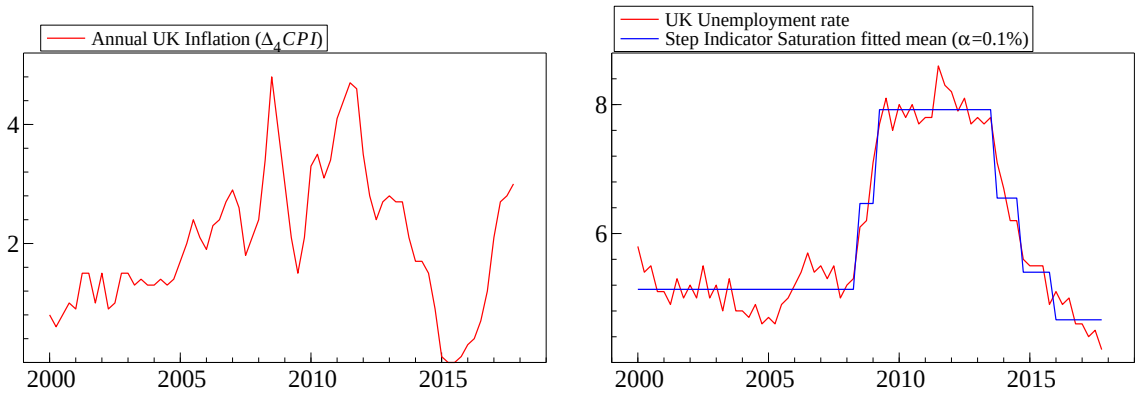

Figure 1: (**a**) Annual UK inflation rate (CPI); (**b**) Annual UK unemployment rate, with SIS detected mean shifts at $\alpha = 0.1\%$.

static, can be applied to these two models where the econometrician wishes to determine whether to augment a univariate forecasting model with the contemporaneous unemployment rate. This 'exogenous' variable is subject to breaks in the form of location shifts, which may occur at or near the forecast horizon. Figure 1 records the annual percentage change in UK consumer price inflation, π, and the UK unemployment rate as a percentage, U_r, along with a broken mean obtained by Step Indicator Saturation (SIS) at $\alpha = 0.1\%$.[1]

The analytics derived below correspond to a Phillips curve formulation (M_1), a univariate AR model (M_2) and selection applied to the unemployment rate using a significance level of 0.16 (M_3):

$$\begin{aligned}
\mathsf{M}_1: \quad & \pi_{t+1} - \pi_t = \mu + \sum_{i=1}^{4} \beta_{\pi,i} \Delta \pi_{t-i} + \sum_{i=0}^{4} \beta_{U_r,i} U_{r,t-i} + \nu_{1,t+1} \\
\mathsf{M}_2: \quad & \pi_{t+1} - \pi_t = \mu + \sum_{i=1}^{4} \beta_{\pi,i} \Delta \pi_{t-i} + \nu_{2,t+1} \\
\mathsf{M}_3: \quad & \pi_{t+1} - \pi_t = \mu + \sum_{i=1}^{4} \beta_{\pi,i} \Delta \pi_{t-i} + \sum_{i=0}^{4} \beta^*_{U_r,i} U_{r,t-i} + \nu_{3,t+1}
\end{aligned}$$

where * denotes selection using *Autometrics* at $\alpha = 16\%$, i.e. $\beta^*_{U_r}$ has a zero when a variable is not selected. Dynamics are included to account for any autocorrelation. The forecasting models are estimated over $2000q1 - 2013q4$, producing 1-quarter ahead inflation forecasts for 2014q1-2017q4 evaluated on MSFE. Selection at 16% results in $U_{r,t-1}$ being retained, with a p−value of 0.149, so would not be retained under a commonly used 5% significance level.

Table 1 reports the pseudo out-of-sample RMSFEs. Three cases are considered corresponding to the analytics below; (a) known $U_{r,t}$, (b) forecast $\widehat{U}_{r,t}$ using the in-sample mean, and (c) forecast $\widehat{U}_{r,t}$ using a random walk, i.e. $U_{r,t-1}$. Although there is little difference across the models with differences in RMSFEs not statistically significant, selection at a loose significance level outperforms if $U_{r,t}$ is known. As this is infeasible, the random walk applied to selecting the regression model using a significance level of 16% matches that of the known $U_{r,t}$, so selection can be beneficial. We now generalize the framework to establish the optimal significance level for selection.

[1] SIS was conducted on the unemployment rate with a forced intercept at a selection significance level of $\alpha = 0.001$. See Castle, Doornik, Hendry, and Pretis (2015) for details of SIS.

Table 1: Root Mean Square Forecast Errors ($\times 100$) for annual inflation over 2014q1-2017q4.

	M_1	M_2	M_3
Known $U_{r,t}$	0.535	0.530	0.515
Mean forecast for $U_{r,t}$	0.519	0.530	0.542
Random Walk forecast for $U_{r,t}$	0.549	0.530	0.515

3 The analytic design

We initially focus on a static DGP with known future exogenous regressors to highlight the main issues, before extending to allow for mean-shifts at the forecast origin in unknown future regressors. We examine the in-sample mean estimator (optimal in the stationary case) and a random walk forecast, both under a location shift, with the aim of determining the best significance level for selecting a regression model based on a single t-test. Although a static DGP may seem restrictive, the formulation is a 3 variable VAR, so the main role of adding dynamics would be slow adjustments to location shifts. Such dynamics are considered in the simulation exercise in section 8.

The DGP is a static VAR given by:

$$\begin{pmatrix} 1 & -\beta_1 & -\beta_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \eta_{1,t} \\ \eta_{2,t} \end{pmatrix}, \tag{1}$$

where $\mathbf{y}_t' = (y_t : x_{1,t} : x_{2,t})$ with:

$$\mathbf{y}_t \sim \mathsf{IN}_3 \left[\boldsymbol{\mu}, \boldsymbol{\Sigma}\right], \tag{2}$$

where $\boldsymbol{\mu}_t' = \left(\mu_y : \mu_1 : \mu_2\right)$ and, without loss of generality we set $\mathsf{V}\left[x_{i,t}\right] = \sigma_{ii}^2 = 1$, such that:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\epsilon^2 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}. \tag{3}$$

While it may be more intuitive to lag the exogenous regressors in the DGP for forecasting purposes, none of the results would change and the current set up naturally leads to analysis of the forecasting models for the contemporaneous exogenous regressors, allowing a comparison of alternative devices and an assessment of open models, see Hendry and Mizon (2012). Throughout, we take the in-sample estimates of the μ_i to be sufficiently precise that their sampling variation can be neglected, and use the population values to focus on the impacts of location shifts. Then (1) implies $\mathsf{E}[y_t] = \mu_y = \beta_0 + \beta_1\mu_1 + \beta_2\mu_2$ with:

$$y_t = \mu_y + \beta_1 \left(x_{1,t} - \mu_1\right) + \beta_2 \left(x_{2,t} - \mu_2\right) + \epsilon_t. \tag{4}$$

3.1 Selecting a model

Considering the conditional model (4) we compare M_1, which includes both weakly exogenous regressors, and M_2, which excludes x_2:

$$\mathsf{M}_1 : y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \epsilon_t \tag{5}$$

$$\mathsf{M}_2 : y_t = \phi_0 + \gamma_1 x_{1,t} + \nu_t, \tag{6}$$

where Appendix A.1 summarises ϕ_0, γ_1, ν_t and σ_ν^2.

The choice between M_1 and M_2 will depend on a test of significance of $x_{2,t}$. The population non-centrality of the t-test, $\mathsf{t}_{\beta_2=0}$, of the null that $\beta_2 = 0$, denoted ψ, is given by:

$$\psi^2 = \frac{T\beta_2^2\left(1-\rho^2\right)}{\sigma_\epsilon^2}. \tag{7}$$

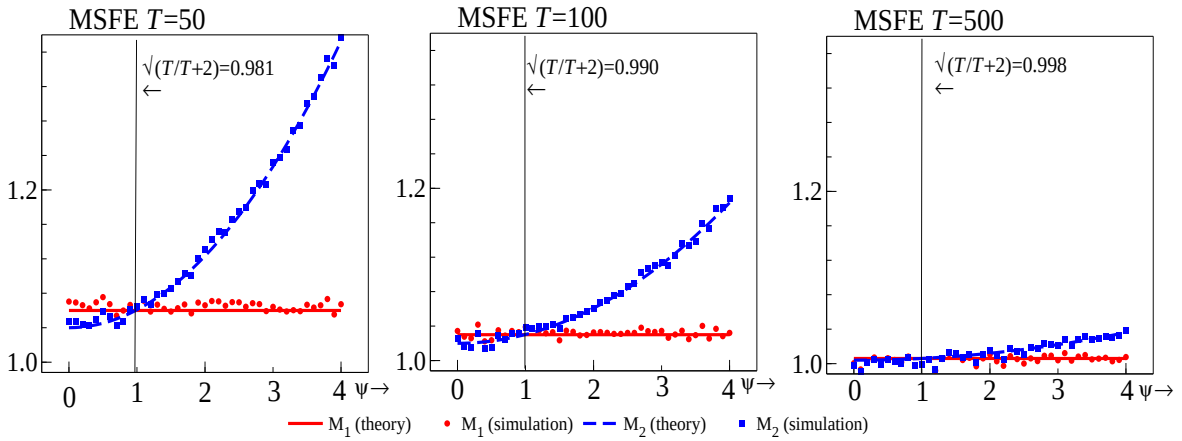

Figure 2: MSFE comparisons between M_1 (solid lines computed from (8) and circles by simulation) and M_2 (dashed line computed from (9) using (48) and squares by simulation), with $\beta_0 = 5$, $\beta_1 = 1$, $\sigma_\epsilon^2 = 1$, $\mu_1 = \mu_2 = 2$, $\rho = 0.5$ and $T = 50, 100, 500$, with an additional observation used to calculate the 1-step ahead MSFEs. (**a**) $T = 50$; (**b**) $T = 100$; (**c**) $T = 500$. Simulations based on $M = 100,000$ replications.

4 Selection when forecasting in a stationary DGP

4.1 Comparing M_1 and M_2 forecast errors

First we compute the 1-step ahead MSFEs from M_1, denoted $\widehat{\cdot}$, and M_2, denoted $\widetilde{\cdot}$, and look at the conditions for $\mathsf{MSFE}_2 \leq \mathsf{MSFE}_1$. An estimated intercept is always retained, which maintains comparability between M_1 and M_2.

When there are no breaks, the parameter estimates for M_1 are unbiased, $\mathsf{E}\left[\widehat{\epsilon}_{T+1|T}\right] = 0$, so the MSFE of M_1 is:

$$\mathsf{MSFE}_1 = \mathsf{E}\left[\widehat{\epsilon}_{T+1|T}^2\right] = \sigma_\epsilon^2\left(1 + \frac{3}{T}\right), \tag{8}$$

which is the unconditional MSFE formula for the impact of estimating 3 parameters, under the assumption of correct model specification and no breaks. For M_2, despite the mis-specification when $\beta_2 \neq 0$, $\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}\right] = 0$ and the MSFE is:

$$\mathsf{MSFE}_2 = \mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}^2\right] = \sigma_\nu^2\left(1 + \frac{2}{T}\right), \tag{9}$$

where $\sigma_\nu^2 = \sigma_\epsilon^2\left(1 + T^{-1}\psi^2\right) \geq \sigma_\epsilon^2$. There is one less parameter to estimate, traded off against a larger equation variance (see Appendix A.2 for derivations).

If the objective is to minimize MSFE, M_2 should be used to forecast when $\mathsf{MSFE}_2 \leq \mathsf{MSFE}_1$, which requires:

$$\sigma_\nu^2\left(1 + \frac{2}{T}\right) - \sigma_\epsilon^2\left(1 + \frac{3}{T}\right) = \frac{\sigma_\epsilon^2}{T}\left[\psi^2\left(1 + \frac{2}{T}\right) - 1\right] \leq 0 \tag{10}$$

which occurs when $\psi^2 \leq T/(T+2)$.

Figure 2 records the 1-step ahead values of MSFE_1 and MSFE_2 for known $x_{i,T+1}$, $i = 1, 2$, for the DGP given by (1) and (2) where β_2 varies along the horizontal axis to get a range of non-centralities in the set $\psi = [0, 4]$ using (7). The results confirm that x_2 should be retained if its non-centrality exceeds approximately 1, and the result converges to 1 as $T \to \infty$, as the information content of the regressor outweighs the parameter estimation cost for 1-step forecasts, regardless of the correlation between x_1 and x_2.

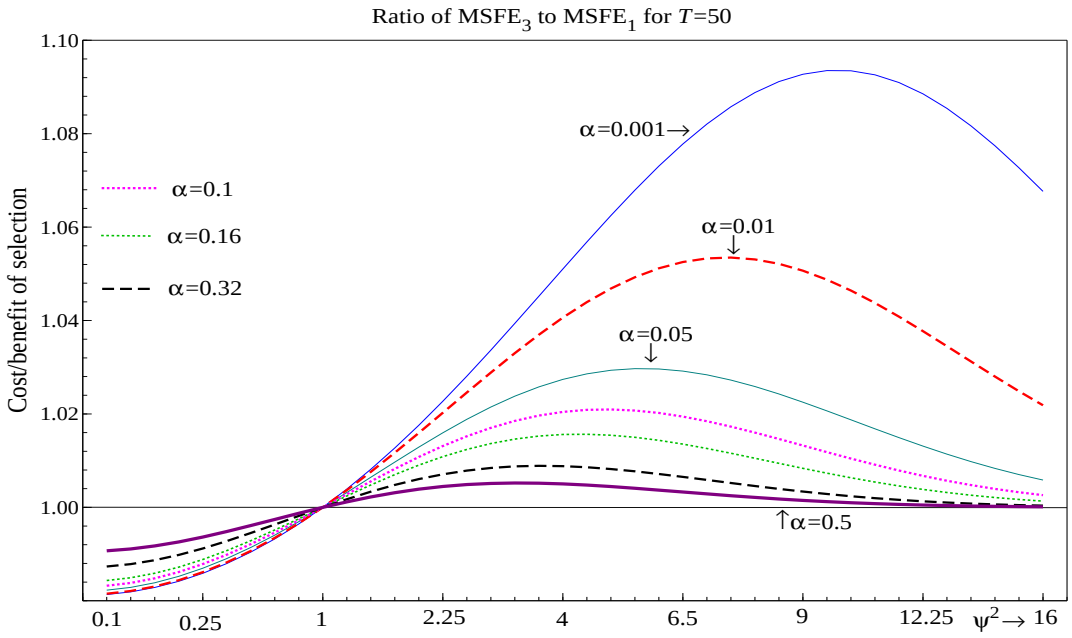

Figure 3: The costs/benefits of selection measured by $\frac{\mathsf{MSFE}_3}{\mathsf{MSFE}_1} = 1 + (T+3)^{-1}\left(1 - \mathsf{p}_\alpha[\psi]\right)\left(\psi^2 - 1\right)$ for $T = 50$ with $\beta_0 = 5$, $\beta_1 = 1$, $\sigma_\epsilon^2 = 1$, $\mu_1 = \mu_2 = 2$, $\rho = 0.5$.

4.2 Selecting regressors

Although M_1 and M_2 provide the extremes of always/never retain x_2, in practice selection will be applied. At a significance level of α and critical value c_α, then $x_{2,t}$ will be omitted if, from (5), $\mathsf{t}^2_{\beta_2=0} < c_\alpha^2$, so using the approximation that:

$$\mathsf{t}_{\beta_2=0} = \frac{\widehat{\beta}_2}{\mathsf{SE}\left[\widehat{\beta}_2\right]} \approx \frac{\sqrt{T\left(1-\rho^2\right)}\widehat{\beta}_2}{\sigma_\epsilon},$$

which implies:

$$\widehat{\beta}_2^2 < \frac{c_\alpha^2 \sigma_\epsilon^2}{T\left(1-\rho^2\right)}. \tag{11}$$

Thus, retention of $x_{2,t}$ will depend on α and ψ^2 for a given draw.

Forecasts in repeated sampling will be based on a mixture of M_1 and M_2 depending on whether $x_{2,t}$ is retained in each draw. The MSFE of the selected model, called M_3, will be a weighted average of the MSFEs of M_1 and M_2, with the weights given by the probability that $x_{2,t}$ is retained:

$$\begin{aligned}
\mathsf{MSFE}_3 &= \mathsf{p}_\alpha[\psi]\mathsf{MSFE}_1 + \left(1 - \mathsf{p}_\alpha[\psi]\right)\mathsf{MSFE}_2 \\
&= \mathsf{MSFE}_1 + \left(1 - \mathsf{p}_\alpha[\psi]\right)\left(\mathsf{MSFE}_2 - \mathsf{MSFE}_1\right) && (12) \\
&\approx \mathsf{MSFE}_1 + \sigma_\epsilon^2 T^{-1}\left(1 - \mathsf{p}_\alpha[\psi]\right)\left(\psi^2 - 1\right), && (13)
\end{aligned}$$

where ψ^2 is given by (7), with:

$$\mathsf{p}_\alpha[\psi] = \Pr\left(\mathsf{t}^2_{\beta_2=0} \geq c_\alpha^2\right). \tag{14}$$

From the last term in (13) it is clear that $\mathsf{MSFE}_3 \leq \mathsf{MSFE}_1$ whenever $\psi^2 \leq 1$. Moreover, $\mathsf{p}_\alpha[\psi]$ will be low when $\psi^2 \leq 1$, so M_2 will usually be selected. Note that $\mathsf{p}_\alpha[\psi] = \alpha$ when $\beta_2 = 0$. However, MSFE_3 is a highly non-linear function of ψ^2 entering directly and indirectly, as well as of α which also influences $\mathsf{p}_\alpha[\psi]$ non-linearly.

Figure 3 records the ratio of MSFE_3 to MSFE_1, for a range of ψ^2, which from (13) is given by:

$$\frac{\mathsf{MSFE}_3}{\mathsf{MSFE}_1} \approx 1 + (T+3)^{-1}\left(1 - \mathsf{p}_\alpha[\psi]\right)\left(\psi^2 - 1\right).$$

Selection delivers a 1.8% improvement in MSFE relative to M_1 when $\psi^2 = 0$ with $\alpha = 0.05$ or tighter, but for looser α, e.g. at 0.5, $\mathsf{p}_\alpha[\psi] = 0.5$ when $x_{2,t}$ is irrelevant, the benefits of selection are halved. Selection is most costly at intermediate non-centralities, where, for example, the largest increase in MSFE relative to M_1 is 3% at $\alpha = 0.05$ for $T = 50$, but is over 9% for $\alpha = 0.001$ at its peak. The hump shape reflects the non-linear trade-off as the non-centrality of $x_{2,t}$ increases, from the cost of omitting $x_{2,t}$ rising as its signal is stronger, but the probability of retaining $x_{2,t}$ also increases. The magnitude of cost/benefit depends on T, so shrinks as the sample size increases.

The selection rule that $x_{2,t}$ should be retained if $\psi^2 > 1$ is evident $\forall\alpha$, but unfortunately the forecaster does not know ψ^2. If it was known, the optimal α is 0 for $\psi^2 < 1$ and 1 for $\psi^2 > 1$. We next look at the choice of α to minimize cost/maximize benefit in terms of improvements in 1-step ahead MSFEs for an unknown ψ^2.

4.3 The choice of selection significance level

Given (11), which is required for x_2 to be excluded at the chosen significance level assuming unbiasedness, on average that inequality requires (when $\mathsf{V}[\cdot]$ denotes variance):

$$\mathsf{E}\left[\widehat{\beta}_2^2\right] = \mathsf{V}\left[\widehat{\beta}_2\right] + \beta_2^2 = \beta_2^2 + \frac{\sigma_\epsilon^2}{T\left(1-\rho^2\right)} < \frac{c_\alpha^2 \sigma_\epsilon^2}{T\left(1-\rho^2\right)}. \tag{15}$$

Equating that inequality, which β_2^2 must satisfy, with $\psi^2 < 1$ from (10) gives the boundary for the critical value c_α in which selection results in a smaller MSFE due to the omission–estimation trade-off:

$$\beta_2^2 = \frac{\sigma_\epsilon^2\left(c_\alpha^2 - 1\right)}{T\left(1-\rho^2\right)} \leq \frac{\sigma_\epsilon^2}{T\left(1-\rho^2\right)}.$$

This implies that $c_\alpha^2 = 2$ at the boundary (also see Clements and Hendry, 1998, Ch.12), or an approximate significance level of $\alpha = 0.16$.

Computing the theoretical probability of retaining x_2 for $\beta_2 > 0$ at $\alpha = 0.16$ using $\mathsf{E}[\mathsf{t}_{\widehat{\beta}_2}] = \psi$:

$$\Pr\left(\mathsf{t}_{\widehat{\beta}_2} \geq c_\alpha\right) = \Pr\left(\mathsf{t}_{\widehat{\beta}_2} - \psi \geq c_\alpha - \psi\right),$$

we obtain the retention probabilities in table 2, with the corresponding retention probabilities for $\alpha = 0.05$ recorded for comparison. These results are close to the implied significance level for the AIC in

Table 2: Retention probabilities for individual $\mathsf{t}-$tests given $\mathsf{E}[\mathsf{t}_{\widehat{\beta}_2}] = \psi$, and five independent regressors with the same non-centrality, where bold cells indicate the grey dots on Figure 4, recorded as $1-(\mathsf{p}_\alpha[\psi])^5$.

