International Journal of Forecasting 40 (2024) 1085–1100

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Improving models and forecasts after equilibrium-mean shifts☆

Jennifer L. Castle[c,b], Jurgen A. Doornik[a,b], David F. Hendry[a,b,*]

[a] Nuffield College, Oxford, UK
[b] Climate Econometrics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK
[c] Magdalen College, Oxford, UK

ARTICLE INFO

Keywords:
Forecasting
Equilibrium-mean shifts
Step-indicator saturation (SIS)
Multiplicative indicators
Model selection

ABSTRACT

Equilibrium-mean shifts can result from changes in intercepts with constant dynamics, or be induced by shifts in dynamics with non-zero data means, or both. Induced shifts distort parameter estimates and create a discrepancy between the forecast origin and the equilibrium mean, leading to forecast failure and requiring modifications to previous forecast-error taxonomies. Step-indicator saturation can detect induced shifts, but that does not correct forecast failure. To discriminate direct from induced equilibrium-mean shifts, we augment the model by multiplicative indicators where all selected step indicators interact with the lagged regressand. Forecasts can be markedly improved after induced shifts by including these interactive indicators.

1. Introduction

There is a large body of research on forecasting in the presence of structural breaks, including selection over observations (Pesaran & Timmermann, 2007, and Pesaran, Pick, & Pranovich, 2013), expanding or rolling windows (Giacomini & Rossi, 2009, and Inoue, Jin, & Rossi, 2017), testing for and conditioning on breaks (Bai & Perron, 1998), and using robust predictors to avoid forecast failure following shifts (Hendry, 2006, Castle, Clements, & Hendry, 2015a, and Martinez, Castle, & Hendry, 2022). The forecast error taxonomy in Clements and Hendry (1998) analyses forecast-origin shifts, but changes to the equilibrium mean, or location, that occur well before the forecast origin can also be problematic, as the initial conditions for forecasts could still be far from the new equilibrium mean. Moreover, the source of equilibrium-mean shifts matters. Step-indicator saturation (SIS; see Castle, Doornik, Hendry, & Pretis, 2015b, and Section 3.2 for more details) works well