ψ	1	2	3	4
$\mathsf{p}_{0.16}[\psi]$	0.34	0.72	0.94	0.995
$\mathsf{p}_{0.05}[\psi]$	0.16	0.51	0.85	0.98
$(\mathsf{p}_{0.16}[\psi])^5$	0.004	0.19	**0.75**	0.98
$(\mathsf{p}_{0.05}[\psi])^5$	0.000	0.03	**0.43**	0.89

Campos, Hendry, and Krolzig (2003). This can have a cumulative effect, as shown in figure 4 which records values of the term $(1 - \mathsf{p}_\alpha[\psi])$ from table 2, where there are five independent regressors, all with the same ψ^2. The probability of retaining all five variables is low even at loose significance levels

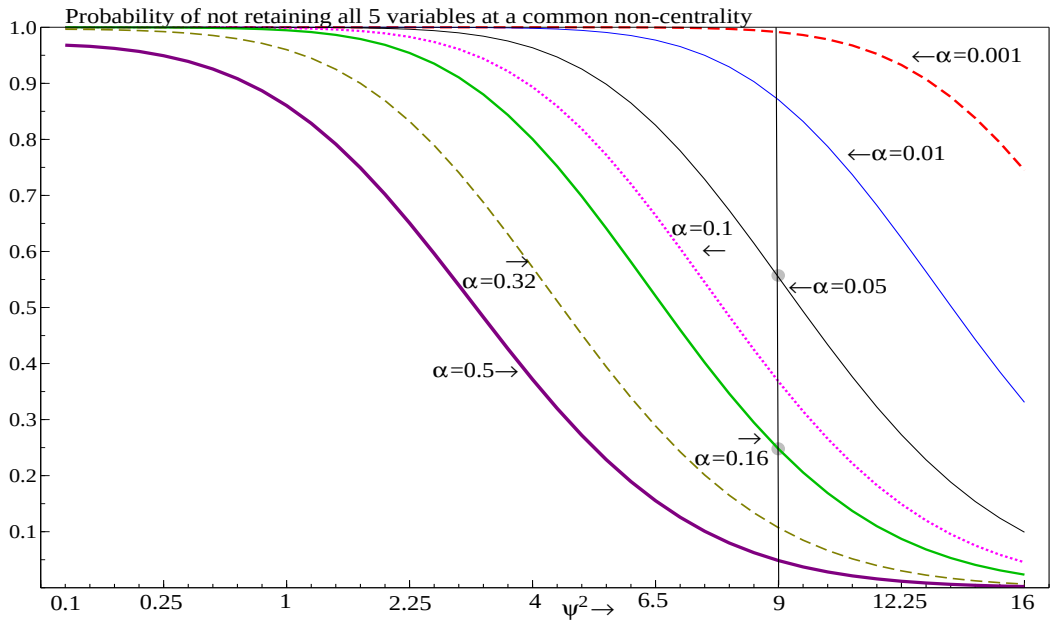

Figure 4: Values of $(1-\mathsf{p}_\alpha[\psi])$ for five independent regressors with the same non-centrality for the range of α and ψ^2.

unless the non-centralities are large, but the gap between $\alpha = 0.05$ and $\alpha = 0.16$ at $\psi^2 = 9$ is 29%, demonstrating large benefits of a looser significance level for the retention of relevant regressors. The trade-off is that more irrelevant variables will be retained, and this can be costly if those variables are subject to breaks, which we next explore.

5 An out-of-sample shift in the regressors

We now consider a mean shift in x_2 at $T+1$ with the forecast origin at T, so the shift coincides with the 1-step ahead forecast. The DGP has the same structure as (1) and (2), but with the mean shifting at time $T+1$ where (3) still holds:

$$\begin{aligned} x_{1,t} &= \mu_1 + \eta_{1,t} && t = 1,\ldots,T+1. \\ x_{2,t} &= \begin{cases} \mu_2 + \eta_{2,t} & t = 1,\ldots,T. \\ \mu_2 + \delta + \eta_{2,t} & t = T+1. \end{cases} \end{aligned} \tag{16}$$

We first evaluate the trade-off when omitting $x_{2,t}$ for known future exogenous regressors, so the break which occurs in the forecast period is modelled in the known $x_{2,T+1}$. Then, the trade-off is examined for the case where the exogenous regressors are unknown, so must be forecast based on in-sample observations. Forecasting devices based on full in-sample information and just the last in-sample observation are considered, which are the extremes of the class in Castle, Clements, and Hendry (2015), but there is no information in-sample regarding the break to help either device.

5.1 Known future values of regressors

The 1-step ahead forecasts for M_1 given (16), in which values of $\mathbf{x}_{T+1}$ are assumed to be known at T, are unbiased when the parameter estimates are unbiased. The MSFE of M_1 (see Appendix A.3 for derivations) is:

$$\mathsf{E}\left[\tilde{\epsilon}^2_{T+1|T+1}\right] = \sigma^2_\epsilon \left[1 + \frac{1}{T\left(1-\rho^2\right)}\left(\delta^2 + 2 - \rho\right)\right], \tag{17}$$

which does not depend on ψ^2. Comparison with (8) highlights the effects of the location shifts: δ^2 enters the MSFE despite the shift being 'known' given $x_{2,T+1}$, and MSFE_1 is no longer independent of ρ. (17) also reveals the additional costs of including an irrelevant regressor which shifts out-of-sample as δ^2 enters even when $\beta_2 = 0$, although it is scaled by $T\left(1-\rho^2\right)$ so larger samples mitigate its effect.

The forecast error for M_2 is $\mathsf{E}[\widetilde{\epsilon}_{T+1|T+1}] = \beta_2\delta$, so the forecasts are biased by the shift in the omitted variable. The 1-step ahead MSFE for M_2 is:

$$\mathsf{E}\left[\widetilde{\epsilon}^2_{T+1|T+1}\right] = \sigma^2_\epsilon + \beta_2^2\left(1-\rho^2+\delta^2\right) + 2T^{-1}\sigma^2_\epsilon\left(1+T^{-1}\psi^2\right). \tag{18}$$

β_2^2 enters directly so the MSFE is a function of ψ^2, unlike for M_1. Comparison with (9) reveals the role that ρ and δ^2 play, and when $\beta_2 = 0$ so M_2 is the correct model (18) collapses to (9).

Again, assuming a criterion of minimizing 1-step ahead MSFE, using (10), $\mathsf{MSFE}_2 \leq \mathsf{MSFE}_1$ requires:

$$\delta^2\left(\psi^2-1\right) + \psi^2\left(1-\rho^2\right)\left(1+2T^{-1}\right) - \rho \leq 0, \tag{19}$$

which depends on estimation uncertainty and therefore doesn't simplify neatly. However, the solution is close to 1 for reasonable values of ρ. For example, when $\rho = 0.5$, $T = 50$ and $\delta = 4$, then $\psi^2 < 0.983$ or $|\psi| < 0.991$ results in a smaller MSFE_2 compared to MSFE_1.

Figure 5 demonstrates the close approximation to a trade-off at $\psi = 1$ which holds regardless of the break. Thus, even knowing there is a shift in x_2 does not affect the choice of forecasting model between including or omitting x_2: always (never) include for $\psi^2 \geq 1$ ($\psi^2 < 1$).

5.2 The impact of selection

Following §4.2, a t−test for statistical significance will be conducted on $x_{2,t}$ in-sample and a decision to retain or exclude $x_{2,t}$ will be made at c_α for a given draw. Hence, M_3 will be a weighted average of M_1 and M_2, using (12):

$$\mathsf{MSFE}_3 = \mathsf{MSFE}_1 + (1-\mathsf{p}_\alpha[\psi])\left\{\sigma^2_\epsilon T^{-1}\left[\psi^2\left(1+\frac{\delta^2}{(1-\rho^2)}\right) - \frac{\delta^2+2-\rho}{(1-\rho^2)}\right]\right\}. \tag{20}$$

(20) is scaled by T so, as the sample size increases, the difference between $\mathsf{MSFE}_{\mathsf{M}_1}$ and $\mathsf{MSFE}_{\mathsf{M}_2}$ diminishes, as before. When $\psi^2 = 0$ the first term in (20) drops out and the benefits of selection relative to M_1 are evident as the second term must be negative. The magnitude of δ^2 affects both M_1 and M_2 but, from (20), the first δ^2 term is multiplied by ψ^2 whereas the second offsetting term is not, so if $\psi^2 > 1$ the effect of the location shift is exacerbated.

Figure 5 compares the MSFEs of M_1 from (17), M_2 from (18), and M_3 using (20) at three illustrative values of α for $T = 50$ and $\delta = 4$. The profiles of the MSFEs mirror the analytical results for the no break case. Selection outperforms the estimated DGP for $\psi^2 < 1$ despite a break, and remains close to the MSFE_1 at $\alpha = 0.16$ for $\psi^2 > 1$.

5.3 Unknown future values of regressors

Now consider that the future values of the regressors are unknown. We use two alternative devices to obtain forecasts of $x_{i,T+1}$, $i = 1, 2$, including using the in-sample mean and a random walk. The random walk can be thought of as a robust device after a location shift, where robustness refers to improved forecasting properties following a location shift. However, the random walk is biased for unanticipated location shifts so is not robust in this setting. The two devices comprise the two extremes of using either the full in-sample data ($t = 1, \ldots, T$) or the last observation ($t = T$) to produce the forecasts of the weakly exogenous regressors.

Although the link between y and the x_i stays constant, forecasts when the $x_{i,T+1}$ are unknown will fail if the shift at $T+1$ is not anticipated, inducing a shift in y_{T+1}. This will lead to forecast failure as the in-sample mean μ_y shifts to $(\mu_y + \beta_2\delta)$ at $T+1$, but would be forecast to be μ_y.

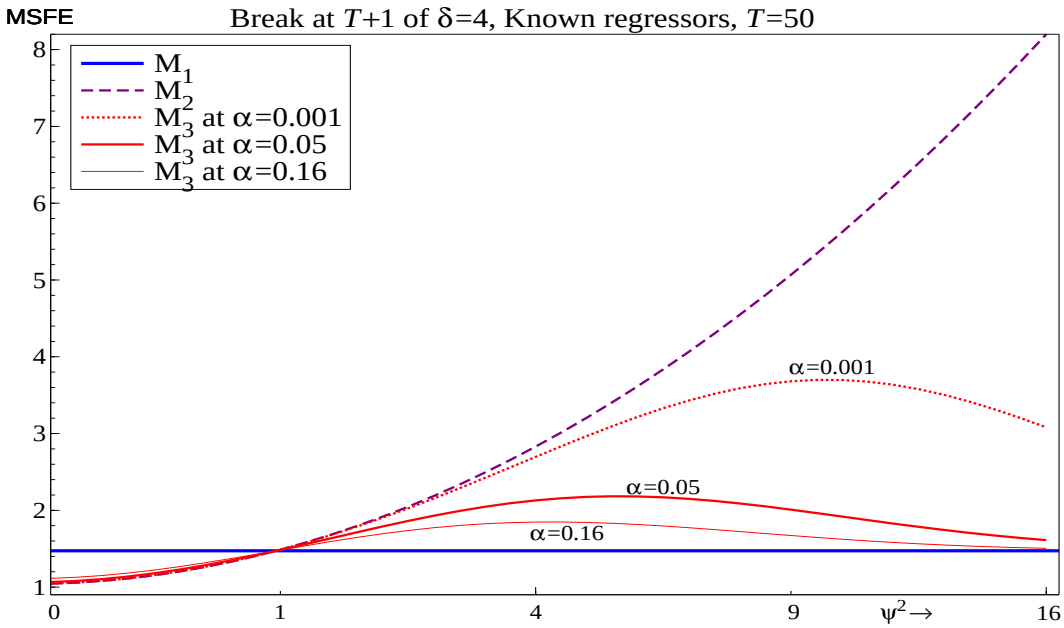

Figure 5: MSFE comparisons of M_1, M_2 and M_3 at 3 illustrative values of α for known future exogenous regressors where the break occurs in the mean of x_2 at $T+1$. $\sigma_\epsilon^2 = 1$, $\beta_0 = 5$, $\beta_1 = 1$, β_2 varies using (7) with ψ ranging from 0 to 4, $\mu_1 = \mu_2 = 2$, $\delta = 4$, $\rho = 0.5$ and $T = 50$

The forecasts based on in-sample estimates from (16) when μ_1 and μ_2 are non-zero are given by:

$$\overline{x}_{1,T+1|T} = \widehat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} x_{1,t} = \mu_1 + \overline{\eta}_1, \tag{21}$$

$$\overline{x}_{2,T+1|T} = \widehat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} x_{2,t} = \mu_2 + \overline{\eta}_2, \tag{22}$$

so will miss the unknown break. When the break occurs in x_2, the MSFEs will worsen for $\beta_2 \neq 0$. As before, we consider the sampling variation in estimating the means as small compared to the impact of shifts, so we approximate by taking T sufficiently large that $\widehat{\mu}_i \approx \mu_i$.

Replacing the unknown $x_{i,T+1}$ by μ_i leads to forecasting y_{T+1} by the in-sample mean for both M_1 and M_2, see Appendix A.4. Both face the same forecast bias, $\mathsf{E}\left[\widehat{\widetilde{\epsilon}}_{T+1|T}\right] = \mathsf{E}\left[\widetilde{\widetilde{\epsilon}}_{T+1|T}\right] = \beta_2\delta$ which is the same bias as M_2 with known regressors. Parameter estimation adds terms of $\mathsf{O}_p\left(T^{-1}\right)$. Hence, ignoring $\mathsf{O}_p\left(T^{-1}\right)$ terms, MSFE_1 is equal to MSFE_2:

$$\mathsf{E}\left[\widehat{\widetilde{\epsilon}}^2_{T+1|T}\right] = \mathsf{E}\left[\widetilde{\widetilde{\epsilon}}^2_{T+1|T}\right] = \beta_2^2\delta^2 + \sigma_\epsilon^2 + \left(\beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2\right). \tag{23}$$

When $\beta_2 = 0$ the MSFE is $\sigma_\epsilon^2 + \beta_1^2$, so is inflated relative to the known regressors case as $x_{1,T+1}$ must also be forecast. However, the in-sample mean forecast is the best forecast device for $x_{1,T+1}$ in this setting (in terms of minimum MSFE) as $x_{1,T+1}$ is stationary and not subject to a location shift. Selection will have little or no noticeable impact when $\mathsf{MSFE}_2 \approx \mathsf{MSFE}_1$, as this will also result in $\mathsf{MSFE}_3 \approx \mathsf{MSFE}_1$.

Figure 6 records the MSFEs for M_1 and M_2 when there is a break in x_2 at $T+1$, comparing known and unknown regressors using the in-sample mean to forecast $x_{i,T+1}$, $i = 1, 2$ in the unknown regressor case, i.e. the figure records (17), (18) and (23), (solid/dashed/dotted lines). Simulations outcomes are used to capture $\mathsf{O}_p\left(T^{-1}\right)$ effects but such effects are negligible so are not recorded in the figure (random walk forecasts also included).

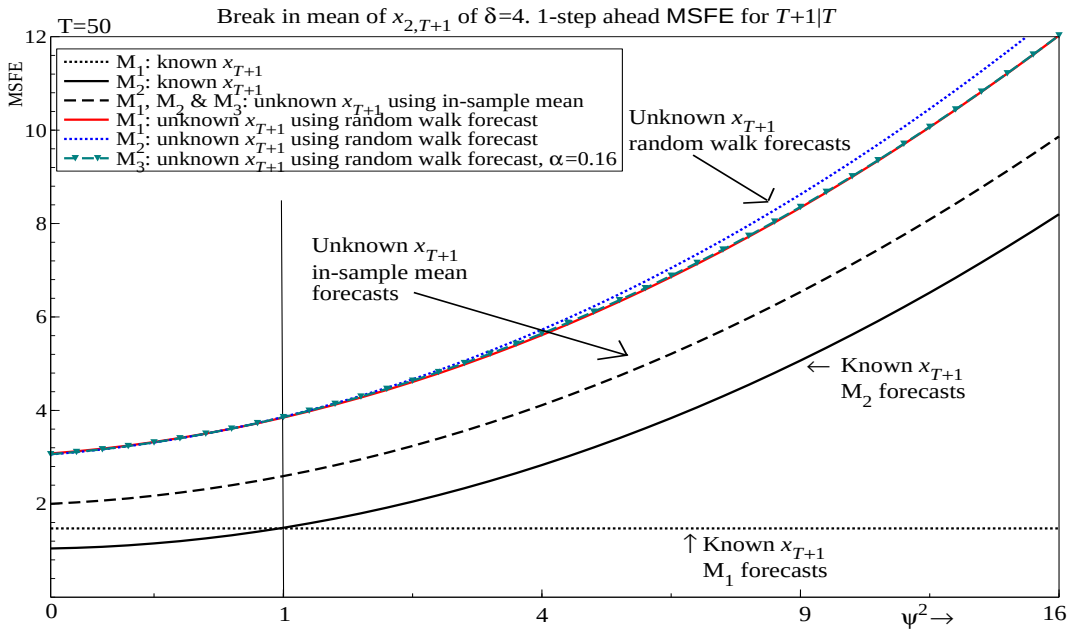

Figure 6: MSFE comparisons between M_1, M_2 and M_3 for known and unknown future exogenous regressors including in-sample mean and random walk forecasts, where the break occurs in the mean of x_2 at $T+1$. $\sigma_\epsilon^2 = 1$, $\beta_0 = 5$, $\beta_1 = 1$, β_2 varies using (7) with ψ ranging from 0 to 4, $\mu_1 = \mu_2 = 2$, $\delta = 4$ and $\rho = 0.5$. Simulations with $M = 100,000$ replications, $T = 50$.

For known regressors for MSFE_1, the break in μ_2 does not affect the MSFE as it is captured in $x_{2,T+1}$: even at $\delta = 4$ for $T = 100$, $\mathsf{MSFE}_1 = 1.23$ for the parameters given in the figure which is only slightly greater than σ_ϵ^2. However, when $\mathbf{x}_{T+1}$ is unknown both M_1 and M_2 are affected by the break in $x_{2,T+1}$. Simulation outcomes again closely match the theory for the unknown break case, and show that the choice of whether to retain or exclude $x_{2,t}$ is not important in a forecasting context. The unanticipated break dominates any forecast error resulting from model mis-specification. Increasing the sample size does mitigate the MSFE costs but the increase in MSFE relative to known regressors is maintained for all ψ^2.

These results show that in this static setting of location shifts, if the break occurs in the forecast period and is unknown and unpredicted, then the retention of x_2 is irrelevant (other than parameter estimation uncertainty), as neither M_1 nor M_2 capture the shift which dominates the MSFE. **Parsimony, or lack thereof, neither helps nor hinders much in this setting**. Moreover, selection does not substantively affect the outcome as $\mathsf{MSFE}_3 \approx \mathsf{MSFE}_1$.

5.4 Random Walk forecast for a post-origin break

We now consider using a random walk as the forecasting device for the exogenous variables, given by:

$$\overline{\overline{x}}_{1,T+1|T} = x_{1,T}, \tag{24}$$

$$\overline{\overline{x}}_{2,T+1|T} = x_{2,T}. \tag{25}$$

Although the last in-sample observation is an imprecise measure of the out-of-sample mean, it is unbiased when there are no location shifts (as there are no dynamics in the DGP), so $\mathsf{E}[x_{1,T}] = \mu_1$ and $\mathsf{E}[x_{2,T}] = \mu_2$, and hence $\mathsf{E}[\Delta x_{1,T+1}] = 0$ and $\mathsf{E}[\Delta x_{2,T+1}] = \delta$.

The forecasts from M_1 will be biased by the bias in the random walk forecast of $x_{2,T+1}$, so (see appendix A.5 for derivations) neglecting the small impact of $\eta_{i,T}$ on $\beta_i - \widehat{\beta}_i$:

$$\mathsf{E}\left[\overline{\overline{\epsilon}}_{T+1|T}\right] = \beta_2 \delta,$$

and the resulting MSFE_1 is:

$$\mathsf{E}\left[\bar{\bar{\epsilon}}^2_{T+1|T}\right] = \beta_2^2\delta^2 + 2\left(\beta_1^2+\beta_2^2\right) + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1+2T^{-1}\right). \tag{26}$$

Comparison with (23) highlights the additional cost of using the random walk relative to the in-sample mean when neither forecasting device can predict the break, since:

$$\mathsf{E}[\widehat{\bar{\epsilon}}^2_{T+1|T}] - \mathsf{E}\left[\bar{\bar{\epsilon}}^2_{T+1|T}\right] = -\left(\beta_1^2+\beta_2^2+2\rho\beta_1\beta_2+2\sigma_\epsilon^2T^{-1}\right).$$

The in-sample mean of x_1 is the optimal forecast of $x_{1,T+1}$ given its in-sample stationarity, so irrespective of the value of β_2, the in-sample mean forecasts dominate when the shift is during the forecast period. When $\beta_2 = 0$, (26) collapses to $\approx \sigma_\epsilon^2 + 2\beta_1^2$, ignoring $\mathsf{O}_p\left(T^{-1}\right)$ terms, compared to $\sigma_\epsilon^2+\beta_1^2$ for the in-sample mean forecasts. A random walk doubles the error variance for the variable being forecast so can be costly if there are no breaks or if the break occurs after the forecast origin. As for the in-sample mean case, the MSFE of M_1 is a function of the break magnitude.

The forecast bias for M_2 is the same as that for M_1 by the same argument, although the MSFE_2 (reported in Appendix A.5) does deviate from that for M_1 as ψ^2 increases. This is due to the correlation parameter ρ which is picking up part of the omitted variable $x_{2,T+1}$ in M_2 and has more effect as ψ^2 increases. When $\beta_2 = 0$, $\mathsf{MSFE}_2 \approx \sigma_\epsilon^2 + 2\beta_1^2$, which is the same as M_1. Despite small but increasing deviations as ψ^2 increases, MSFE_2 follows a similar trajectory to MSFE_1 so the mis-specification is less relevant for the random walk forecasts of the marginal processes relative to the effect of the break, similar to the results for the in-sample mean forecasts.

5.5 The impact of selection

In practice, selection will be applied to determine whether to include $x_{2,t}$ or not, so from (12) we can obtain the MSFE_3 as:

$$\mathsf{MSFE}_{\mathsf{M}_3} = \mathsf{MSFE}_1 + (1-\mathsf{p}_\alpha[\psi])\left\{\sigma_\epsilon^2T^{-1}\left[\psi^2\left(\frac{(1+\rho^2)}{(1-\rho^2)}+T^{-1}\right)+1\right]\right\}.$$

The trade-off between parameter estimation uncertainty and including x_2 is essentially the same as in the known variable case: if x_2 has a non-centrality of zero, so $\beta_2 = \psi^2 = 0$, then the 1-step MSFE is minimized by excluding x_2 from the forecasting model. It should be included if $\psi^2 > 1$. However, depending on the values of ρ and T, the switch point can be smaller than $\psi^2 = 1$, although the impact is likely to be small given the scale factor $\sigma_\epsilon^2T^{-1}$. Even though the random walk forecast is highly uncertain by using just one observation, if the variable that breaks is quite significant then it pays to include that variable when using the random walk forecast.

Figure 6 also records the MSFEs for the random walk forecast using the same parameter values. The increase in MSFE over the in-sample mean forecasts is evident. Both MSFE_1 and MSFE_2 follow similar trajectories, although they do start to diverge for large ψ^2, with MSFE_3 at $\alpha = 0.16$ close to MSFE_1.

6 An in-sample shift in the regressors

The break is now assumed to occur at T, so there is information available regarding the break from the last in-sample observation. The DGP is adapted from (16) but the shift in μ_2 occurs at T, rather than $T+1$:

$$\begin{aligned} x_{1,t} &= \mu_1 + \eta_{1,t} && t = 1,\ldots,T+1. \\ x_{2,t} &= \begin{cases} \mu_2 + \eta_{2,t} & t = 1,\ldots,T-1. \\ \mu_2 + \delta + \eta_{2,t} & t = T, T+1. \end{cases} \end{aligned} \tag{27}$$

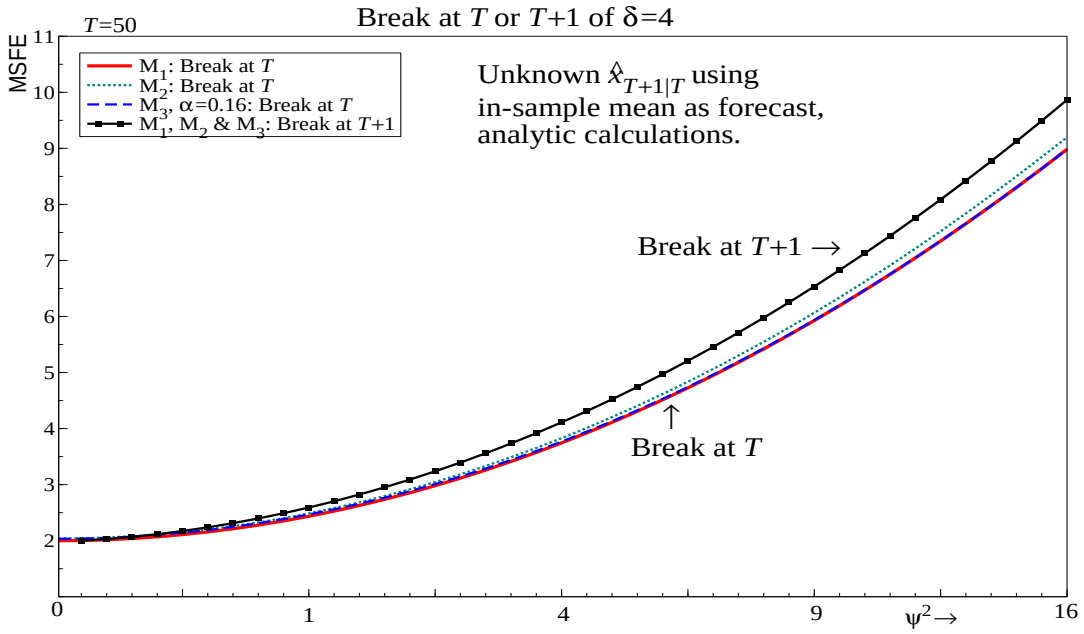

Figure 7: MSFE comparisons between M_1, M_2 and M_3 for unknown future exogenous regressors where the break occurs in the mean of x_2 at T and the in-sample mean is used as the forecast for the conditioning regressors, recorded with the same models when the break occurs at $T+1$, where M_1, M_2 and M_3 coincide in the analytic calculations. $T = 50$, using the same parameter values as previous figures.

6.1 Forecasts using in-sample means

The relationship of interest, the conditional equation for y_{T+1}, remains constant but the in-sample mean μ_y is shifted to $(\mu_y + \beta_2\delta)$ at T. Although the only DGP parameter to shift is μ_2 to $\mu_2 + \delta$, sample calculations will be altered as now $\mathsf{E}[x_2] = \mu_2 + T^{-1}\delta$, see Appendix A.6 for derivations.

The impact on the estimated in-sample mean of $\{x_{2,t}\}$ will be small from the break, unless δ is very large, so by using the in-sample means for their future unknown values, the forecasted mean of y_{T+1} for M_1 will still be close to μ_y, and the resulting forecast error bias is:

$$\mathsf{E}\left[\widehat{\widehat{\epsilon}}_{T+1|T+1}\right] \approx \beta_2\delta\left(1 - T^{-1}\right).$$

This is unbiased when $\beta_2 = 0$, but could be badly biased if $\beta_2\delta$ is large. The MSFE for M_1 is:

$$\mathsf{E}\left[\widehat{\widehat{\epsilon}}^2_{T+1|T+1}\right] = \beta_2^2\delta^2\left(1 - T^{-1}\right)^2 + \beta_1^2 + \beta_2^2 + \sigma_\epsilon^2. \tag{28}$$

This is very similar to the MSFE_1 in (23) for an out-of-sample break using the in-sample means to forecast the exogenous regressors, and hence MSFE_2 and MSFE_3 as well, although the correlation between the two regressors does not enter.

When $\beta_2 = 0$, both MSFEs collapse to $\sigma_\epsilon^2 + \beta_1^2$, but the dampening of the squared location shift by $\left(1 - T^{-1}\right)^2$ slightly improves the MSFE for the in-sample shift relative to an out-of-sample shift at larger ψ^2, as shown in figure 7.

For a break out-of-sample we found the analytic results for M_2 are identical to those for M_1 (see §5.3). For the in-sample break, the forecast error and MSFE for M_2 does differ to that of M_1 (see Appendix A.6 for analytic results). This is because the in-sample location shift affects ρ which introduces a term similar to the squared location shift scaled by T in (28). Therefore, $\mathsf{MSFE}_1 \neq \mathsf{MSFE}_2$ unless $\beta_2 = 0$, with M_2 incurring a larger MSFE cost as ψ^2 increases due to misspecification, although the divergence is small even for small T, and disappears asymptotically.

6.2 Selecting regressors

Selection follows from (12) and hence:

$$\mathsf{MSFE}_3 \approx \mathsf{MSFE}_1 + (1 - \mathsf{p}_\alpha[\psi]) \left[\sigma_\epsilon^2 - \beta_1^2 - \rho^2\beta_2^2 + 2T^{-1}\left(\sigma_\nu^2 + \beta_2^2\delta^2\right)\right]. \tag{29}$$

The cost of omitting x_2 rises with $\beta_2^2\delta^2$, although increases in β_2 will raise ψ^2 and hence raise the probability of retaining x_2, albeit unconnected with the magnitude of δ^2. As the location shift is scaled by T, $\mathsf{MSFE}_3 \to \mathsf{MSFE}_1$ as $T \to \infty$, as can be seen in Figure 7.

6.3 Forecasts using the random walk

From the previous analysis in §6.1, knowledge of the break, where the break occurs at T, brought little benefit. However, the random walk should do better when the break occurs at T as opposed to $T+1$ as it is now a robust forecasting device. As before:

$$\widetilde{\widetilde{x}}_{1,T+1|T} = x_{1,T} \quad \text{and } \widetilde{\widetilde{x}}_{2,T+1|T} = x_{2,T},$$

but now $\mathsf{E}[x_{1,T}] = \mu_1$ and $\mathsf{E}[x_{2,T}] = \mu_2 + \delta$, and hence $\mathsf{E}[\Delta x_{1,T+1}] = 0$ and $\mathsf{E}[\Delta x_{2,T+1}] = 0$ as well.

Given the unbiased forecasts of the exogenous regressors, it follows that the forecasts for M_1 are unbiased (see Appendix A.7) when the parameter estimates are unbiased. The MSFE for M_1 is:

$$\mathsf{E}\left[\widetilde{\widetilde{\epsilon}}_{T+1|T}^2\right] = 2\left(\beta_1^2 + \beta_2^2\right) + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1 + \frac{2}{T} + \frac{\delta^2}{T\left(1-\rho^2\right)}\right). \tag{30}$$

When $\beta_2 = 0$, the MSFE is similar to that of the out-of-sample break case, where the random walk is costly as forecasts of both $x_{1,T+1}$ and $x_{2,T+1}$ are inefficient. However, (30) does depend on the magnitude of the shift independently of β_2, unlike (26). MSFE_1 is a function of ψ^2, increasing as ψ^2 increases unlike in the known regressor case, but it does so more slowly than for breaks out-of-sample, or breaks in-sample using the in-sample mean. As ψ^2 increases, the break at T in μ_2 has a larger effect on the dependent variable, and hence the benefits of using a random walk forecast of $x_{2,T+1}$ are larger.

M_2 will suffer when $\beta_2 \neq 0$ as the forecasts will be biased. The MSFE for M_2 is:

$$\mathsf{E}\left[\widetilde{\widetilde{\epsilon}}_{T+1|T}^2\right] = \beta_2^2\left(\delta^2 + \rho^2 + 1\right) + 2\beta_1^2 + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1 + T^{-1} + T^{-2}\psi^2\right). \tag{31}$$

so no robustness to the break is achieved unless $\beta_2 = 0$. When $\beta_2 = 0$, $\mathsf{MSFE}_2 < \mathsf{MSFE}_1$, but the bias from not including a robust, and hence unbiased, forecast of $x_{2,T+1}$ quickly outweighs parameter estimation costs at ψ^2 increases.

Solving for $\mathsf{MSFE}_2 < \mathsf{MSFE}_1$ results in:

$$\psi^2 < \frac{\left(1-\rho^2\right) + \delta^2}{\left(1-\rho^2\right)\left(T^{-1}-1\right) + \delta^2}. \tag{32}$$

The break term dominates and enters on the numerator and denominator, leading to a trade-off at ≈ 1 with deviations scaled by T^{-1}. For $\rho = 0.5$, $T = 100$ and $\delta = 4$, MSFE_2 dominates when $\psi = 1.05$. Interestingly, the cut-off is slightly above 1 for this case, compared to slightly below 1 for the known breaks out-of-sample case, but the results still imply that a selection significance level of approximately 16% would be optimal to trade-off the cost of estimating an additional parameter.

Figure 8 records the MSFEs from M_1 (30), M_2 (31) and three values of M_3 (73) for the analytic results. There is a clear trade-off at $\psi^2 \approx 1$, just as in the known breaks case.

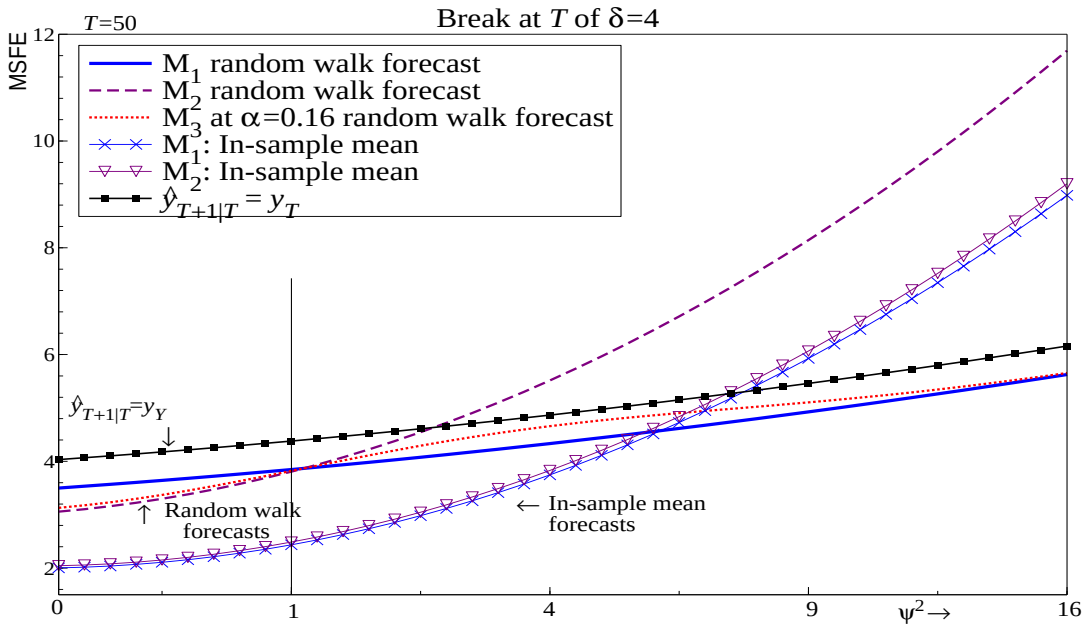

Figure 8: MSFE comparisons between M_1, M_2 and M_3 at $\alpha = 0.16$ for unknown future exogenous regressors where the break occurs in the mean of x_2 at T and the last in-sample observation is used as the forecast for the conditioning regressors. Also recorded is the MSFE for M_1 and M_2 using in-sample means and a misspecified random walk for y_{T+1} directly. $T = 50$.

6.4 Selecting regressors

The MSFE for M_3 for the random walk is reported in Appendix A.7. As with Figure 5, selection between M_1 and M_2 can be advantageous even for the random walk as seen in Figure 8. Selection outperforms M_1 for $\psi^2 < 1$, and remains close to the MSFE_1 at $\alpha = 0.05$ and $\alpha = 0.16$, again in all cases matching or outperforming always using M_2.

A comparison with the MSFE for the in-sample mean forecasts, also recorded in figure 8, suggest a possible forecast improvement. If the regressor that breaks at T is known, combining the in-sample mean forecast for M_1 with the random walk forecast for M_2 will improve forecast performance (shifting the MSFE curves for the random walk forecast down by approximately 1). As the number of regressors increase, the forecasting method for each contemporaneous regressor will have a cumulative impact. However, as the break occurs in-sample, methods to detect breaks at the forecast origin such as impulse indicator saturation (IIS) could be used to guide the forecaster to the most appropriate forecasting device.[2] Selection between non-robust and robust forecasting devices requires pre-testing and would only help for in-sample shifts, see, e.g., Chu, Stinchcombe, and White (1996).

Thus, selection can be valuable for forecasting to the extent that it retains relevant regressors that shift (here, x_2), and also if it eliminates irrelevant regressors that shift, as considered in Section 8.

6.5 A misspecified random walk forecasting device

If a break is suspected, an alternative to selecting a conditional model that aims to approximate the DGP is to use a knowingly misspecified model of the conditional DGP. A random walk forecast could be obtained directly for y, with the advantage that y_T is known and avoids the need to forecast $x_{i,T+1}$ $i = 1, 2$. Hendry and Mizon (2012) derive a forecast error taxonomy for open models that demonstrates

[2]See Hendry, Johansen, and Santos (2008), Johansen and Nielsen (2009), Johansen and Nielsen (2016) and Hendry and Doornik (2014) for details of IIS. Castle, Doornik, and Hendry (2012) demonstrate the ability of IIS to detect breaks in the form of location shift at any point in the sample.

the numerous additional forecast errors that arise from forecasting regressors offline in open models and show that in some cases it can pay to use a misspecified model rather than to forecast the regressors offline. The forecast device is:

$$\widetilde{\widetilde{y}}_{T+1|T} = y_T, \tag{33}$$

where

$$y_T = \mu_y + \beta_2\delta + \beta_1\eta_{1,T} + \beta_2\eta_{2,T} + \epsilon_T,$$

is a noisy 1-observation estimator of $\left(\mu_y + \beta_2\delta\right)$, The outturn at $T+1$ is:

$$y_{T+1} = \left(\mu_y + \beta_2\delta\right) + \beta_1\Delta\eta_{1,T+1} + \beta_2\Delta\eta_{2,T+1} + \epsilon_{T+1} + \beta_1\eta_{1,T} + \beta_2\eta_{2,T},$$

and so the forecast error is given by:

$$\begin{aligned} \widetilde{\widetilde{\epsilon}}_{T+1|T} &= y_{T+1} - \widetilde{\widetilde{y}}_{T+1|T} \\ &= \beta_1\Delta\eta_{1,T+1} + \beta_2\Delta\eta_{2,T+1} + \Delta\epsilon_{T+1}, \end{aligned} \tag{34}$$

which is unbiased and has a MSFE of:

$$\mathsf{E}\left[\widetilde{\widetilde{\epsilon}}^2_{T+1|T}\right] = 2\left(\beta_1^2 + \beta_2^2\right) + 4\rho\beta_1\beta_2 + 2\sigma_\epsilon^2. \tag{35}$$

This is independent of δ so should perform relatively the best when δ^2 is large, although performs worse than random walk forecasts for the $x_{i,T+1}$ $i = 1, 2$ when ψ^2 is small, see Figure 8. The forecasts are invariant to omitting x_2 since the random walk is independent of the regressors, which is a major advantage and negates the role of selection. However, there is a cost when the model is correctly specified. The results in the simulation below suggest that such an approach should be viewed as complementary, with forecast pooling across selected conditional models and misspecified robust devices frequently outperforming individual methods.

7 Overview of analytic results and the impact of selection

The analytic results in §3-6 have established that:

- Regressors should be retained if $\psi \gtrapprox 1$. This is established for stationary DGPs and DGPs with a break out-of-sample for known regressors and a break in-sample using random walk forecasts.
- For the two regressor case, $\psi = 1$ maps to $\alpha \approx 0.16$. Selection delivers improvements to the 1-step ahead MSFE for $\psi < 1$ and can be close to the correct model specification for $\psi > 1$, with the largest deviation occuring at intermediate values of ψ.
- If there are breaks out-of-sample and contemporaneous regressors need to be forecast, the break dominates the MSFE and selection plays almost no role. Similar results are found even if the break occurs at the end of the sample, but a non-robust in-sample mean is used to forecast to regressors.
- Random walk forecasts are costly if there are no breaks (forecasting $x_{1,T+1}$) or if the breaks are unpredictable (a break at $T+1$ and forecasting $T+1|T$). However, they improve MSFE when the break is predictable (break at T and forecasting $T+1|T$).

Table 3 summarises the full set of analytic results for specific parameters for $T = 50$, with results for $T = 100$ reported in table 9 in the appendix. For each scenario, the ratio of $\mathsf{MSFE}_j/\mathsf{MSFE}_1$ for $j = 2, 3$ is reported, for three values of α ($\alpha = 0.001; 0.05; 0.16$). Benchmarks of $\psi^2 = 0, 1, 4, 9$ and 16 are reported, capturing the full hump shape seen in the figures above. The parameters are $\sigma_\epsilon^2 = 1$, $\beta_0 = 5$, $\beta_1 = 1$, $\mu_1 = \mu_2 = 2$, $\delta = 4$ and $\rho = 0.5$.

Looking down the column labelled $\psi^2 = 1$ highlights the $\psi = 1$ trade-off, with all cases almost exactly equal to 1. (19) found a cut-off slightly lower than 1, which is reflected in the ratio marginally greater than 1, and conversely, (32) found a cut-off slightly larger than 1, resulting in a ratio slightly below 1, but the differences are small.

Moving to the column labelled $\psi^2 = 0$, here M_2 is the correct model, so the ratio of $\mathsf{MSFE}_2/\mathsf{MSFE}_1$ measures the cost of over-specification. The gains can be substantial in some cases, almost 30% for a break out-of-sample with known regressors, but in other cases including $x_{2,t}$ is not at all costly despite its irrelevance. Tighter selection for M_3 is close to M_2 as $x_{2,t}$ will be omitted more frequently, but even at $\alpha = 0.16$ the ratio for M_3 is close to the ratio for M_2, suggesting that selection is not costly.

Next, consider the columns labelled $\psi^2 = 4, 9$, and 16. M_1 is the correct model so the objective is to minimize the ratio. In some cases M_2 performs poorly, but M_3 at $\alpha = 0.16$ is frequently very close to 1, i.e. MSFE_1. Selection forecast performance tends to be worse at $\psi^2 = 4$, but as the signal for x_2 increases the probability of retaining x_2 increases so the selected model is closer to M_1. The benefits of selection vary by case. For example, for a break at T using in-sample means, selection at $\alpha = 0.16$ delivers a 2.4% improvement relative to M_2 for $\psi = 4$, compared to a halving of the ratio for the random walk. In almost every setting, MSFE_3 is close to MSFE_1 so the costs of selection are usually small, irrespective of the non-centrality. In that sense, model selection acts to reduce the risk relative to the worst model. Conversely, the costs of unmodelled shifts are very large, up to almost 8-fold greater than the baseline stationary MSFE_1.

These results show that even facing breaks, the well-known trade-off for selecting variables in forecasting models, namely that variables should be retained if their non-centralities exceed 1, still applies, resulting in much looser significance levels than typically used. The problem with such an approach is that when many $\beta_{2,i} = 0$ but are subject to location shifts, M_1, which erroneously includes $x_{2,t}$ in the model, will perform worse. Loose significance levels increase the chance that irrelevant variables with $\psi = 0$ are retained by being adventitiously significant for that draw. To evaluate this effect, the next section undertakes a simulation study of selection in models with more irrelevant (10) than relevant (5) exogenous regressor variables confronting a variety of shifts.

8 Simulation evidence

We generalize the above analysis to consider larger models with dynamics, evaluating for a range of different significance levels using Monte Carlo analysis. Single $\mathtt{t}-$tests are no longer appropriate as there are many potential regressors to select and correlations between potential regressors are non-zero. The selection algorithm *Autometrics*, see Doornik (2009), is used which is a general-to-specific tree search algorithm that searches feasible reduction paths to allow for collinearity.

8.1 Simulation Design

The data generating process (DGP) is given by:

$$y_t = \beta_0 + \beta_y y_{t-1} + \boldsymbol{\beta}' \mathbf{x}_t + \epsilon_t, \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right], \tag{36}$$

where:

$$\underset{(N\times 1)}{\mathbf{x}_t} = \begin{cases} \boldsymbol{\iota} + \boldsymbol{\lambda}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t & \text{for } t = 1, \ldots, T \\ (\boldsymbol{\iota} + \boldsymbol{\nu}\nabla\iota) + (\boldsymbol{\lambda} + \boldsymbol{\nu}\nabla\lambda)\,\mathbf{x}_{t-1} + \boldsymbol{\eta}_t & \text{for } t = T+1, T+2 \end{cases} \tag{37}$$

where $\boldsymbol{\eta}_t \sim \mathsf{IN}_N\left[\mathbf{0}, \mathbf{I}\right]$, and where $\boldsymbol{\iota}$ is a column vector of ones and $\boldsymbol{\nu}$ is an $(N \times 1)$ vector with elements taking the value 0 or 1 to reflect which elements of $\mathbf{x}_t$ experience a shift (either relevant, irrelevant, or all regressors). $\nabla\iota$ is a (1×1) vector giving the intercept shift magnitude. It is set to give a 4 standard deviation mean shift in $\mathbf{x}_t$ at $T+1$, so $\mathsf{E}\left[\mathbf{x}^*_{T+h}\right] = \mathsf{E}\left[\mathbf{x}_{T-h}\right] + 4\mathsf{V}\left[\mathbf{x}_{T-h}\right]^{1/2}$ when $h > 0$, so:

$$\mathsf{E}\left[\mathbf{x}^*_{T+h}\right] = (1-\lambda)^{-1}\boldsymbol{\iota} + 4\left[\sigma_\eta^2\left(1-\lambda^2\right)^{-1}\right]^{\frac{1}{2}}\boldsymbol{\iota} = 6.62\boldsymbol{\iota},$$

Table 3: $T = 50$, Ratio M_2 reports $\frac{\mathsf{MSFE}_2}{\mathsf{MSFE}_1}$ and Ratio M_3 reports $\frac{\mathsf{MSFE}_3}{\mathsf{MSFE}_1}$. $\sigma_\epsilon^2 = 1$, $\beta_0 = 5$, $\beta_1 = 1$, $\mu_1 = \mu_2 = 2$, $\delta = 4$ and $\rho = 0.5$

Case		Model	$\psi^2 = 0$	$\psi^2 = 1$	$\psi^2 = 4$	$\psi^2 = 9$	$\psi^2 = 16$
Stationary	Ratio M_2		0.981	1.001	1.060	1.158	1.295
($\delta^2 = 0$)	Ratio M_3	$\alpha = 0.001$	0.981	1.000	1.051	1.093	1.068
		$\alpha = 0.05$	0.982	1.000	1.027	1.023	1.006
		$\alpha = 0.16$	0.984	1.000	1.016	1.008	1.001
out-of-sample shift	Ratio M_2		0.709	1.014	1.927	3.450	5.582
known future	Ratio M_3	$\alpha = 0.001$	0.709	1.013	1.836	2.505	2.095
regressors		$\alpha = 0.05$	0.724	1.011	1.449	1.366	1.095
		$\alpha = 0.16$	0.756	1.009	1.256	1.136	1.022
out-of-sample shift	Ratio M_2		1.000	1.000	1.000	1.000	1.000
unknown future	Ratio M_3	$\alpha = 0.001$	1.000	1.000	1.000	1.000	1.000
regressors		$\alpha = 0.05$	1.000	1.000	1.000	1.000	1.000
mean forecast		$\alpha = 0.16$	1.000	1.000	1.000	1.000	1.000
out-of-sample shift	Ratio M_2		0.993	1.004	1.020	1.034	1.043
unknown future	Ratio M_3	$\alpha = 0.001$	0.993	1.004	1.018	1.021	1.010
regressors		$\alpha = 0.05$	0.994	1.003	1.010	1.005	1.001
random walk forecast		$\alpha = 0.16$	0.994	1.002	1.006	1.002	1.000
in-sample shift	Ratio M_2		1.020	1.021	1.022	1.023	1.024
unknown future	Ratio M_3	$\alpha = 0.001$	1.020	1.021	1.020	1.014	1.006
regressors		$\alpha = 0.05$	1.019	1.017	1.011	1.004	1.000
mean forecast		$\alpha = 0.16$	1.017	1.014	1.006	1.001	1.000
in-sample shift	Ratio M_2		0.871	0.990	1.273	1.653	2.078
unknown future	Ratio M_3	$\alpha = 0.001$	0.871	0.990	1.246	1.401	1.258
regressors		$\alpha = 0.05$	0.878	0.991	1.132	1.097	1.022
random walk forecast		$\alpha = 0.16$	0.892	0.993	1.075	1.036	1.005

resulting in $\boldsymbol{\iota}^* = \boldsymbol{\iota} + \nabla\iota = 3.31\boldsymbol{\iota}$ and $\nabla\iota = 2.31$, for those elements that shift.

$\boldsymbol{\lambda} = 0.5\mathbf{I}_N$ is an $(N \times N)$ matrix giving the degree of persistence in the exogenous regressors, with zeros in the off-diagonals so the regressors are uncorrelated in the population, but lags of the regressors are correlated. $\nabla\lambda = 0.45$ is a (1×1) vector giving the autoregressive parameter shift magnitude, and so the degree of persistence increases from 0.5 to 0.95 when the elements of the $\boldsymbol{\nu}$ vector are 1. Shifts in $\boldsymbol{\iota}$ and $\boldsymbol{\lambda}$ are considered separately.

$\sigma_\epsilon^2 = 1$, $\beta_0 = 5$, $\beta_y = 0.5$ and $\boldsymbol{\beta} = \left(\sigma_\epsilon^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{1/2} \boldsymbol{\psi}$, where three alternative experiments are considered, with $N = 15$ and n as the number of relevant variables, for:

$$\underset{(N\times 1)}{\boldsymbol{\psi}} = \begin{cases} (0,0,0,0,0,0,0,0.5,1,1.5,2,2.5,3,3.5,4)'. \\ (0,0,0,0,0,0,0,0,0,0,0,4,4,4,4,4)'. \\ (0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1)'. \end{cases} \tag{38}$$

allowing for some irrelevant variables, some marginally relevant variables and some strongly significant variables.[3]

The in-sample period is $T = 100$, with a sequence of two 1-step ahead forecasts undertaken for $y_{T+1|T}$ and $y_{T+2|T+1}$. As the break occurs at $T + 1$, there is no information on the break for the first forecast, but the second forecast conditions on information at $T + 1$. $M = 1,000$ replications.

The four cases examined are:

(a) No breaks: $\boldsymbol{\nu} = \mathbf{0}_N$.

[3]Results for $N = 10$, with 5 fewer irrelevant variables, for $\rho = 0$, and for $\beta_y = 0$ are similar and are available on request.

(b) Breaks in relevant variables: $\boldsymbol{\nu} = \left(\{\mathbf{0}\}_{N-n} : \{\mathbf{1}\}_n\right)$.
(c) Breaks in irrelevant variables: $\boldsymbol{\nu} = \left(\{\mathbf{1}\}_{N-n} : \{\mathbf{0}\}_n\right)$.
(d) Breaks in all variables: $\boldsymbol{\nu} = \mathbf{1}_N$.
Selection is undertaken from the general unrestricted model (GUM):

$$y_t = \overline{\beta}_0 + \beta_y y_{t-1} + \sum_{i=0}^{1}\sum_{j=1}^{N} \beta_{ij} x_{j,t-i} + \epsilon_t, \quad \text{for} \quad t = h+1, \ldots, T+h-1, \tag{39}$$

for $h = 1, 2$, where $\overline{\cdot}$ denotes the intercept is not selected over so is always retained. A new model is selected for each forecast horizon ($h = 1, 2$) using a rolling window. Selection is applied using *Autometrics* for a range of target significance levels $\alpha = (0.001, 0.01, 0.05, 0.1, 0.16, 0.32, 0.5)$, resulting in a selected model for each replication m and significance level α_k:

$$\widehat{y}_t = \widehat{\beta}_0 + \widehat{\boldsymbol{\beta}}_s' \mathbf{w}_t, \tag{40}$$

where $\mathbf{w}_t$ stacks $(y_{t-1} : \mathbf{x}_t : \mathbf{x}_{t-1})$ and $\widehat{\boldsymbol{\beta}}_s$ has zeros for regressors not retained in selection.

In the experiments, we compute the 1-step ahead MSFEs, given by $\frac{1}{M}\sum_{i=1}^{M} \epsilon^2_{T+h,i|T+h-1,\alpha_k}$, for a forecast horizon ($h = 1, 2$) for various forecast models under the assumption of:

(i) known future exogenous regressors, with forecast error $\overline{\epsilon}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \mathbf{w}_{T+h}$;

(ii) unknown future exogenous regressors using the in-sample mean as the forecast for the retained exogenous regressors, resulting in the forecast error $\overline{\epsilon}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \overline{\mathbf{w}}_{T+h|T+h-1}$.

The lagged retained variables will be known for one-step ahead forecasts, so $\overline{\mathbf{w}}_{T+h|T+h-1}$ consists of the in-sample averages for $\mathbf{x}_{T+h}$ but known y_{T+h-1} and $\mathbf{x}_{T+h-1}$, i.e. $\overline{\mathbf{w}}_{T+h|T+h-1} = \left(y_{T+h-1} : \frac{1}{T}\sum_{t=2}^{T+h-1} \mathbf{x}_t : \mathbf{x}_{T+h-1}\right)$.

The in-sample means for forecasts at $T+1$ do not include the break, but for forecasts at $T+2$ will include the break observation, although this will have a small effect on the mean given the sample size unless the break is extremely large.

(iii) unknown future exogenous regressors, forecasting each variable selected from the GUM by:

$$x_{j,t} = \delta_0 + \delta_j x_{j,t-1} + \sum_{l=1}^{N}\sum_{i=0}^{1} \delta_{l,i} x_{l,t-i} + u_{j,t} \quad u_{j,t} \sim \mathsf{IN}\left[0, \sigma_u^2\right] \tag{41}$$

for $j = 1, \ldots, N$, $l \neq j$ and $t = 2, \ldots, T$, or $t = 3, \ldots, T+1$ depending on the forecast origin, where selection is undertaken at the same α_k as that for (40). The resulting forecast error is $\widehat{\widehat{\epsilon}}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \widehat{\mathbf{w}}_{T+h|T+h-1}$.

One step ahead forecasts are computed for $x_{j,T+h|T+h-1}$ but known values are used for any retained lags: $\widehat{\mathbf{w}}_{T+h|T+h-1} = \left(y_{T+h-1} : \widehat{\mathbf{x}}_{T+h|T+h-1} : \mathbf{x}_{T+h-1}\right)$.

(iv) unknown future exogenous regressors using a robust forecast for the exogenous regressors. Two alternative robust forecasts are evaluated:

(a) random walk, using the last in-sample observation as the forecast for the retained current-dated exogenous regressors:
$\widetilde{\epsilon}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \widetilde{\mathbf{w}}_{T+h-1}$, where $\widetilde{\mathbf{w}}_{T+h|T+h-1} = (y_{T+h-1} : \mathbf{x}_{T+h-1} : \mathbf{x}_{T+h-1})$.

(b) random walk with difference, using the robust device from Hendry (2006) as the forecast for the retained exogenous regressors. For each exogenous regressor, estimate a first-order autoregression (AR(1)):

$$x_{j,t} = \delta_{j,0} + \delta_{j,1} x_{j,t-1} + u_{j,t} \quad u_{j,t} \sim \mathsf{IN}\left[0, \sigma_u^2\right], \quad \text{for } j = 1, \ldots, N, \tag{42}$$

for $t = 2, \ldots, T$, or $t = 3, \ldots, T+1$, and obtain a robust forecast given by $\widehat{x}_{j,T+h|T+h-1} = x_{j,T+h-1} + \widehat{\delta}_{j,1} \Delta x_{j,T+h-1}$. The forecast error is $\widetilde{\widetilde{\epsilon}}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \widetilde{\widetilde{\mathbf{w}}}_{T+h-1}$, where $\widetilde{\widetilde{\mathbf{w}}}_{T+h|T+h-1} = \left(y_{T+h-1} : \mathbf{x}_{T+h-1} + \widehat{\boldsymbol{\delta}}_1 \Delta \mathbf{x}_{T+h-1} : \mathbf{x}_{T+h-1}\right)$.

(v) unknown future exogenous regressors using an AR(1) for each regressor to obtain forecasts for the retained exogenous regressors. Estimate (42) for $j = 1, \ldots, N$ to obtain $\widehat{x}_{j,T+h|T+h-1} = \widehat{\delta}_{j,0} + \widehat{\delta}_{j,1} x_{j,T+h-1}$, such that $\widehat{\mathbf{w}}_{T+h|T+h-1} = \left(y_{T+h-1} : \widehat{\mathbf{x}}_{T+h|T+h-1} : \mathbf{x}_{T+h-1}\right)$. The resulting forecast error is $\widehat{\widehat{\epsilon}}_{T+h|T+h-1} = y_{T+h} - \widehat{\beta}_0 - \widehat{\boldsymbol{\beta}}_s' \widehat{\mathbf{w}}_{T+h|T+h-1}$.

(vi) 'direct' univariate forecasts for y_{T+h} including:[4]

(a) a random walk forecast: $\widehat{y}_{T+h|T+h-1} = y_{T+h-1}$; and

(b) an AR(1) forecast: $\widehat{y}_{T+h|T+h-1} = \widehat{\gamma}_0 + \widehat{\gamma}_1 y_{T+h-1}$;

such that no exogenous variables are used to forecast y_{T+h} for $h = 1, 2$.

(vii) pooling, computed using an equally weighted average of:
(iii) forecasts of exogenous regressors using a selected model from the GUM (41),
(iv,a) the robust random walk for the exogenous regressors, and
(vi,b) a direct univariate forecast of the endogenous variable using an AR(1).[5]

Cases (ii) and (iv) are the two extremes of the class of robust forecasting devices proposed by Castle, Clements, and Hendry (2015) whereby the fundamental parameters (equilibrium mean and growth rate) are estimated by varying amounts of past data. The full in-sample mean as a forecasting device, used in (ii), is the least robust, and the instantaneous estimate of the mean in (iv) is the most robust. Intermediate cases, such as the average of the last r observations to obtain an estimate of the mean of $\mathbf{w}$, could also be considered.

Results are also reported for estimation of the GUM, equation (39), and estimation of the DGP (36). Forecasts are obtained by plugging in values for all regressors in the GUM/DGP using methods (i)–(v). Note that for (iii), the in-sample selected forecast, the GUM and DGP have different forecasts for different values of α as the α refers to the significance level for selection of the models to produce forecasts of the regressors (41) as well as the significance level for selection of the forecasting model (i.e. $\alpha = 1$ for the DGP and GUM). The DGP provides the infeasible benchmark as it cannot be known in practice, but it allows the costs of selection to be measured by comparing the forecasts from the selected model relative to had the DGP been known. Comparison with the GUM is also informative to measure the costs and benefits of search.

8.2 Results

Tables 10-15 record the MSFEs for the range of experiments, with each table corresponding to a set of experiments for a given $\boldsymbol{\psi}$ specification and a given horizon; $T+1|T$ or $T+2|T+1$, where $T+1|T$

[4] 'Direct' refers to forecasts that ignore conditioning variables and just use the endogenous variable in producing forecasts.

[5] Forecast averages are also computed for an equally weighted average of: {(iv,a) and (v)}, {(iv,b) and (v)}, {(iv,a) and (iii)}, {(iv,b) and (iii)}, {(iv,a), (iii) and (v)}, {(vi,a) and (vi,b)}, {(iv,a), (v) and (vi,b)}, and {(iv,a), (v) and (iii)}. The pooled forecast for {(iv,a), (vi,b) and (iii)} was found to outperform, but results for the other pooled forecasts are available on request.

implies the break occurs out-of-sample (as the location shift occurs at $T+1$) and $T+2|T+1$ is where the break occurs in-sample. Where a number is only reported in the first row of the table corresponding to $\alpha = 0.001$, the same MSFE applies to all values of α.

8.2.1 Potency and Gauge

Before considering forecast performance, selection is examined using potency and gauge, given by:

$$\begin{array}{lcll} \text{retention rate: } \widetilde{\mathsf{p}}_j & = & \frac{1}{M}\sum_{i=1}^{M} 1_{\{\widehat{\beta}_{j,i}\neq 0\}}, & j=1,\ldots,2N+1, \\ \text{potency} & = & \frac{1}{n+1}\sum_{j=1}^{n+1}\widetilde{\mathsf{p}}_j, & \\ \text{gauge} & = & \frac{1}{2N-n}\sum_{j=n+2}^{2N+1}\widetilde{\mathsf{p}}_j, & \end{array} \tag{43}$$

where the GUM includes lags of the regressors and the lagged dependent variable so there are $2N+1$ possible variables, of which $n+1$ are relevant. The intercept is forced so is excluded from the potency calculations.

Figure 9 records the gauge averaged across all experiments, recorded at each target significance level. The very narrow range for the gauge at each α is evident, despite wide variations in DGP specifications. The gauge is slightly above the target significance level at very tight significance levels, and is too low at very loose significance levels, but is well calibrated for the intermediate target significance levels, almost exactly for a significance level of 16%.

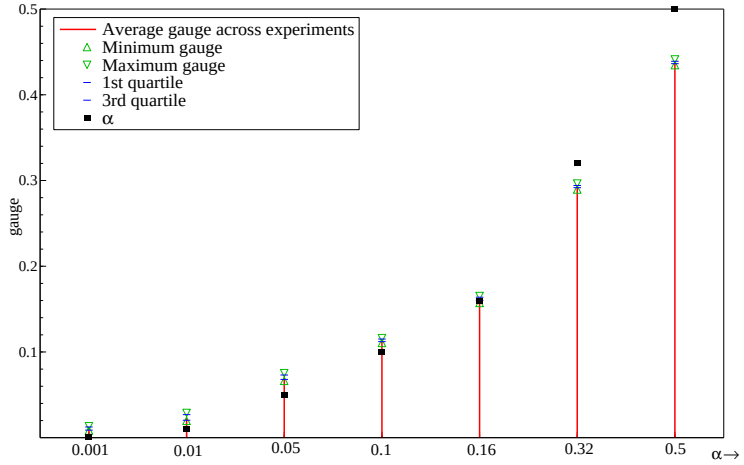

Figure 9: Gauge across all experiments (two forecast horizons, three ψ specifications and all types of break) with target significance level α.

Figure 10 records the potency against the theoretical probability. The theoretical retention rate is calculated as the average of the probability of retaining a single relevant variable assuming a single independent $\mathtt{t}-$test at a given significance level for each relevant variable, so includes the probability of retaining the lagged dependent variable (which is 0.983 for $\alpha = 0.001$, increasing to 1 for larger α). Retention probabilities for the exogenous regressors are given in Table 8. The variation in potency is narrow across all experiments, and the potency is very close to the theoretical probabilities. At tighter significance levels there are cases where the potency exceeds the retention rate for a single test, due to a looser gauge than the significance level. Nevertheless, the probability of retaining the DGP is much smaller, as Table 8 shows.

To summarise, selection using *Autometrics* has the expected properties with a well-calibrated null rejection frequency close to the chosen significance level for a wide range of values of α, and with non-null rejections close to the powers of one-off $\mathtt{t}-$tests with the same non-centralities. Consequently, it

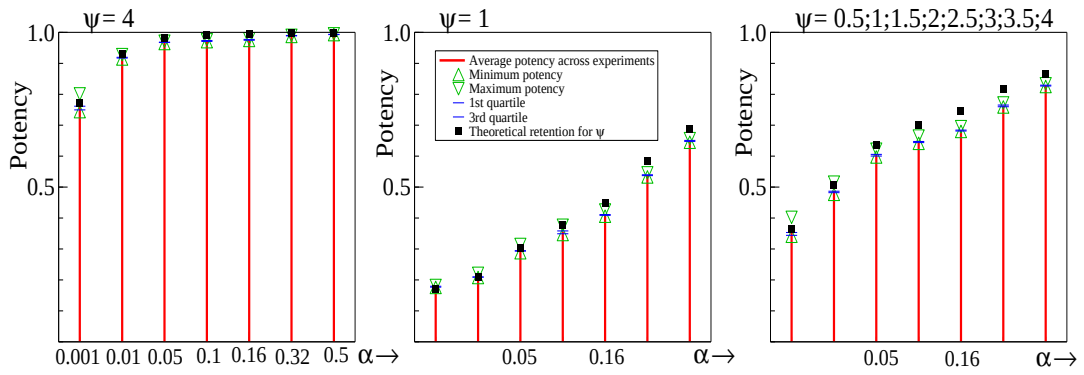

Figure 10: Potency across experiments for each ψ specification in equation (38), i.e. (**a**) $\boldsymbol{\psi = 4}$; (**b**) $\boldsymbol{\psi = 1}$; (**c**) $\boldsymbol{\psi} = 0.5, \ldots, 4$. Theoretical retention is the average probability of retaining each of the relevant regressors including the lagged dependent variable and either five variables with a non-centralities of 4 or 1 or eight regressors ranging from $0.5, \ldots 4$.

is appropriate to use *Autometrics* to evaluate the theoretical results derived in sections 3-6 by simulation, without concern that the selection algorithm will influence the results relative to the single $\mathsf{t}-$test approach analyzed above.

8.2.2 Minimum MSFE methods

We can infer some general results from the appendix tables 10–15. First, when the break occurs out-of-sample, so forecasts are computed for $T+1$ when the break occurs at $T+1$, the two dominant methods across all experiments are using the in-sample mean to forecast the exogenous regressors, and the pooled forecast. For the experiment having 5 relevant regressors with non-centralities of 4, a slightly tighter significance level of between 1% and 5% dominates, but for the other two experiments (non-centralities of 1 or a range of non-centralities), a selection significance level of 10% often out-performs, with 16% frequently ranked second.

Moving to the case when the break occurs in-sample, so the forecasts are computed for $T+2$ when the break occurs at $T+1$, the robust device given by a random walk augmented with the difference weighted by the persistence parameter is preferred when the break occurs in the relevant or all regressors. Looser significance levels tend to do well here, at roughly 10% for the experiments with non-centralities of 4, and even looser at 32%–50% for a range of non-centralities. If the breaks occur in the irrelevant regressors, pooling works well, but here tighter significance levels are preferred. When the non-centralities are small ($\psi_i = 1 \forall i$), pooling outperforms the robust device, and just using the sample mean still works well for breaks in irrelevant variables.

There are substantial differences in the forecast performance of the two robust devices. When there is a break in the relevant regressors, the random walk plus the difference notably improves on the random walk forecasts, cutting the MSFE by a half to two-thirds in some cases. If there are no breaks, or breaks occur in the irrelevant regressors, the random walk forecast is preferred to the random walk with difference. The benefits of using the random walk device with the difference are so large if there are breaks that this method dominates if the source of the break is unknown.

The variation in MSFEs across α is very small for intermediate values of α relative to the variation in MSFEs across break types and DGP designs. Too tight or too loose α (0.1% or 50%) can worsen the MSFE substantially, but for moderate α the selection significance level does not have a huge impact on forecast performance. This is an encouraging finding showing that forecast performance is relatively unaffected by the precise choice of significance level for selection when using *Autometrics*, despite a wide range of non-centralities and numbers of relevant and irrelevant exogenous variables.

8.2.3 Forecast rankings

The appendix tables 10–15 are summarised in tables 5-7 which report outcomes when selecting at $\alpha = 5\%$ and 16%, comparing to the DGP and known regressors as infeasible benchmarks to evaluate the costs of selection against knowing the model. Highlighted cells are the minimum MSFE for selection comparing between the sample mean, robust random walk with difference and the pooled forecast: bold shows cases where knowing the DGP, but not the future values of the regressors, dominates.

While the results are mixed, selection at 5% is preferred for the non-centralities of 4, but 16% often dominates for the non-centralities of 1, or mixed non-centralities. If the signal coming from the relevant regressors is strong, a tighter significance level enables elimination of many more irrelevant variables at low cost, but with a weak signal the trade-off between retaining irrelevant variables and omitting relevant variables is more finely balanced, and a significance level of 16% as derived in the theory above can dominate. Perhaps the most surprising finding is that knowing the DGP would only have been preferred in 4 out of the 14 cases, irrespective of the non-centralities of the relevant variables, although also knowing the future values of the regressors (and hence the breaks) would always have dominated.

Table 4 ranks the forecast performance of each method for selection at $\alpha = 10\%$, where **1** is the minimum MSFE and *8* is the worst MSFE for the experiment design. A 10% significance level is a reasonable choice for 15 variables in the GUM when at least half are likely to be not very relevant, and the largest number of minimum MSFEs occurs at 10% reflecting the balance discussed above. Forecast pooling is consistently ranked in the top half of methods, suggesting that it is a successful insurance policy. An AR(1) for y also performs well across the board, matching an oft-found outcome. This model is mis-specified, ignoring all information from the exogenous regressors, but mis-specification need not entail forecast failure. Indeed, the costs of forecasting the exogenous regressors can outweight their inclusion. However, the DGP design is an AR(1) in y so this forecasting device has the advantage of correctly specifying the dynamics. Such a naive device may not perform so well if the DGP contained more complex dynamics.

However, using the AR(1) for the exogenous regressors performs worst across all experiments, despite the DGP dynamics for the exogenous regressors being an AR(1) process. The random walk with difference forecasts oscillate between being ranked first for relevant breaks in-sample to one of the worst for breaks out-of-sample. In contrast, the in-sample mean switches from being the best forecast for breaks out-of-sample, but the worst for breaks in-sample. Unfortunately the forecaster does not know which world they will be in when computing forecasts, although IIS may help.

8.2.4 Is selection costly when forecasting?

Comparing selection to the DGP at 5% or 16% in tables 5–7, the costs of selection can be very small when the regressors are unknown and must be forecast. As knowing the precise formulation of the DGP is always infeasible, selection must be undertaken. Bold cells indicate where knowing the DGP, but not future values of regressors, would have dominated, and there are many quadrants where selection delivers a smaller MSFE. These results demonstrate that selection incurs almost no cost relative to the DGP, with the caveat that a very tight significance level of $\alpha = 0.1\%$ can increase the MSFE relative to the DGP. Known future regressors (i) are omitted from the tables but almost always deliver the best forecast performance, although would be infeasible in practice.

Figure 11 records a scatter plot of the MSFEs for the selected model at three alternative significance levels against the MSFEs for the DGP, for three alternative forecasting devices including selecting a forecasting model for the regressors, using the in-sample mean and the robust device augmented with the difference, across all experiments conducted. The solid black line is the 45° line, so if the MSFEs lie on the line then there is no cost to selecting from a more general model compared to **knowing** the DGP. At very tight significance levels, selection does increase MSFEs, particularly if the regressors are forecast by selecting a model from the GUM, but at 5% or 16% almost all observations lie on or very close to the

Table 4: Simulation summary rankings for $\alpha = 10\%$. 'Out' refers to forecasts for $T+1|T$, i.e. the break is out-of-sample. 'In' refers to forecasts for $T+2|T+1$ where the break is in-sample. (1) is for the case with $\psi = (0,0,0,0,0,0,0,0.5,1,1.5,2,2.5,3,3.5,4)'$, (2) is case $\psi = (0,0,0,0,0,0,0,0,0,0,4,4,4,4,4)'$, and (3) is for $\psi = (0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1)'$. Lower case Roman numerals respectively denote forecasting the unknown future exogenous regressors by: (ii) the in-sample mean; (iii) selecting from the GUM (41); (iva) a random walk; (ivb) that with the added difference; (v) an AR(1); (via) a direct random walk forecast of y; (vib) a direct AR(1) forecast of y; and (vii) pooling.

			(ii)	(iii)	(iva)	(ivb)	(v)	(via)	(vib)	(vii)
No Break										
	(1)	Out	3	4	5	7	*8*	6	**1**	2
		In	3	4	5	7	*8*	6	**1**	2
	(2)	Out	4	3	5	7	*8*	6	2	**1**
		In	3	4	5	7	*8*	6	2	**1**
	(3)	Out	2	4	5	6	*8*	7	**1**	3
		In	2	4	5	6	*8*	7	**1**	3
Break Relevant										
	(1)	Out	1	3	6	7	*8*	5	2	4
ι		In	*8*	3	2	**1**	6	5	7	4
	(2)	Out	1	4	5	7	*8*	6	2	3
		In	*8*	4	2	**1**	6	5	7	3
	(3)	Out	1	4	5	6	*8*	7	2	3
		In	*8*	4	2	**1**	7	3	6	5
	(1)	Out	5	2	4	7	*8*	6	3	**1**
λ		In	*8*	4	2	**1**	6	5	7	3
	(2)	Out	7	3	2	6	*8*	5	4	**1**
		In	*8*	4	2	**1**	6	5	7	3
	(3)	Out	2	4	5	6	*8*	7	**1**	3
		In	7	4	3	2	*8*	5	6	**1**
Break Irrelevant										
	(1)	Out	3	4	5	7	*8*	6	**1**	2
ι		In	3	6	4	7	*8*	5	**1**	2
	(2)	Out	3	4	5	7	*8*	6	2	**1**
		In	3	6	5	7	*8*	4	**1**	2
	(3)	Out	2	5	4	6	*8*	7	**1**	3
		In	2	6	4	7	*8*	5	**1**	3
	(1)	Out	3	4	5	7	*8*	6	2	**1**
λ		In	3	4	5	7	*8*	6	**1**	2
	(2)	Out	4	3	5	7	*8*	6	2	**1**
		In	3	6	4	7	*8*	5	**1**	2
	(3)	Out	2	4	5	6	*8*	7	**1**	3
		In	2	5	4	6	*8*	7	**1**	3
Break All										
	(1)	Out	1	4	5	7	*8*	6	2	3
ι		In	*8*	3	2	**1**	6	5	7	4
	(2)	Out	1	4	5	7	*8*	6	2	3
		In	*8*	3	2	**1**	6	5	7	4
	(3)	Out	1	4	5	6	*8*	7	2	3
		In	*8*	5	2	**1**	7	3	6	4
	(1)	Out	5	2	4	7	*8*	6	3	**1**
λ		In	*8*	3	2	**1**	6	5	7	4
	(2)	Out	7	3	2	6	*8*	5	4	**1**
		In	*8*	4	2	**1**	6	5	7	3
	(3)	Out	2	4	5	6	*8*	7	**1**	3
		In	7	4	3	2	*8*	5	6	**1**
Average			4.2	4.0	3.9	5.1	7.6	5.6	3.1	2.5

Table 5: Simulation summary for 8 relevant variables with non-centralities of $0.5; 1; 1.5; 2; 2.5; 3; 3.5; 4$ and 7 irrelevant variables. Underlined cells indicate minimum MSFE for selection across methods listed; bold where knowing the DGP, but not the future values of the regressors, would have dominated.

Break type & case		Out: $T+1\|T$			In: $T+2\|T+1$		
		DGP	$\alpha=0.05$	$\alpha=0.16$	DGP	$\alpha=0.05$	$\alpha=0.16$
No break							
	(ii) sample mean	1.59	1.63	1.64	1.57	1.59	1.62
	(iv,b) RW with diff.	2.12	2.26	2.30	2.02	2.11	2.18
	(vii) pooling		1.52	1.52		1.49	1.53
Break Relevant							
ι	(ii) sample mean	17.56	17.58	17.54	39.50	40.40	41.28
	(iv,b) RW with diff.	18.61	18.77	18.96	**2.53**	3.91	3.46
	(vii) pooling		17.75	17.79		17.99	16.54
λ	(ii) sample mean	4.46	4.51	4.50	11.96	12.12	12.35
	(iv,b) RW with diff.	4.38	4.62	4.66	**2.43**	3.40	3.11
	(vii) pooling		4.16	4.16		7.07	6.76
Break Irrelevant							
ι	(ii) sample mean	1.59	1.63	1.64	1.57	1.59	1.60
	(iv,b) RW with diff.	2.11	2.25	2.31	2.01	2.21	2.30
	(vii) pooling		1.52	1.54		1.54	1.57
λ	(ii) sample mean	1.59	1.63	1.64	1.57	1.59	1.61
	(iv,b) RW with diff.	2.12	2.25	2.31	2.02	2.11	2.20
	(vii) pooling		1.52	1.53		1.51	1.55
Break All							
ι	(ii) sample mean	17.88	17.90	17.86	40.01	40.93	41.69
	(iv,b) RW with diff.	18.86	18.99	19.12	**2.53**	3.76	3.46
	(vii) pooling		18.02	18.00		17.30	15.50
λ	(ii) sample mean	4.50	4.55	4.55	12.06	12.23	12.45
	(iv,b) RW with diff.	4.40	4.63	4.68	**2.42**	3.37	3.15
	(vii) pooling		4.19	4.18		6.99	6.64

$45°$ line, so not knowing the correct model and undertaking selection at fairly loose significance levels is not costly in a forecasting context.

8.2.5 Explaining the variation in forecast performance

There are 2142 distinct MSFE observations excluding results for the GUM and DGP, with a mean of 5.15 and a standard deviation of 7.50. Attempts to explain the main characteristics of the results in a response surface highlighted the high degree of non-linearity, large number of interaction terms and many indicator variables retained using impulse indicator saturation needed to obtain a congruent specification. The results do not lend themselves to a parsimonious response surface specification.

However, the analysis does highlight some important aspects that explain forecast performance across experiments. Some characteristics are self-evident: breaks in relevant variables, breaks in the intercept and breaks in experiments with regressors with large non-centralities result in the largest MSFEs. Features that matter across specifications are potency, gauge and the theoretical retention probability given ψ. Higher potency improves forecast performance as does higher retention probability, so both the theoretical and empirical measures of retaining relevant variables matter. The overall effect of gauge varies by forecast method, with the retention of more irrelevant variables less problematic for known regressors than if the regressors need to be forecast. The retention probabilities are recorded in Table 8,

Table 6: Simulation summary for 5 relevant variables with non-centralities of 4 and 10 irrelevant variables. Underlined cells indicate minimum MSFE for selection across methods listed; bold where knowing the DGP, but not the future values of the regressors, would have dominated.

Break type & case		Out: $T+1\|T$			In: $T+2\|T+1$		
		DGP	$\alpha=0.05$	$\alpha=0.16$	DGP	$\alpha=0.05$	$\alpha=0.16$
No break							
	(ii) sample mean	1.90	1.96	1.97	1.89	1.93	1.95
	(iv,b) RW with diff.	2.59	2.68	2.77	2.51	2.62	2.71
	(vii) pooling		1.73	1.77		1.76	1.80
Break Relevant							
ι	(ii) sample mean	21.22	21.20	21.22	47.18	48.42	49.52
	(iv,b) RW with diff.	22.60	22.77	22.78	**2.92**	3.37	3.41
	(vii) pooling		21.51	21.56		20.13	19.51
λ	(ii) sample mean	5.46	5.52	5.54	14.52	14.80	15.11
	(iv,b) RW with diff.	5.17	5.35	5.44	**2.87**	3.20	3.28
	(vii) pooling		4.91	4.96		8.04	7.99
Break Irrelevant							
ι	(ii) sample mean	1.89	1.95	1.96	1.88	1.92	1.94
	(iv,b) RW with diff.	2.57	2.67	2.75	2.49	2.73	2.90
	(vii) pooling		1.74	1.78		1.82	1.94
λ	(ii) sample mean	1.90	1.96	1.97	1.89	1.93	1.95
	(iv,b) RW with diff.	2.58	2.67	2.77	2.51	2.62	2.65
	(vii) pooling		1.73	1.77		1.79	1.84
Break All							
ι	(ii) sample mean	21.88	21.88	21.89	48.27	49.58	50.64
	(iv,b) RW with diff.	23.14	23.30	23.32	**2.94**	3.43	3.59
	(vii) pooling		22.10	22.14		18.92	17.70
λ	(ii) sample mean	5.55	5.60	5.64	14.71	15.01	15.28
	(iv,b) RW with diff.	5.21	5.40	5.47	**2.87**	3.17	3.29
	(vii) pooling		4.98	5.01		7.88	7.77

which compute the probability of retaining all relevant regressors assuming independent t−tests. While the probability of retaining one variable may be quite large, the joint probability of retaining all relevant variables can be extremely low. Thus, even using a significance level of 16%, many relevant variables will be omitted if their non-centralities are small. However, their contribution to explaining the dependent variable is also small and therefore breaks in such variables will have a smaller effect. The break magnitude will be a function of the coefficient in the conditional model, the degree of persistence governing the break evolution, and the size of the break in the marginal model.

The choice of α interacts with whether the break occurs in the relevant or irrelevant regressors, such that a tighter significance level is preferred for breaks in irrelevant regressors, effectively increasing the chance of their removal, but a looser significance level is preferred for breaks in relevant variables, consistent with retaining such variables being important when the variables shift.

If the relevant variables have large non-centralities, standard significance levels of 1% or 5% are preferred to minimize MSFE, but if the relevant regressors range in non-centrality from small to large, a looser significance level is preferred. For the practitioner who was uncertain of the nature of the unknown DGP, a moderate selection significance level of $\alpha = 10\%$–16% insures against the extremes, although there will be cases when such a choice is not optimal.

Table 7: Simulation summary for 5 relevant variables with non-centralities of 1 and 10 irrelevant variables. Underlined cells indicate minimum MSFE for selection across methods listed; bold where knowing the DGP, but not the future values of the regressors, would have dominated.

Break type & case		Out: $T+1\|T$			In: $T+2\|T+1$		
		DGP	$\alpha=0.05$	$\alpha=0.16$	DGP	$\alpha=0.05$	$\alpha=0.16$
No break							
	(ii) sample mean	**1.06**	1.09	1.09	**1.06**	1.08	1.08
	(iv,b) RW with diff.	1.19	1.29	1.37	1.20	1.23	1.33
	(vii) pooling		1.11	1.12		1.08	1.10
Break Relevant							
ι	(ii) sample mean	**2.32**	2.34	2.34	4.07	4.22	4.26
	(iv,b) RW with diff.	2.52	2.61	2.73	**1.53**	2.60	2.18
	(vii) pooling		2.39	2.42		3.09	2.73
λ	(ii) sample mean	**1.30**	1.33	1.33	1.88	1.93	1.94
	(iv,b) RW with diff.	1.38	1.53	1.60	**1.28**	1.70	1.63
	(vii) pooling		1.35	1.36		1.70	1.62
Break Irrelevant							
ι	(ii) sample mean	**1.06**	1.09	1.09	**1.06**	1.07	1.08
	(iv,b) RW with diff.	1.19	1.28	1.37	1.20	1.40	1.62
	(vii) pooling		1.11	1.12		1.10	1.14
λ	(ii) sample mean	**1.06**	1.09	1.09	**1.06**	1.08	1.08
	(iv,b) RW with diff.	1.19	1.29	1.38	1.20	1.27	1.34
	(vii) pooling		1.11	1.12		1.10	1.10
Break All							
ι	(ii) sample mean	**2.36**	2.38	2.38	4.14	4.28	4.34
	(iv,b) RW with diff.	2.56	2.66	2.74	**1.53**	2.71	2.33
	(vii) pooling		2.43	2.45		3.05	2.67
λ	(ii) sample mean	**1.31**	1.34	1.34	1.90	1.94	1.95
	(iv,b) RW with diff.	1.39	1.53	1.59	**1.28**	1.72	1.66
	(vii) pooling		1.35	1.36		1.70	1.61

Table 8: Retention probabilities $\mathsf{P}\left(\text{reject } H_0|\boldsymbol{\psi},\alpha\right)$; $\times 5$ refers to the probability of retaining all 5 regressors with a given non-centrality and "Joint" is the probability of retaining all regressors with non-centralities of $(0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4)$, assuming independence.

α	$\psi=4$	$\times 5$	$\psi=1$	$\times 5$	$\psi=0.5$	$\psi=1.5$	$\psi=2$	$\psi=2.5$	$\psi=3$	$\psi=3.5$	Joint
0.001	0.729	0.206	0.008	0.000	0.002	0.029	0.082	0.187	0.348	0.544	0.000
0.01	0.915	0.642	0.052	0.000	0.017	0.130	0.266	0.450	0.646	0.809	0.000
0.05	0.978	0.895	0.163	0.000	0.069	0.314	0.506	0.697	0.845	0.935	0.001
0.1	0.990	0.953	0.255	0.001	0.123	0.436	0.633	0.799	0.910	0.967	0.006
0.16	0.995	0.976	0.339	0.004	0.180	0.534	0.721	0.861	0.943	0.981	0.019
0.32	0.999	0.993	0.500	0.031	0.309	0.692	0.841	0.933	0.977	0.994	0.081
0.5	1.000	0.998	0.627	0.097	0.430	0.795	0.907	0.966	0.990	0.998	0.185

9 Conclusion

The paper investigates the choice of significance level and its associated critical value when selecting forecasting models, both analytically in a static bivariate setting where there are location shifts at the forecast origin, and in more general simulation experiments. The theory suggests that variables should

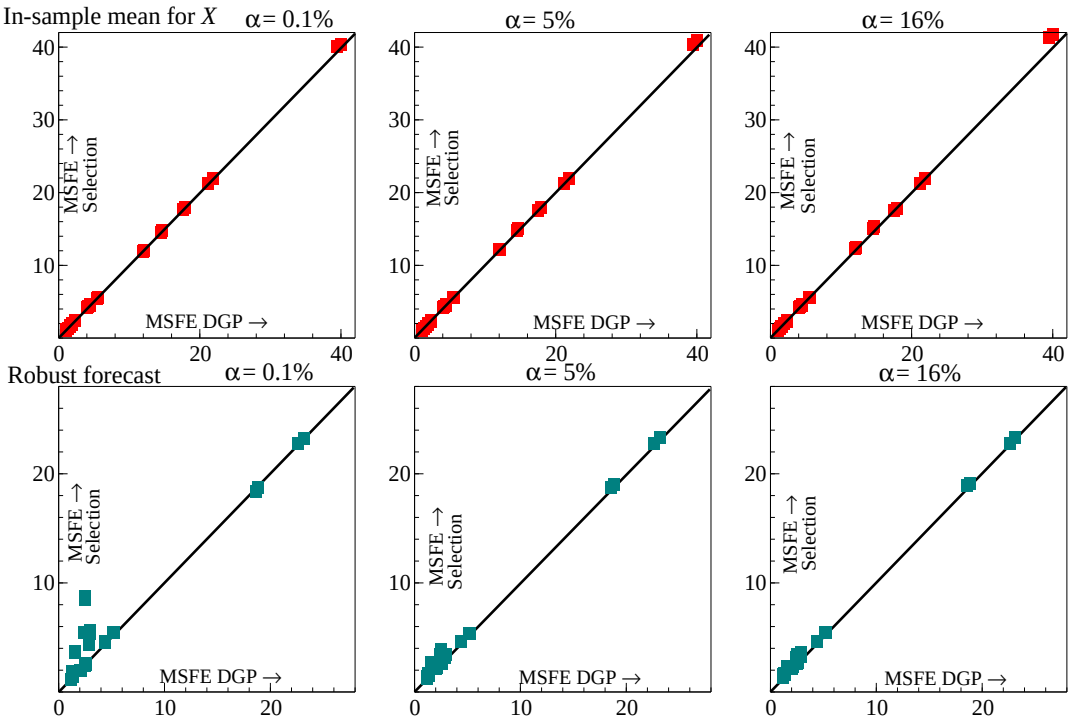

Figure 11: Scatter plots of the MSFE for model selection against the MSFE for DGP for unknown future regressors which are forecast using: (top row) the in-sample mean of the exogenous regressors, and (bottom row) the robust device augmented with the difference, for significance levels of 0.1% (left hand panel), 5% (middle panel) and 16% (right hand panel). The solid black line is the 45° line. The experiments include the three ψ specifications, two forecast horizons, and all 7 break specifications.

be retained if their non-centralities exceed 1, which translates to $c_\alpha^2 = 2$ at the boundary. This result holds regardless of whether or not location shifts affect the variable about which a retention decision is made. Undertaking selection at such loose significance levels implies that fewer relevant variables will be excluded when they contribute to forecast accuracy, but that more variables will be retained by chance because they happen to be in a draw that results in statistical significance at the proposed critical value. Although retaining irrelevant variables that are subject to location shifts usually worsens forecast performance, their coefficient estimates will be driven towards zero when updating estimates as the horizon moves forward. Indeed, in a progressive research strategy of learning sequentially from evidence, large breaks in irrelevant variables will rapidly lead to their being omitted and focus the specification on relevant variables.

Although the static design is simple, it is not restrictive. The analytic results hold regardless of whether the regressors are contemporaneous or lagged, although the timing of location shifts is fundamental. Dynamics will slow adjustment to new equilibria, but would not change the essence of the results. The inflation forecasting illustration demonstrates the analytic results, with a loose selection significance level of 16% being preferred for both the known regressors and the random walk forecasts for unknown regressors case.

The simulation evidence examines a wide range of experimental designs and despite the disparate outcomes, they provide some guidance for forecasting. The ideal scenario is obviously to have **complete** knowledge of the DGP, such that the empirical modeller knows the number and magnitude of both relevant and irrelevant regressors, and their future values, and hence whether and where breaks are likely to occur. In practice, no-one has the benefit of omniscience, and once the future values of regressors need to be forecast, not knowing the precise specification of the DGP may not be costly relative to selecting

from a GUM that nests it. Indeed, simply knowing the specification of the DGP, but needing to forecast future values of the exogenous regressors, rarely delivered the best MSFE outcome.

The simulation results suggest that if the model is being used primarily for 1-step ahead forecasting with the aim of minimizing MSFE, selection at looser than standard selection significance levels may well help, and doing so will rarely hinder forecast performance. The results provide some support for selecting models at around 10% when there are approximately 15 regressors, many of which are irrelevant. This is close to the 16% derived theoretically in the paper when the number of irrelevant regressors is small. The simulation results also highlight the degree of complexity in pinning down the optimal selection rule for forecasting, with results depending on all aspects of the experimental design. A take-away for the forecaster is that pooling works well across many settings, suggesting a combination of model-based forecast, robust device and univariate methods provides a good insurance policy. Moreover, methods that did not nest the DGP, such as the direct AR(1) forecast of the dependent variable, also performed well, both matching commonly found empirical outcomes.

A Analytic Calculations

A.1 Derivations for the equations reported in §3.1

The DGP given in (1), (2) and (3) results in

$$\sqrt{T}\begin{pmatrix}\widehat{\beta}_1-\beta_1\\ \widehat{\beta}_2-\beta_2\end{pmatrix}\sim \mathsf{N}_2\left[\begin{pmatrix}0\\0\end{pmatrix},\frac{\sigma_\epsilon^2}{\sigma_{11}^2\sigma_{22}^2\left(1-\rho^2\right)}\begin{pmatrix}\sigma_{22}^2 & -\rho\sigma_{11}\sigma_{22}\\ -\rho\sigma_{11}\sigma_{22} & \sigma_{11}^2\end{pmatrix}\right], \tag{44}$$

with:

$$\sqrt{T}\left(\mu_y-\widehat{\mu}_y\right)\sim \mathsf{N}\left[0,\sigma_\epsilon^2\right], \tag{45}$$

where we subsequently set $\sigma_{11}=\sigma_{22}=0$ without loss of generality.

M_2 in (6) partials out $x_{2,t}$. From (2) we can write in deviations from means for $t=1,\ldots,T$:

$$x_{2,t}-\mu_2=\rho\left(x_{1,t}-\mu_1\right)+e_t, \tag{46}$$

such that $e_t=\eta_{2,t}-\rho\eta_{1,t}$, so $\gamma_1=(\beta_1+\beta_2\rho)$ and $\phi_0=\mu_y-\gamma_1\mu_1$. Hence M_2 is:

$$\begin{aligned} y_t &= \mu_y+(\beta_1+\beta_2\rho)\left(x_{1,t}-\mu_1\right)+\beta_2 e_t+\epsilon_t\\ &= \gamma_0+\gamma_1\left(x_{1,t}-\mu_1\right)+\nu_t, \end{aligned}$$

with $\gamma_0=\mu_y$. The error for M_2 is given by:

$$\nu_t=\beta_2\left(\eta_{2,t}-\rho\eta_{1,t}\right)+\epsilon_t, \tag{47}$$

where:

$$\sigma_\nu^2=\sigma_\epsilon^2+\beta_2^2\left(1-\rho^2\right)=\sigma_\epsilon^2\left(1+T^{-1}\psi^2\right)\geq\sigma_\epsilon^2. \tag{48}$$

Also:

$$\sqrt{T}\begin{pmatrix}\widetilde{\gamma}_0-\gamma_0\\ \widetilde{\gamma}_1-\gamma_1\end{pmatrix}\sim \mathsf{N}_2\left[\begin{pmatrix}0\\0\end{pmatrix},\sigma_\nu^2\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right]. \tag{49}$$

A.2 Derivations for the equations reported in §4

The 1-step ahead forecast error from M_1 is:

$$\begin{aligned} \widehat{\epsilon}_{T+1|T} &= y_{T+1}-\widehat{y}_{T+1|T}\\ &= \left(\mu_y-\widehat{\mu}_y\right)+\left(\beta_1-\widehat{\beta}_1\right)\left(x_{1,T+1}-\mu_1\right)+\left(\beta_2-\widehat{\beta}_2\right)\left(x_{2,T+1}-\mu_2\right)+\epsilon_{T+1}. \end{aligned} \tag{50}$$

When there are no breaks, the parameter estimates are unbiased, $\mathsf{E}\left[\widehat{\epsilon}_{T+1|T}\right]=0$ so the MSFE of M_1 is:

$$\mathsf{E}\left[\widehat{\epsilon}_{T+1|T}^2\right]=\sigma_\epsilon^2\left(1+\frac{1}{T}+\frac{2}{T\left(1-\rho^2\right)}-\frac{2\rho^2}{T\left(1-\rho^2\right)}\right)=\sigma_\epsilon^2\left(1+\frac{3}{T}\right). \tag{51}$$

The 1-step ahead forecast error from M_2 in which $x_{2,t}$ is omitted is:

$$\begin{aligned}\widetilde{\epsilon}_{T+1|T} &= y_{T+1}-\widetilde{y}_{T+1|T}\\ &= \beta_2\eta_{2,T+1}+\epsilon_{T+1}+\left(\gamma_0-\widetilde{\gamma}_0\right)+\left(\beta_1-\widetilde{\gamma}_1\right)\eta_{1,T+1}.\end{aligned} \tag{52}$$

Therefore, despite the mis-specification, $\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}\right]=0$ and the MSFE is:

$$\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}^2\right]=\mathsf{E}\left[\left(\beta_2\eta_{2,T+1}+\epsilon_{T+1}+\left(\gamma_0-\widetilde{\gamma}_0\right)+\left(\beta_1-\widetilde{\gamma}_1\right)\eta_{1,T+1}\right)^2\right]=\sigma_\nu^2\left(1+\frac{2}{T}\right). \tag{53}$$

A.3 Derivations for the equations reported in §5.1

The regression equation itself stays constant so:

$$y_{T+1}=\left(\mu_y+\beta_2\delta\right)+\beta_1\left(x_{1,T+1}-\mu_1\right)+\beta_2\left(x_{2,T+1}-\mu_2-\delta\right)+\epsilon_{T+1}. \tag{54}$$

Consequently, using $\widehat{\beta}_0=\mu_y-\widehat{\beta}_1\mu_1-\widehat{\beta}_2\mu_2$ to match the formulation of M_2, the forecast for M_1 is:

$$\overline{\widehat{y}}_{T+1|T+1}=\mu_y+\widehat{\beta}_2\delta+\widehat{\beta}_1\left(x_{1,T+1}-\mu_1\right)+\widehat{\beta}_2\left(x_{2,T+1}-\mu_2-\delta\right), \tag{55}$$

and the 1-step ahead forecast error for M_1 is:

$$\begin{aligned}\overline{\widehat{\epsilon}}_{T+1|T+1} &= y_{T+1}-\overline{\widehat{y}}_{T+1|T+1}\\ &= \left(\beta_2-\widehat{\beta}_2\right)\delta+\left(\beta_1-\widehat{\beta}_1\right)\eta_{1,T+1}+\left(\beta_2-\widehat{\beta}_2\right)\eta_{2,T+1}+\epsilon_{T+1},\end{aligned} \tag{56}$$

and a 1-step ahead MSFE of:

$$\mathsf{E}\left[\overline{\widehat{\epsilon}}_{T+1|T+1}^2\right]=\sigma_\epsilon^2\left(1+\frac{\delta^2+2-\rho}{T\left(1-\rho^2\right)}\right). \tag{57}$$

Next consider the 1-step ahead forecast for M_2, given $\gamma_0=\mu_y$ and $\gamma_1=(\beta_1+\beta_2\rho)$:

$$\overline{\widetilde{y}}_{T+1|T+1}=\widetilde{\gamma}_0+\widetilde{\gamma}_1\left(x_{1,T+1}-\mu_1\right).$$

The 1-step ahead forecast error is given by:

$$\begin{aligned}\overline{\widetilde{\epsilon}}_{T+1|T+1} &= y_{T+1}-\overline{\widetilde{y}}_{T+1|T+1}\\ &= \beta_2\delta+\left(\gamma_0-\widetilde{\gamma}_0\right)+\left(\gamma_1-\widetilde{\gamma}_1\right)\eta_{1,T+1}-\beta_2\rho\eta_{1,T+1}+\beta_2\eta_{2,T+1}+\epsilon_{T+1},\end{aligned}$$

and the 1-step ahead MSFE for M_2 is:

$$\mathsf{E}\left[\overline{\widetilde{\epsilon}}_{T+1|T+1}^2\right]=\sigma_\epsilon^2+\beta_2^2\left(1-\rho^2+\delta^2\right)+2T^{-1}\sigma_\nu^2. \tag{58}$$

A.4 Derivations for the equations reported in §5.3

For $\widehat{\beta}_0 = \mu_y - \widehat{\beta}_1\mu_1 - \widehat{\beta}_2\mu_2$, replacing the unknown $x_{i,T+1}$ by μ_i leads to forecasting y_{T+1} by the in-sample mean:

$$\widehat{\widehat{y}}_{T+1|T} = \mu_y,$$

so the forecast error for M_1 is:

$$\begin{aligned}\widehat{\widehat{\epsilon}}_{T+1|T} &= y_{T+1} - \widehat{\widehat{y}}_{T+1|T} \\ &= \beta_2\delta + \beta_1\eta_{1,T+1} + \beta_2\eta_{2,T+1} + \epsilon_{T+1},\end{aligned} \tag{59}$$

and the forecast error bias is:

$$\mathsf{E}\left[\widehat{\widehat{\epsilon}}_{T+1|T}\right] = \beta_2\delta. \tag{60}$$

The MSFE_1 is:

$$\mathsf{E}\left[\widehat{\widehat{\epsilon}}^2_{T+1|T}\right] = \beta_1^2 + \beta_2^2\left(1+\delta^2\right) + 2\rho\beta_1\beta_2 + \sigma_\epsilon^2. \tag{61}$$

Parameter estimation adds terms of $\mathsf{O}_p\left(T^{-1}\right)$.

Similarly, for M_2, from (6) forecasting $x_{1,T+1}$ by μ_1 leads to:

$$\widetilde{\widetilde{y}}_{T+1|T} = \mu_y,$$

and hence for 'known' μ_y the forecast error is:

$$\widetilde{\widetilde{\epsilon}}_{T+1|T} = \beta_2\delta + \beta_1\eta_{1,T+1} + \beta_2\eta_{2,T+1} + \epsilon_{T+1} = \widehat{\widehat{\epsilon}}_{T+1|T},$$

with:

$$\mathsf{E}\left[\widetilde{\widetilde{\epsilon}}_{T+1|T}\right] = \beta_2\delta,$$

and MSFE_2 is given by (23). Hence, ignoring $\mathsf{O}_p\left(T^{-1}\right)$ terms, $\mathsf{MSFE}_2 = \mathsf{MSFE}_1$.

A.5 Derivations for the equations reported in §5.4

From (54) the regression equation for y_{T+1} can also be written as:

$$y_{T+1} = \left(\mu_y + \beta_2\delta\right) + \beta_1\Delta x_{1,T+1} + \beta_2\left(\Delta x_{2,T+1} - \delta\right) + \epsilon_{T+1} + \beta_1\eta_{1,T} + \beta_2\eta_{2,T}.$$

Furthermore, the forecast for M_1 using (24) and (25) is:

$$\overline{\overline{y}}_{T+1|T} = \mu_y + \widehat{\beta}_1\left(x_{1,T} - \mu_1\right) + \widehat{\beta}_2\left(x_{2,T} - \mu_2\right),$$

so the forecast error for M_1 is:

$$\begin{aligned}\overline{\overline{\epsilon}}_{T+1|T} &= y_{T+1} - \overline{\overline{y}}_{T+1|T} \\ &= \beta_2\delta + \beta_1\Delta x_{1,T+1} + \beta_2\left(\Delta x_{2,T+1} - \delta\right) + \left(\beta_1 - \widehat{\beta}_1\right)\eta_{1,T} + \left(\beta_2 - \widehat{\beta}_2\right)\eta_{2,T} + \epsilon_{T+1}.\end{aligned}$$

Consequently, neglecting the small impact of $\eta_{i,T}$ on $\beta_i - \widehat{\beta}_i$:

$$\mathsf{E}\left[\overline{\overline{\epsilon}}_{T+1|T}\right] = \beta_2\delta,$$

and hence MSFE_1 is:

$$\mathsf{E}\left[\overline{\overline{\epsilon}}^2_{T+1|T}\right] = 2\beta_1^2 + \beta_2^2\left(2+\delta^2\right) + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1+2T^{-1}\right). \tag{62}$$

Next, we compute the equivalent bias and MSFE for M_2, noting $\gamma_1 = \beta_1 + \beta_2\rho$, so that the forecast is given by:

$$\widetilde{\overline{y}}_{T+1|T} = \widetilde{\gamma}_0 + \widetilde{\gamma}_1 \left(x_{1,T} - \mu_1\right)$$

As $\widetilde{\gamma}_0 = \gamma_0 = \mu_y$, the forecast error for M_2 using the random walk is:

$$\begin{aligned} \widetilde{\overline{\epsilon}}_{T+1|T} &= y_{T+1} - \widetilde{\overline{y}}_{T+1|T} \\ &= \beta_2\delta + \beta_1\Delta\eta_{1,T+1} + \beta_2\Delta\eta_{2,T+1} + \epsilon_{T+1} + \left(\beta_1 - \widetilde{\gamma}_1\right)\eta_{1,T} + \beta_2\eta_{2,T}, \end{aligned} \tag{63}$$

where, as before:

$$\mathsf{E}\left[\widetilde{\overline{\epsilon}}_{T+1|T}\right] = \beta_2\delta.$$

Neglecting the small impact of $\eta_{1,T}$ on $\widetilde{\gamma}_1$ the MSFE for M_2 is:

$$\mathsf{E}\left[\widetilde{\overline{\epsilon}}^2_{T+1|T}\right] = 2\beta_1^2 + \beta_2^2\left(3 + \rho^2 + \delta^2\right) + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1 + T^{-1} + T^{-2}\psi^2\right). \tag{64}$$

A.6 Derivations for the equations reported in §6.1

The conditional DGP for the forecast observation is:

$$\begin{aligned} y_{T+1} &= \beta_0 + \beta_1 x_{1,T+1} + \beta_2 x_{2,T+1} + \epsilon_{T+1} \\ &= \left(\mu_y + \beta_2\delta\right) + \beta_1\left(x_{1,T+1} - \mu_1\right) + \beta_2\left(x_{2,T+1} - \mu_2 - \delta\right) + \epsilon_{T+1}, \end{aligned} \tag{65}$$

where the in-sample mean μ_y is shifted to $(\mu_y + \beta_2\delta)$ at T. Sample calculations will be altered as now $\mathsf{E}\left[\overline{x}_2\right] = \mu_2 + T^{-1}\delta$ from:

$$\overline{x}_2 = \frac{1}{T}\sum_{t=1}^{T} x_{2,t} = \mu_2 + T^{-1}\delta + \overline{\eta}_2,$$

and neglecting terms of T^{-2} or smaller:

$$\left(\sigma^*_{22}\right)^2 \approx \sigma^2_{22} + T^{-1}\delta^2,$$

with $\sigma^*_{12} = \sigma_{12}$ implying that:

$$\rho^* = \frac{\sigma_{12}}{\sigma_{11}\sigma^*_{22}}.$$

The intercept is again included with $\widehat{\beta}_0 = \mu_y - \widehat{\beta}_1\mu_1 - \widehat{\beta}_2\mu_2$ to match the formulation of M_2.

$$\widehat{\overline{y}}_{T+1|T+1} \approx \widehat{\beta}_0 + \widehat{\beta}_1\mu_1 + \widehat{\beta}_2\left(\mu_2 + T^{-1}\delta\right) = \mu_y + \widehat{\beta}_2 T^{-1}\delta,$$

and hence neglecting terms of T^{-2} or smaller, the forecast error for M_1 is:

$$\begin{aligned} \widehat{\overline{\epsilon}}_{T+1|T+1} &= y_{T+1} - \widehat{\overline{y}}_{T+1|T+1} \\ &\approx \beta_2\delta\left(1 - T^{-1}\right) + \beta_1\eta_{1,T+1} + \beta_2\eta_{2,T+1} + \epsilon_{T+1}, \end{aligned} \tag{66}$$

so the forecast error bias is given by:

$$\mathsf{E}\left[\widehat{\overline{\epsilon}}_{T+1|T+1}\right] \approx \beta_2\delta\left(1 - T^{-1}\right).$$

The MSFE for M_1 is:

$$\mathsf{E}\left[\widehat{\overline{\epsilon}}^2_{T+1|T+1}\right] = \beta_2^2\delta^2\left(1 - T^{-1}\right)^2 + \beta_1^2 + \beta_2^2 + \sigma_\epsilon^2. \tag{67}$$

Omitting x_2 from the forecasting equation leads to a forecast error of:

$$\begin{aligned}\widehat{\widetilde{\epsilon}}_{T+1|T+1} &= y_{T+1} - \widehat{\widetilde{y}}_{T+1|T+1} \\ &\approx \beta_2\delta + (\gamma_0 - \widetilde{\gamma}_0) + (\gamma_1 - \widetilde{\gamma}_1)\,\eta_{1,T+1} + v_{T+1}\end{aligned} \tag{68}$$

with a MSFE for M_2 given by:

$$\mathsf{E}\left[\widehat{\widetilde{\epsilon}}^2_{T+1|T+1}\right] = \beta_2^2\delta^2 + \sigma_\epsilon^2 + \sigma_\nu^2\left(1 + \frac{2}{T}\right) \tag{69}$$

where σ_ν^2 is given in (48).

A.7 Derivations for the equations reported in §6.3 and §6.4

Following a similar strategy as the previous analysis, including the intercept for comparability where $\widehat{\beta}_0 = \mu_y - \widehat{\beta}_1\mu_1 - \widehat{\beta}_2\mu_2$, then the forecast for M_1 is:

$$\widehat{\overline{y}}_{T+1|T+1} = \widehat{\beta}_0 + \widehat{\beta}_1\widehat{\widetilde{x}}_{1,T+1|T} + \widehat{\beta}_2\widehat{\widetilde{x}}_{2,T+1|T} = \mu_y + \widehat{\beta}_2\delta + \widehat{\beta}_1\eta_{1,T+1} + \widehat{\beta}_2\eta_{2,T+1},$$

so that the forecast error for M_1 is:

$$\begin{aligned}\widehat{\overline{\epsilon}}_{T+1|T} &= y_{T+1} - \widehat{\overline{y}}_{T+1|T} \\ &= \left(\beta_2 - \widehat{\beta}_2\right)\delta + \beta_1\Delta\eta_{1,T+1} + \beta_2\Delta\eta_{2,T+1} + \epsilon_{T+1} + \left(\beta_1 - \widehat{\beta}_1\right)\eta_{1,T} + \left(\beta_2 - \widehat{\beta}_2\right)\eta_{2,T},\end{aligned}$$

with $\mathsf{E}\left[\widehat{\overline{\epsilon}}_{T+1|T}\right] = 0$ when the parameter estimates are unbiased. The MSFE for M_1 is:

$$\mathsf{E}\left[\widehat{\overline{\epsilon}}^2_{T+1|T}\right] = 2\left(\beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2\right) + \sigma_\epsilon^2\left[1 + T^{-1}\left(2 + \frac{\delta^2}{(1-\rho^2)}\right)\right]. \tag{70}$$

Next we compute the random walk forecast for M_2 so $\gamma_1 = \beta_1 + \beta_2\rho$ and $\gamma_0 = \mu_y$, leading to the forecast given by:

$$\widehat{\widetilde{y}}_{T+1|T} = \widetilde{\gamma}_0 + \widetilde{\gamma}_1\left(x_{1,T} - \mu_1\right),$$

and the forecast error for M_2 is:

$$\begin{aligned}\widehat{\widetilde{\epsilon}}_{T+1|T} &= y_{T+1} - \widehat{\widetilde{y}}_{T+1|T} \\ &= \beta_2\delta + \beta_1\Delta\eta_{1,T+1} + \beta_2\Delta\eta_{2,T+1} + \epsilon_{T+1} + (\beta_1 - \widetilde{\gamma}_1)\,\eta_{1,T} + \beta_2\eta_{2,T},\end{aligned} \tag{71}$$

which is now biased for $\beta_2\delta \neq 0$. The MSFE for M_2 is:

$$\mathsf{E}\left[\widehat{\widetilde{\epsilon}}^2_{T+1|T}\right] = 2\beta_1^2 + \beta_2^2\left(\delta^2 + 1 + \rho^2\right) + 4\rho\beta_1\beta_2 + \sigma_\epsilon^2\left(1 + T^{-1} + T^{-2}\psi^2\right). \tag{72}$$

From (12):

$$\mathsf{MSFE}_3 = \mathsf{MSFE}_1 + (1 - \mathsf{p}_\alpha[\psi])\left[\beta_2^2\left(\delta^2 + \rho^2 - 1\right) + \sigma_\epsilon^2\left(\frac{-\delta^2}{T\left(1-\rho^2\right)} - T^{-1} + T^{-2}\psi^2\right)\right] \tag{73}$$

References

Akaike, A.. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. L. Csaki (Eds.), *Second International Symposium of Information Theory*, pp. 267–281. Budapest: Akademiai Kiado.

Bontemps, C. and G. E. Mizon. 2003. Congruence and encompassing. In B. P. Stigum (Ed.), *Econometrics and the Philosophy of Economics*, pp. 354–378. Princeton: Princeton University Press.

Breiman, L.. 1996. Bagging predictors. *Machine Learning 24*, 123–140.

Campos, J., D. F. Hendry, and H.-M. Krolzig. 2003. Consistent model selection by an automatic *Gets* approach. *Oxford Bulletin of Economics and Statistics 65*, 803–819.

Castle, J. L., M. P. Clements, and D. F. Hendry. 2015. Robust approaches to forecasting. *International Journal of Forecasting 31*, 99–112.

Castle, J. L., J. A. Doornik, and D. F. Hendry. 2012. Model selection when there are multiple breaks. *Journal of Econometrics 169(2)*, 239–246.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis. 2015. Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3(2)*, 240–264.

Castle, J. L. and N. Shephard (Eds.). 2009. *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Chu, C. S., M. Stinchcombe, and H. White. 1996. Monitoring structural change. *Econometrica 64*, 1045–1065.

Clements, M. P. and D. F. Hendry. 1998. *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P. and D. F. Hendry. 2001. Explaining the results of the M3 forecasting competition. *International Journal of Forecasting 17*, 550–554.

Doornik, J. A.. 2009. Autometrics. See Castle and Shephard (2009), pp. 88–121.

Fildes, R. and K. Ord. 2002. Forecasting competitions – their role in improving forecasting practice and research. In M. P. Clements and D. F. Hendry (Eds.), *A Companion to Economic Forecasting*, pp. 322–253. Oxford: Blackwells.

Hendry, D. F.. 1995. *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F.. 2006. Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics 135*, 399–426.

Hendry, D. F. and J. A. Doornik. 2014. *Empirical Model Discovery and Theory Evaluation*. Cambridge, Mass.: MIT Press.

Hendry, D. F., S. Johansen, and C. Santos. 2008. Automatic selection of indicators in a fully saturated regression. *Computational Statistics 33*, 317–335. Erratum, 337–339.

Hendry, D. F. and G. E. Mizon. 2012. Open-model forecast-error taxonomies. In X. Chen and N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

Ing, C-K. and C-Z. Wei. 2003. On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis 85*, 130–155.

Inoue, A. and L. Kilian. 2008. How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of the American Statistical Association 103*, 511–522.

Johansen, S. and B. Nielsen. 2009. An analysis of the indicator saturation estimator as a robust regression estimator. See Castle and Shephard (2009), pp. 1–36.

Johansen, S. and B. Nielsen. 2016. Outlier detection algorithms for least squares time series regression. *Scandinavian Journal of Statistics 43*(2), 321–348. With Discussion.

Leeb, H. and B. M. Pötscher. 2009. Model selection. In T. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 889–926. Berlin: Springer.

Makridakis, S. and M. Hibon. 2000. The M3-competition: Results, conclusions and implications. *International Journal of Forecasting 16*, 451–476.

Pötscher, B. M.. 1991. Effects of model selection on inference. *Econometric Theory 7*, 163–185.

Shibata, R.. 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics 8*, 147–164.

Stock, James and Mark W. Watson. 2009. Phillips curve inflation forecasts. In J. Fuhrer, Y. Kodrzycki, J. Little, and G. Olivei (Eds.), *Understanding Inflation and the Implications for Monetary Policy*, pp. 99–202. Cambridge: MIT Press.

B Supplementary Tables

Table 9: $T = 100$, Ratio M_2 reports $\frac{\mathsf{MSFE}_2}{\mathsf{MSFE}_1}$ and Ratio M_3 reports $\frac{\mathsf{MSFE}_3}{\mathsf{MSFE}_1}$

Case		Model	$\psi^2 = 0$	$\psi^2 = 1$	$\psi^2 = 4$	$\psi^2 = 9$	$\psi^2 = 16$
Stationary	Ratio M_2		0.990	1.000	1.030	1.079	1.149
($\delta^2 = 0$)	Ratio M_3	$\alpha = 0.001$	0.990	1.000	1.026	1.048	1.035
		$\alpha = 0.05$	0.991	1.000	1.014	1.012	1.003
		$\alpha = 0.16$	0.992	1.000	1.008	1.004	1.001
out-of-sample shift	Ratio M_2		0.827	1.008	1.551	2.457	3.724
known future	Ratio M_3	$\alpha = 0.001$	0.827	1.008	1.497	1.895	1.651
regressors		$\alpha = 0.05$	0.836	1.007	1.267	1.217	1.056
		$\alpha = 0.16$	0.855	1.005	1.152	1.081	1.013
out-of-sample shift	Ratio M_2		1.000	1.000	1.000	1.000	1.000
unknown future	Ratio M_3	$\alpha = 0.001$	1.000	1.000	1.000	1.000	1.000
regressors		$\alpha = 0.05$	1.000	1.000	1.000	1.000	1.000
mean forecast		$\alpha = 0.16$	1.000	1.000	1.000	1.000	1.000
out-of-sample shift	Ratio M_2		0.997	1.002	1.013	1.024	1.033
unknown future	Ratio M_3	$\alpha = 0.001$	0.997	1.002	1.012	1.015	1.008
regressors		$\alpha = 0.05$	0.997	1.002	1.006	1.004	1.001
random walk forecast		$\alpha = 0.16$	0.997	1.001	1.004	1.001	1.000
in-sample shift	Ratio M_2		1.010	1.009	1.008	1.007	1.007
unknown future	Ratio M_3	$\alpha = 0.001$	1.010	1.009	1.008	1.005	1.002
regressors		$\alpha = 0.05$	1.010	1.008	1.004	1.001	1.000
mean forecast		$\alpha = 0.16$	1.008	1.006	1.002	1.000	1.000
in-sample shift	Ratio M_2		0.931	0.994	1.155	1.386	1.661
unknown future	Ratio M_3	$\alpha = 0.001$	0.931	0.994	1.140	1.237	1.158
regressors		$\alpha = 0.05$	0.934	0.995	1.075	1.058	1.014
random walk forecast		$\alpha = 0.16$	0.942	0.996	1.043	1.021	1.003

Table 10: Forecasts for $T+1|T$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0, 0, 0, 0, 0, 0, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4)$. $N = 15$, so there are $n = 8$ relevant regressors and $N - n = 7$ irrelevant regressors. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Intercept forced and not included in measure of potency. Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast						(v) AR(1) forecast			(vi) direct forecast		(vii) Pool	selection	
											(a) RW			(b) RW with diff										
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(1)	Potency	Gauge
	No break																							
	0.1	1.519	1.128	1.528	1.758	1.590	1.678	2.359	1.474	1.617	2.655	1.694	1.755	2.816	2.118	2.015	5.368	1.438	1.720	1.848	1.506	1.541	0.346	0.012
	1			1.388			1.626	2.366	1.484	1.579			1.782			2.164			2.116			1.513	0.487	0.028
	5			1.381			1.634	2.374	1.518	1.635			1.822			2.258			2.797			*1.521*	0.604	0.073
	10			1.378			1.630	2.381	1.550	1.650			1.813			2.255			3.342			**1.506**	0.645	0.116
	16			1.397			1.640	2.380	1.577	1.704			1.840			2.295			3.807			*1.521*	0.681	0.164
	32			1.430			1.666	2.413	1.650	1.856			1.806			2.266			4.564			1.530	0.760	0.296
	50			1.481			1.675	2.428	1.674	1.906			1.836			2.324			5.009			1.533	0.829	0.435
	Break in relevant regressors																							
ι	0.1	2.110	1.545	6.748	17.678	17.561	17.662	18.292	17.649	17.702	18.451	18.140	18.083	18.525	18.607	18.396	20.730	17.632	18.074	18.371	17.651	17.723	0.344	0.012
	1			3.631			17.653	18.309	17.661	17.652			18.224			18.690			18.869			17.726	0.484	0.028
	5			2.553			17.585	18.330	17.755	17.763			18.303			18.774			19.736			17.751	0.599	0.074
	10			2.321			**17.534**	18.477	17.847	17.713			18.383			18.861			20.355			17.737	0.643	0.116
	16			2.217			17.541	18.477	17.948	17.898			18.421			18.956			20.906			17.789	0.680	0.164
	32			2.120			17.567	18.474	18.027	18.019			18.395			18.936			21.838			17.763	0.759	0.294
	50			2.085			*17.558*	18.465	18.165	18.340			18.441			19.004			22.174			17.823	0.825	0.437
λ	0.1	1.638	1.245	2.598	4.669	4.455	4.534	4.806	4.124	4.354	5.244	4.103	4.366	5.357	4.379	4.554	7.669	4.083	4.494	4.502	4.294	4.253	0.344	0.012
	1			1.879			4.520	4.817	4.125	4.265			4.302			4.588			5.046			4.176	0.484	0.028
	5			1.667			4.507	4.822	4.149	4.279			4.311			4.622			5.826			4.157	0.601	0.074
	10			1.589			4.484	4.864	4.189	4.249			4.309			4.618			6.238			**4.130**	0.645	0.116
	16			1.592			4.502	4.863	4.234	4.347			4.340			4.664			6.799			*4.160*	0.681	0.163
	32			1.598			4.544	4.877	4.308	4.515			4.342			4.691			7.628			4.183	0.760	0.296
	50			1.613			4.553	4.875	4.338	4.587			4.355			4.715			8.058			4.180	0.828	0.436
	Break in irrelevant regressors																							
ι	0.1	2.130	1.128	1.566	1.754	1.587	1.674	2.332	1.505	1.619	2.647	1.691	1.743	2.808	2.112	2.001	5.353	1.436	1.745	1.847	1.504	1.535	0.344	0.012
	1			1.443			1.624	2.346	1.493	1.584			1.779			2.165			2.157			1.508	0.482	0.029
	5			1.548			1.630	2.359	1.582	1.687			1.825			2.254			2.827			*1.524*	0.599	0.073
	10			1.605			1.626	2.387	1.634	1.728			1.813			2.265			3.257			**1.504**	0.643	0.115
	16			1.697			1.635	2.368	1.703	1.842			1.853			2.313			3.812			1.536	0.681	0.163
	32			1.859			1.663	2.425	1.811	2.052			1.806			2.261			4.560			1.541	0.761	0.295
	50			2.000			1.672	2.469	1.899	2.217			1.836			2.323			5.017			1.557	0.826	0.436
λ	0.1	1.653	1.128	1.539	1.758	1.590	1.674	2.346	1.487	1.620	2.653	1.694	1.748	2.814	2.117	2.009	5.367	1.438	1.746	1.848	1.506	1.539	0.344	0.012
	1			1.408			1.627	2.356	1.490	1.586			1.781			2.164			2.158			1.515	0.485	0.029
	5			1.439			1.634	2.374	1.544	1.648			1.820			2.247			2.798			*1.523*	0.602	0.073
	10			1.433			1.631	2.390	1.573	1.665			1.817			2.267			3.331			**1.506**	0.645	0.115
	16			1.491			1.639	2.384	1.615	1.740			1.849			2.308			3.866			1.528	0.682	0.164
	32			1.534			1.667	2.424	1.690	1.901			1.803			2.259			4.519			1.531	0.760	0.295
	50			1.605			1.677	2.449	1.736	1.991			1.838			2.328			5.068			1.541	0.828	0.436
	Break in all regressors																							
ι	0.1	2.707	1.545	6.843	17.993	17.878	17.975	18.595	17.931	17.956	18.775	18.404	18.383	18.847	18.860	18.706	21.057	17.922	18.316	18.656	17.957	18.002	0.345	0.012
	1			3.685			17.937	18.622	17.917	*17.879*			18.432			18.903			19.190			17.965	0.486	0.029
	5			2.726			17.896	18.630	18.094	18.041			18.552			18.994			19.997			18.015	0.603	0.073
	10			2.554			**17.854**	18.827	18.241	18.054			18.637			19.097			20.589			18.011	0.644	0.116
	16			2.502			17.860	18.770	18.287	18.126			18.628			19.124			21.325			17.999	0.681	0.163
	32			2.574			17.891	18.755	18.473	18.360			18.649			19.181			22.209			18.011	0.761	0.296
	50			2.616			17.885	18.762	18.578	18.726			18.710			19.256			22.603			18.069	0.828	0.436
λ	0.1	1.765	1.245	2.612	4.718	4.503	4.579	4.848	4.166	4.388	5.292	4.130	4.396	5.405	4.401	4.583	7.720	4.120	4.541	4.536	4.337	4.287	0.344	0.012
	1			1.899			4.565	4.866	4.155	4.292			4.324			4.594			5.066			4.206	0.485	0.028
	5			1.713			4.550	4.876	4.201	4.319			4.331			4.629			5.833			*4.189*	0.603	0.074
	10			1.668			4.535	4.929	4.236	4.305			4.347			4.654			6.268			**4.169**	0.644	0.116
	16			1.674			4.551	4.921	4.280	4.375			4.355			4.677			6.905			4.179	0.680	0.164
	32			1.693			4.597	4.935	4.355	4.549			4.358			4.700			7.674			4.199	0.760	0.295
	50			1.735			4.603	4.941	4.386	4.644			4.388			4.739			8.059			4.206	0.829	0.435

Table 11: Forecasts for $T+2|T+1$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0, 0, 0, 0, 0, 0, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4)$. $N = 15$, so there are $n = 8$ relevant regressors and $N - n = 7$ irrelevant regressors. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Intercept forced and not included in measure of potency. Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast (a) RW			(b) RW with diff			(v) AR(1) forecast			(vi) direct forecast		(vii) Pool	selection	
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(1)	Potency	Gauge
	No break																							
	0.1	1.497	1.090	1.392	1.691	1.573	1.629	2.300	1.474	1.532	2.911	1.683	1.723	3.081	2.018	1.984	5.493	1.450	1.753	1.842	1.501	1.503	0.344	0.012
	1			1.313			1.582	2.302	1.487	1.541			1.733			2.027			2.125			**1.491**	0.481	0.027
	5			1.296			1.590	2.294	1.521	1.600			1.759			2.112			2.684			1.495	0.601	0.074
	10			1.311			1.594	2.291	1.568	1.666			1.769			2.135			2.904			*1.502*	0.645	0.114
	16			1.362			1.620	2.325	1.604	1.741			1.802			2.179			3.459			1.526	0.684	0.163
	32			1.440			1.634	2.321	1.671	1.853			1.816			2.225			4.221			1.536	0.765	0.297
	50			1.484			1.642	2.317	1.646	1.878			1.843			2.252			4.677			1.532	0.827	0.437
	Break in relevant regressors																							
ι	0.1	2.518	1.661	7.683	43.373	39.504	40.018	12.516	19.880	25.317	43.421	6.423	14.839	43.320	2.531	8.732	18.419	19.386	25.675	18.350	31.740	23.049	0.400	0.013
	1			4.148			40.163	12.533	18.716	21.753			11.314			5.321			23.821			20.332	0.516	0.027
	5			2.787			40.398	12.551	16.566	17.910			9.213			3.911			22.812			17.986	0.622	0.074
	10			2.440			40.781	12.691	15.438	16.215			8.755			3.597			22.471			17.144	0.665	0.113
	16			2.446			41.280	12.703	14.415	14.813			8.624			*3.464*			22.407			16.539	0.695	0.163
	32			2.369			41.986	12.701	12.786	12.595			8.303			3.298			22.474			15.452	0.771	0.293
	50			2.416			42.422	12.720	11.921	11.458			8.173			**3.177**			22.502			14.888	0.831	0.438
λ	0.1	1.846	1.331	3.997	12.918	11.963	11.924	6.356	7.394	8.893	13.246	3.986	6.793	13.339	2.426	5.422	10.006	7.153	9.186	7.161	10.025	8.351	0.357	0.013
	1			2.513			11.983	6.351	7.278	8.106			5.727			4.160			9.050			7.661	0.486	0.028
	5			1.918			12.117	6.342	6.899	7.305			5.005			3.399			9.225			7.066	0.607	0.075
	10			1.793			12.236	6.340	6.797	7.056			4.816			3.198			9.318			6.884	0.648	0.114
	16			1.776			12.346	6.393	6.726	6.880			4.724			*3.105*			9.599			6.760	0.689	0.161
	32			1.813			12.558	6.379	6.421	6.483			4.637			3.009			10.403			6.540	0.766	0.295
	50			1.809			12.661	6.410	6.339	6.286			4.609			**2.958**			10.548			6.419	0.830	0.438
	Break in irrelevant regressors																							
ι	0.1	2.429	1.090	1.449	1.689	1.569	1.601	4.075	1.493	1.570	3.254	1.677	1.732	3.853	2.011	1.992	9.067	1.445	1.721	1.835	1.496	**1.507**	0.341	0.012
	1			1.512			1.571	4.118	1.546	1.688			1.742			2.072			2.225			1.517	0.477	0.026
	5			1.611			1.589	4.200	1.697	1.878			1.807			2.211			2.704			*1.538*	0.597	0.073
	10			1.696			1.587	4.305	1.823	2.057			1.817			2.238			3.090			1.564	0.641	0.112
	16			1.841			1.604	4.336	1.918	2.203			1.831			2.303			3.576			1.575	0.682	0.163
	32			2.108			1.623	4.429	2.042	2.500			1.892			2.463			4.499			1.613	0.765	0.294
	50			2.296			1.640	4.518	2.204	2.743			1.937			2.571			5.058			1.653	0.826	0.435
λ	0.1	1.783	1.090	1.416	1.692	1.573	1.628	2.768	1.470	1.554	3.112	1.682	1.714	3.411	2.017	1.969	6.356	1.449	1.693	1.841	1.500	**1.504**	0.342	0.012
	1			1.395			1.579	2.779	1.512	1.591			1.745			2.045			2.154			1.507	0.477	0.027
	5			1.389			1.589	2.805	1.586	1.676			1.763			2.114			2.682			1.507	0.600	0.073
	10			1.447			1.597	2.863	1.662	1.791			1.797			2.166			3.037			*1.531*	0.647	0.114
	16			1.562			1.613	2.879	1.710	1.894			1.804			2.197			3.494			1.550	0.684	0.164
	32			1.639			1.630	2.918	1.819	2.059			1.811			2.230			4.379			1.565	0.765	0.296
	50			1.740			1.640	2.938	1.891	2.169			1.850			2.288			4.781			1.581	0.827	0.435
	Break in all regressors																							
ι	0.1	3.269	1.662	7.546	43.937	40.014	40.392	7.302	19.097	24.667	44.548	6.439	14.575	44.941	2.534	8.448	9.857	19.580	25.658	18.506	32.094	22.820	0.402	0.013
	1			4.190			40.684	7.335	17.558	20.883			11.236			5.293			23.933			20.062	0.514	0.027
	5			2.843			40.926	7.482	14.542	16.130			9.014			3.755			22.638			17.297	0.622	0.073
	10			2.710			41.378	7.548	12.821	13.877			8.734			3.612			22.479			16.245	0.665	0.113
	16			2.675			41.690	7.741	11.586	12.398			8.503			*3.458*			22.911			15.496	0.692	0.163
	32			2.906			42.439	7.806	9.664	10.130			8.335			**3.421**			23.272			14.325	0.772	0.293
	50			3.077			42.958	7.910	8.747	8.893			8.337			3.428			22.711			13.646	0.834	0.438
λ	0.1	2.047	1.331	3.988	13.031	12.064	12.054	5.226	7.300	8.818	13.665	3.991	6.796	13.940	2.423	5.417	8.149	7.195	9.255	7.191	10.095	8.346	0.356	0.013
	1			2.497			12.097	5.218	7.117	8.012			5.698			4.111			8.977			7.634	0.487	0.027
	5			1.921			12.230	5.232	6.594	7.056			4.990			3.369			9.115			6.987	0.607	0.075
	10			1.863			12.305	5.288	6.348	6.693			4.914			3.292			9.228			6.793	0.648	0.114
	16			1.891			12.445	5.344	6.237	6.488			4.781			*3.153*			9.646			6.644	0.684	0.163
	32			1.920			12.649	5.385	5.851	5.898			4.670			3.027			10.526			6.334	0.765	0.296
	50			1.986			12.747	5.416	5.661	5.705			4.648			**3.011**			10.695			6.201	0.829	0.435

Table 12: Forecasts for $T+1|T$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0,0,0,0,0,0,0,0,0,0,4,4,4,4,4)$. $N = 15$, so there are $n = 5$ relevant regressors and $N-n = 10$ irrelevant regressors. Intercept forced and not included in measure of potency. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Pool (2) given by an equally weighted average of (iv,a) Robust RW and (iii) in-sample forecast for X. Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast						(v) AR(1) forecast			(vi) direct forecast		(vii) Pool		selection	
											(a) RW			(b) RW with diff											
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(2)	(1)	Potency	Gauge
	No break																								
	0.1	1.514	1.093	1.363	2.142	1.904	1.940	3.134	1.706	1.787	3.445	2.019	2.106	3.613	2.589	2.621	6.228	1.670	1.935	2.179	1.794	1.853	1.760	0.753	0.008
	1			1.244			1.941	3.137	1.723	1.783			2.070			2.643			2.225			1.819	1.733	0.919	0.020
	5			1.241			1.956	3.144	1.764	1.843			2.068			2.680			3.164			1.827	**1.727**	0.969	0.068
	10			1.293			1.966	3.144	1.808	1.914			2.131			2.751			3.624			1.872	*1.755*	0.973	0.115
	16			1.330			1.971	3.140	1.845	1.963			2.151			2.765			4.058			1.897	1.765	0.975	0.165
	32			1.411			2.006	3.172	1.927	2.154			2.165			2.794			4.984			1.952	1.797	0.990	0.292
	50			1.455			2.043	3.193	1.978	2.222			2.199			2.851			5.577			1.974	1.803	0.993	0.440
	Break in relevant regressors																								
ι	0.1	1.883	1.321	3.761	21.348	21.222	21.301	22.152	21.303	21.371	22.491	22.016	22.146	22.556	22.599	22.727	25.568	21.312	21.739	22.197	21.379	21.663	21.495	0.744	0.009
	1			1.861			21.215	22.165	21.322	21.366			22.053			22.656			22.251			21.599	21.445	0.914	0.020
	5			1.616			**21.204**	22.178	21.481	21.588			22.150			22.770			23.577			21.723	21.514	0.965	0.068
	10			1.640			**21.204**	22.292	21.549	21.683			22.152			22.771			24.091			21.740	21.521	0.971	0.115
	16			1.692			*21.219*	22.299	21.695	21.837			22.173			22.780			24.692			21.801	21.556	0.974	0.164
	32			1.733			21.237	22.268	21.723	22.018			22.220			22.851			25.808			21.839	21.570	0.989	0.293
	50			1.799			21.245	22.237	21.776	22.143			22.343			23.017			26.260			21.907	21.600	0.993	0.437
λ	0.1	1.601	1.156	1.887	5.763	5.457	5.499	6.150	4.910	5.081	6.712	4.852	5.139	6.833	5.172	5.457	9.405	4.877	5.309	5.370	5.209	5.018	5.010	0.749	0.008
	1			1.363			5.482	6.156	4.917	4.990			4.958			5.302			5.599			**4.867**	4.903	0.918	0.020
	5			1.335			5.516	6.161	4.971	5.087			4.980			5.348			6.661			*4.904*	4.915	0.968	0.068
	10			1.379			5.531	6.196	5.021	5.161			5.037			5.416			7.080			4.947	4.940	0.972	0.114
	16			1.398			5.542	6.197	5.073	5.244			5.065			5.443			7.646			4.988	4.963	0.976	0.164
	32			1.471			5.591	6.204	5.152	5.428			5.098			5.499			8.618			5.048	4.997	0.990	0.293
	50			1.534			5.631	6.201	5.166	5.474			5.163			5.587			9.172			5.074	5.003	0.993	0.438
	Break in irrelevant regressors																								
ι	0.1	2.358	1.093	1.451	2.128	1.893	1.943	3.080	1.780	1.862	3.419	2.006	2.086	3.587	2.566	2.583	6.192	1.663	1.939	2.169	1.786	1.872	1.767	0.745	0.008
	1			1.329			1.931	3.090	1.759	1.820			2.052			2.621			2.239			1.817	**1.724**	0.914	0.020
	5			1.451			1.946	3.096	1.870	1.972			2.062			2.670			3.184			1.862	1.736	0.966	0.068
	10			1.578			1.956	3.120	1.957	2.091			2.130			2.748			3.641			1.915	*1.761*	0.970	0.114
	16			1.700			1.960	3.107	2.052	2.220			2.141			2.754			4.042			1.948	1.777	0.974	0.165
	32			1.981			1.991	3.179	2.265	2.547			2.160			2.784			4.959			2.028	1.817	0.989	0.293
	50			2.145			2.029	3.258	2.418	2.811			2.185			2.824			5.567			2.101	1.849	0.993	0.439
λ	0.1	1.685	1.093	1.387	2.141	1.903	1.937	3.112	1.733	1.818	3.441	2.017	2.108	3.608	2.585	2.618	6.224	1.670	1.948	2.177	1.794	1.870	1.769	0.754	0.008
	1			1.274			1.940	3.117	1.736	1.797			2.065			2.637			2.224			1.823	**1.733**	0.918	0.020
	5			1.296			1.957	3.131	1.801	1.887			2.066			2.675			3.166			1.845	1.735	0.969	0.068
	10			1.359			1.966	3.138	1.855	1.961			2.136			2.762			3.615			1.891	*1.760*	0.972	0.115
	16			1.418			1.966	3.131	1.918	2.047			2.149			2.767			4.039			1.926	1.775	0.975	0.165
	32			1.527			2.005	3.169	2.016	2.248			2.164			2.792			4.981			1.975	1.800	0.990	0.293
	50			1.604			2.041	3.207	2.109	2.396			2.207			2.860			5.575			2.031	1.827	0.993	0.439
	Break in all regressors																								
ι	0.1	2.695	1.321	3.734	22.011	21.883	21.957	22.797	21.878	21.951	23.174	22.579	22.687	23.236	23.142	23.233	26.269	21.923	22.355	22.796	22.016	22.211	22.066	0.753	0.008
	1			1.883			**21.873**	22.815	21.879	21.900			22.608			23.216			22.816			22.127	22.003	0.920	0.020
	5			1.793			21.876	22.815	22.160	22.245			22.703			23.304			24.199			22.297	22.096	0.968	0.068
	10			1.911			21.876	23.012	22.362	22.416			22.741			23.340			24.596			22.346	22.116	0.972	0.115
	16			2.025			*21.890*	22.957	22.479	22.628			22.738			23.318			25.399			22.399	22.144	0.975	0.165
	32			2.236			21.895	22.917	22.554	22.824			22.804			23.426			26.377			22.409	22.132	0.990	0.293
	50			2.451			21.915	22.928	22.681	23.024			22.892			23.549			26.941			22.453	22.138	0.993	0.440
λ	0.1	1.756	1.156	1.904	5.858	5.548	5.594	6.230	4.977	5.145	6.808	4.901	5.180	6.928	5.213	5.487	9.503	4.947	5.343	5.433	5.289	5.069	5.068	0.751	0.008
	1			1.400			5.578	6.242	4.968	5.046			5.015			5.362			5.681			**4.923**	*4.964*	0.918	0.020
	5			1.386			5.605	6.250	5.045	5.162			5.031			5.395			6.732			4.963	4.976	0.967	0.068
	10			1.438			5.623	6.297	5.105	5.239			5.085			5.453			7.170			5.001	4.997	0.972	0.115
	16			1.486			5.638	6.287	5.163	5.336			5.105			5.470			7.708			5.040	5.015	0.975	0.165
	32			1.580			5.682	6.297	5.214	5.496			5.158			5.550			8.653			5.090	5.040	0.990	0.293
	50			1.665			5.725	6.308	5.255	5.579			5.220			5.637			9.310			5.122	5.047	0.993	0.440

Table 13: Forecasts for $T+2|T+1$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0,0,0,0,0,0,0,0,0,0,4,4,4,4,4)$. $N = 15$, so there are $n = 5$ relevant regressors and $N-n = 10$ irrelevant regressors. Intercept forced and not included in measure of potency. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast (a) RW			((iv) robust forecast (b) RW with diff			(v) AR(1) forecast			(vi) direct forecast		(vii) Pool	selection	
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(1)	Potency	Gauge
	No break																							
	0.1	1.495	1.056	1.302	2.078	1.893	1.918	3.002	1.747	1.793	3.730	2.040	2.073	3.898	2.512	2.530	6.381	1.706	1.940	2.180	1.799	*1.759*	0.756	0.008
	1			1.176			1.909	3.004	1.758	1.790			2.078			2.587			2.254			**1.741**	0.920	0.020
	5			1.190			1.931	2.990	1.812	1.882			2.086			2.617			2.871			1.756	0.969	0.067
	10			1.225			1.938	2.992	1.875	1.966			2.103			2.637			3.455			1.775	0.974	0.114
	16			1.295			1.949	3.028	1.912	2.043			2.168			2.707			3.883			1.805	0.978	0.164
	32			1.395			1.974	3.028	1.987	2.221			2.193			2.767			4.668			1.833	0.988	0.293
	50			1.436			2.003	3.029	1.966	2.239			2.201			2.768			5.175			1.819	0.994	0.440
	Break in relevant regressors																							
ι	0.1	2.159	1.431	3.984	52.259	47.182	47.823	27.793	24.191	26.788	53.270	7.632	11.630	53.390	2.918	5.653	31.625	23.137	26.309	22.385	36.993	23.294	0.791	0.009
	1			2.056			47.937	27.823	23.269	23.928			8.960			3.661			24.724			21.100	0.925	0.020
	5			1.779			48.423	27.822	21.339	21.640			8.555			3.367			25.248			20.129	0.968	0.068
	10			1.764			48.996	27.870	20.207	20.288			8.669			**3.344**			25.741			19.727	0.975	0.115
	16			1.856			49.522	27.854	19.308	19.344			8.921			*3.415*			25.865			19.506	0.979	0.162
	32			1.975			50.351	27.864	17.909	17.561			9.155			3.473			26.920			18.934	0.990	0.293
	50			2.057			51.027	27.758	17.125	16.514			9.526			3.636			27.131			18.660	0.994	0.441
λ	0.1	1.720	1.193	2.463	15.844	14.523	14.583	10.698	8.939	9.706	16.555	4.756	6.176	16.701	2.871	4.398	14.213	8.601	9.636	8.676	11.897	8.838	0.763	0.008
	1			1.493			14.683	10.706	8.852	9.061			5.203			3.325			9.534			8.218	0.917	0.020
	5			1.434			14.803	10.690	8.532	8.720			5.094			3.199			10.169			8.041	0.968	0.068
	10			1.452			14.955	10.671	8.440	8.612			5.117			**3.186**			10.785			8.001	0.976	0.115
	16			1.513			15.108	10.692	8.365	8.525			5.227			*3.281*			11.133			7.989	0.976	0.162
	32			1.606			15.312	10.690	8.168	8.282			5.281			3.339			11.945			7.889	0.990	0.293
	50			1.653			15.496	10.685	8.033	8.046			5.401			3.443			12.561			7.819	0.993	0.440
	Break in irrelevant regressors																							
ι	0.1	2.773	1.056	1.439	2.068	1.879	1.886	12.058	1.797	1.851	4.260	2.021	2.095	5.080	2.489	2.582	15.899	1.692	1.971	2.163	1.784	**1.760**	0.745	0.009
	1			1.416			1.893	12.140	1.861	1.985			2.091			2.645			2.248			1.766	0.914	0.020
	5			1.665			1.916	12.280	2.125	2.406			2.117			2.734			3.005			*1.821*	0.964	0.066
	10			1.855			1.928	12.402	2.362	2.730			2.174			2.827			3.534			1.880	0.970	0.113
	16			2.010			1.936	12.474	2.542	3.015			2.223			2.903			4.050			1.937	0.976	0.163
	32			2.344			1.968	12.665	2.848	3.530			2.339			3.173			5.180			2.033	0.989	0.292
	50			2.618			1.990	12.800	3.034	3.843			2.343			3.192			5.805			2.071	0.993	0.439
λ	0.1	1.901	1.056	1.353	2.079	1.891	1.909	5.144	1.749	1.793	4.019	2.037	2.087	4.378	2.509	2.558	8.512	1.703	1.906	2.177	1.796	1.755	0.753	0.008
	1			1.230			1.907	5.169	1.789	1.842			2.076			2.584			2.249			**1.748**	0.918	0.020
	5			1.352			1.929	5.220	1.933	2.039			2.095			2.622			2.925			*1.787*	0.968	0.068
	10			1.430			1.942	5.291	2.054	2.204			2.098			2.608			3.551			1.813	0.971	0.116
	16			1.480			1.953	5.294	2.141	2.345			2.127			2.655			3.840			1.840	0.976	0.164
	32			1.703			1.979	5.410	2.286	2.581			2.213			2.829			4.848			1.890	0.987	0.293
	50			1.804			2.001	5.442	2.386	2.743			2.200			2.783			5.354			1.898	0.995	0.439
	Break in all regressors																							
ι	0.1	3.272	1.431	3.749	53.481	48.274	48.843	8.715	23.024	25.524	55.694	7.683	11.361	56.781	2.936	5.373	11.710	23.567	26.460	22.753	37.726	22.931	0.801	0.009
	1			2.219			49.088	8.796	21.222	21.992			8.892			3.661			24.961			20.561	0.928	0.020
	5			2.038			49.577	8.984	17.588	17.946			8.557			3.432			25.583			18.921	0.972	0.067
	10			2.218			50.060	9.066	15.503	16.059			8.670			**3.399**			26.295			18.191	0.976	0.116
	16			2.409			50.645	9.308	14.117	14.640			8.969			*3.590*			26.302			17.703	0.979	0.162
	32			2.749			51.666	9.408	11.808	12.057			9.510			3.797			27.000			16.738	0.990	0.293
	50			3.059			52.259	9.526	10.705	10.864			9.848			3.984			27.395			16.214	0.995	0.439
λ	0.1	2.046	1.193	2.504	16.058	14.712	14.795	6.917	8.755	9.555	17.222	4.767	6.168	17.653	2.868	4.395	10.126	8.680	9.695	8.733	12.023	8.815	0.765	0.009
	1			1.550			14.867	6.909	8.569	8.805			5.157			*3.275*			9.468			8.142	0.919	0.020
	5			1.522			15.014	6.939	7.978	8.183			5.084			**3.171**			10.330			7.875	0.969	0.067
	10			1.575			15.132	6.998	7.681	7.933			5.136			3.182			10.865			7.770	0.974	0.115
	16			1.664			15.284	7.060	7.581	7.834			5.270			3.295			11.210			7.769	0.978	0.165
	32			1.828			15.572	7.136	7.061	7.192			5.358			3.416			12.159			7.505	0.990	0.294
	50			1.927			15.699	7.177	6.837	6.950			5.423			3.447			12.463			7.392	0.994	0.441

Table 14: Forecasts for $T+1|T$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1)$. $N = 15$, so there are $n = 5$ relevant regressors and $N-n = 10$ irrelevant regressors. Intercept forced and not included in measure of potency. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast (a) RW			(b) RW with diff			(v) AR(1) forecast			(vi) direct forecast		(vii) Pool	selection	
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(1)	Potency	Gauge
	No break																							
	0.1	1.532	1.098	1.126	1.142	1.058	1.119	1.648	1.059	1.126	1.767	1.128	1.130	1.962	1.194	1.150	2.958	1.059	1.151	1.387	1.060	1.094	0.177	0.012
	1			1.131			1.107	1.654	1.065	1.123			1.157			1.184			1.327			1.095	0.209	0.021
	5			1.240			1.088	1.665	1.076	1.177			1.220			1.288			1.731			1.111	0.294	0.069
	10			1.294			**1.078**	1.677	1.081	1.218			1.239			1.337			1.956			1.117	0.352	0.112
	16			1.307			1.088	1.671	1.091	1.248			1.257			1.373			2.200			1.121	0.409	0.160
	32			1.406			*1.093*	1.700	1.107	1.312			1.294			1.443			2.585			1.129	0.540	0.291
	50			1.470			*1.093*	1.720	1.112	1.381			1.312			1.489			2.847			1.141	0.650	0.438
	Break in relevant regressors																							
ι	0.1	1.899	1.316	2.293	2.407	2.320	2.372	2.883	2.346	2.399	2.992	2.456	2.419	3.162	2.525	2.453	4.192	2.351	2.443	2.752	2.334	2.373	0.177	0.012
	1			2.191			2.373	2.896	2.350	2.390			2.451			2.484			2.682			2.373	0.208	0.021
	5			2.023			*2.342*	2.908	2.378	2.438			2.540			2.615			3.114			2.393	0.293	0.069
	10			1.986			**2.332**	2.956	2.380	2.481			2.572			2.670			3.441			2.405	0.348	0.111
	16			1.890			2.338	2.948	2.405	2.535			2.606			2.725			3.574			2.418	0.406	0.161
	32			1.848			2.343	2.975	2.424	2.594			2.630			2.779			4.039			2.419	0.535	0.291
	50			1.864			2.343	2.999	2.432	2.736			2.662			2.844			4.360			2.450	0.649	0.438
λ	0.1	1.617	1.163	1.359	1.394	1.302	1.360	1.850	1.282	1.373	1.978	1.333	1.380	2.162	1.383	1.404	3.156	1.283	1.406	1.633	1.304	1.341	0.176	0.013
	1			1.347			1.351	1.858	1.287	1.361			1.403			1.433			1.593			1.338	0.209	0.021
	5			1.424			*1.331*	1.870	1.302	1.408			1.466			1.535			2.004			1.350	0.293	0.069
	10			1.476			**1.323**	1.894	1.304	1.440			1.484			1.579			2.281			1.354	0.351	0.112
	16			1.450			1.330	1.886	1.319	1.484			1.492			1.602			2.459			1.358	0.408	0.160
	32			1.536			1.336	1.912	1.337	1.539			1.528			1.670			2.885			1.364	0.537	0.291
	50			1.550			1.338	1.933	1.342	1.622			1.545			1.711			3.191			1.379	0.649	0.438
	Break in irrelevant regressors																							
ι	0.1	2.398	1.098	1.188	1.141	1.057	1.119	1.623	1.075	1.126	1.765	1.128	1.130	1.960	1.193	1.149	2.953	1.059	1.149	1.387	1.059	1.093	0.176	0.012
	1			1.194			1.107	1.639	1.075	1.133			1.162			1.190			1.330			1.099	0.208	0.021
	5			1.414			*1.088*	1.650	1.092	1.190			1.221			1.285			1.752			1.113	0.294	0.069
	10			1.537			**1.080**	1.683	1.096	1.263			1.230			1.330			1.962			1.123	0.349	0.111
	16			1.658			1.087	1.678	1.107	1.290			1.256			1.370			2.196			1.125	0.409	0.160
	32			1.979			1.093	1.752	1.144	1.409			1.282			1.423			2.591			1.136	0.538	0.291
	50			2.213			1.093	1.815	1.170	1.560			1.310			1.484			2.873			1.160	0.647	0.436
λ	0.1	1.711	1.098	1.142	1.142	1.058	1.120	1.634	1.066	1.127	1.766	1.128	1.130	1.962	1.194	1.149	2.957	1.059	1.151	1.387	1.060	1.094	0.177	0.012
	1			1.141			1.107	1.644	1.069	1.128			1.161			1.189			1.327			1.098	0.210	0.021
	5			1.284			1.088	1.660	1.084	1.184			1.221			1.286			1.726			1.113	0.295	0.069
	10			1.345			**1.078**	1.679	1.089	1.228			1.238			1.339			1.952			1.119	0.351	0.111
	16			1.379			1.088	1.671	1.100	1.257			1.261			1.378			2.210			1.124	0.411	0.160
	32			1.532			1.094	1.713	1.121	1.330			1.287			1.433			2.590			1.129	0.539	0.291
	50			1.631			*1.093*	1.743	1.135	1.421			1.305			1.478			2.878			1.143	0.649	0.437
	Break in all regressors																							
ι	0.1	2.729	1.316	2.364	2.449	2.362	2.418	2.902	2.396	2.445	3.035	2.492	2.459	3.204	2.559	2.493	4.239	2.390	2.491	2.791	2.376	2.415	0.177	0.012
	1			2.253			2.415	2.923	2.392	2.441			2.502			2.538			2.726			2.420	0.209	0.021
	5			2.228			*2.383*	2.932	2.427	2.483			2.584			2.658			3.143			2.434	0.294	0.069
	10			2.220			**2.370**	3.015	2.435	2.543			2.604			2.708			3.433			2.443	0.351	0.112
	16			2.150			2.376	2.993	2.459	2.585			2.624			2.737			3.664			2.445	0.410	0.160
	32			2.397			2.384	3.045	2.486	2.676			2.694			2.858			4.098			2.458	0.539	0.291
	50			2.567			2.386	3.107	2.503	2.884			2.696			2.874			4.386			2.484	0.650	0.438
λ	0.1	1.781	1.163	1.375	1.400	1.307	1.367	1.844	1.292	1.380	1.984	1.336	1.384	2.168	1.386	1.409	3.163	1.288	1.411	1.637	1.310	1.346	0.177	0.012
	1			1.369			1.359	1.857	1.293	1.374			1.408			1.436			1.603			1.345	0.209	0.021
	5			1.467			*1.338*	1.872	1.311	1.414			1.465			1.531			2.007			1.354	0.293	0.069
	10			1.511			**1.328**	1.904	1.314	1.459			1.481			1.574			2.274			1.359	0.351	0.111
	16			1.494			1.337	1.892	1.329	1.487			1.489			1.594			2.485			1.358	0.409	0.160
	32			1.644			1.343	1.928	1.347	1.560			1.535			1.675			2.868			1.371	0.538	0.292
	50			1.703			1.345	1.959	1.358	1.654			1.545			1.711			3.178			1.382	0.649	0.438

Table 15: Forecasts for $T+2|T+1$ where break occurs at $T+1$. DGP contains lagged dependent variable with persistence of 0.5. $\psi = (0,0,0,0,0,0,0,0,0,0,1,1,1,1,1)$. $N = 15$, so there are $n = 5$ relevant regressors and $N-n = 10$ irrelevant regressors. Intercept forced and not included in measure of potency. Pool (1) given by an equally weighted average of (iv,a) Robust RW, (iii) in-sample forecast for X, and (vi,b) Direct y forecast using an AR(1). Pool (3) given by an equally weighted average of (iv,b) Robust RW with Diff and (iii) in-sample forecast for X. Bold indicates minimum MSFE for selection with unknown regressors, with underline highlighting next smallest MSFE and italic highlighting third smallest MSFE.

		(i) known regressors			(ii) in-sample mean			(iii) in-sample forecast			((iv) robust forecast (a) RW			((iv) robust forecast (b) RW with diff			(v) AR(1) forecast			(vi) direct forecast		(vii) Pool		selection	
	α (%)	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	GUM	DGP	Select	(a) RW	(b) AR(1)	(3)	(1)	Potency	Gauge
	No break																								
	0.1	1.506	1.071	1.161	1.109	1.062	1.096	1.559	1.062	1.104	1.823	1.130	1.118	1.996	1.197	1.154	2.964	1.062	1.169	1.362	1.054	1.118	1.080	0.179	0.013
	1			1.189			1.090	1.564	1.061	1.112			1.119			1.149			1.324			1.109	**1.075**	0.209	0.023
	5			1.222			1.076	1.555	1.069	1.148			1.161			1.231			1.567			1.143	1.080	0.291	0.071
	10			1.266			1.076	1.546	1.077	1.181			1.188			1.276			1.787			1.165	1.087	0.355	0.113
	16			1.296			*1.078*	1.573	1.084	1.221			1.220			1.326			2.026			1.198	1.100	0.406	0.160
	32			1.412			1.080	1.571	1.091	1.310			1.275			1.435			2.309			1.260	1.121	0.533	0.292
	50			1.460			1.080	1.576	1.079	1.347			1.318			1.512			2.609			1.292	1.132	0.645	0.436
	Break in relevant regressors																								
ι	0.1	2.182	1.436	3.603	4.440	4.066	4.213	3.489	2.578	3.882	5.177	1.622	3.715	5.561	1.528	3.616	4.705	2.504	4.003	2.664	3.849	3.690	3.786	0.182	0.013
	1			3.153			4.191	3.503	2.517	3.558			3.372			3.230			3.973			3.288	3.528	0.220	0.022
	5			2.492			4.221	3.537	2.397	3.083			2.798			2.598			3.892			2.638	3.092	0.315	0.069
	10			2.237			4.235	3.571	2.317	2.869			2.548			2.339			3.940			2.349	2.889	0.377	0.111
	16			2.067			4.256	3.587	2.276	2.657			2.382			2.178			4.044			2.140	2.729	0.424	0.159
	32			2.140			4.302	3.641	2.186	2.630			2.146			2.023			4.204			*1.969*	2.551	0.544	0.293
	50			2.089			4.325	3.636	2.147	2.582			2.044			1.961			4.415			**1.857**	2.447	0.656	0.436
λ	0.1	1.742	1.206	1.865	2.011	1.884	1.943	2.161	1.540	1.886	2.708	1.341	1.871	2.917	1.284	1.879	3.526	1.516	1.947	1.782	1.825	1.863	1.845	0.180	0.013
	1			1.809			1.933	2.167	1.531	1.834			1.804			1.796			2.105			1.783	1.794	0.210	0.023
	5			1.640			1.927	2.172	1.510	1.755			1.705			1.697			2.303			1.652	1.698	0.295	0.070
	10			1.600			1.925	2.165	1.508	1.740			1.664			1.662			2.531			1.601	1.660	0.360	0.113
	16			1.622			1.936	2.170	1.505	1.716			1.620			1.632			2.718			*1.555*	1.618	0.416	0.159
	32			1.686			1.947	2.193	1.495	1.759			1.600			1.645			2.975			1.545	1.590	0.540	0.292
	50			1.705			1.955	2.218	1.484	1.777			1.593			1.664			3.195			**1.528**	1.569	0.651	0.438
	Break in irrelevant regressors																								
ι	0.1	2.793	1.071	1.194	1.111	1.062	1.096	2.917	1.067	1.116	2.354	1.129	1.124	3.169	1.196	1.179	4.273	1.062	1.191	1.361	1.053	1.122	1.079	0.178	0.013
	1			1.304			1.086	2.976	1.071	1.174			1.139			1.212			1.381			1.143	1.084	0.209	0.022
	5			1.603			**1.072**	3.091	1.118	1.343			1.221			1.398			1.738			1.245	1.102	0.288	0.069
	10			1.779			1.074	3.204	1.146	1.465			1.277			1.523			2.032			1.319	1.121	0.348	0.111
	16			1.928			1.084	3.258	1.171	1.563			1.327			1.621			2.310			1.377	1.137	0.410	0.157
	32			2.353			*1.079*	3.462	1.229	1.885			1.416			1.848			2.872			1.542	1.194	0.536	0.290
	50			2.612			1.084	3.578	1.254	2.086			1.476			1.972			3.178			1.645	1.237	0.646	0.437
λ	0.1	1.908	1.071	1.178	1.110	1.062	1.097	2.008	1.061	1.108	2.076	1.130	1.120	2.417	1.197	1.158	3.379	1.062	1.189	1.362	1.054	1.119	1.081	0.179	0.013
	1			1.228			1.091	2.029	1.063	1.129			1.123			1.162			1.345			1.120	*1.079*	0.209	0.023
	5			1.373			1.080	2.046	1.090	1.209			1.186			1.273			1.637			1.186	1.096	0.290	0.071
	10			1.433			1.080	2.091	1.104	1.265			1.215			1.325			1.869			1.212	1.104	0.347	0.113
	16			1.488			**1.076**	2.107	1.122	1.292			1.214			1.340			2.122			1.219	1.097	0.411	0.160
	32			1.692			1.077	2.204	1.138	1.370			1.295			1.494			2.387			1.281	1.115	0.539	0.291
	50			1.803			1.085	2.227	1.160	1.505			1.333			1.571			2.753			1.347	1.144	0.648	0.437
	Break in all regressors																								
ι	0.1	3.306	1.436	3.684	4.521	4.138	4.300	2.887	2.519	3.938	5.951	1.625	3.777	7.064	1.529	3.691	4.064	2.533	4.104	2.687	3.913	3.748	3.839	0.181	0.014
	1			3.278			4.274	2.950	2.405	3.686			3.382			3.268			4.068			3.342	3.567	0.221	0.023
	5			2.713			4.283	3.153	2.201	3.091			2.821			2.710			3.927			2.639	3.050	0.311	0.069
	10			2.581			4.292	3.257	2.087	2.927			2.527			2.431			4.083			2.363	2.809	0.373	0.112
	16			2.589			4.336	3.434	2.002	2.784			2.416			2.334			4.134			2.196	2.667	0.426	0.157
	32			2.853			4.379	3.583	1.888	2.798			2.214			2.238			4.542			2.067	2.460	0.546	0.291
	50			3.067			4.411	3.730	1.829	2.821			*2.135*			2.235			4.935			**2.017**	2.357	0.657	0.438
λ	0.1	2.064	1.206	1.894	2.026	1.896	1.963	2.234	1.526	1.896	3.019	1.342	1.883	3.424	1.283	1.885	3.533	1.520	1.988	1.785	1.836	1.870	1.854	0.181	0.013
	1			1.846			1.948	2.239	1.508	1.847			1.820			1.815			2.147			1.797	1.803	0.210	0.023
	5			1.744			1.942	2.270	1.485	1.774			1.706			1.715			2.359			1.665	1.695	0.296	0.070
	10			1.727			1.936	2.322	1.477	1.765			1.666			1.666			2.518			1.610	1.653	0.361	0.112
	16			1.740			1.950	2.355	1.480	1.723			1.633			1.659			2.727			1.563	1.606	0.412	0.159
	32			1.911			1.958	2.434	1.456	1.775			1.625			1.692			3.005			**1.562**	1.573	0.546	0.293
	50			1.945			1.969	2.472	1.457	1.842			1.624			1.716			3.354			*1.566*	1.566	0.652	0.435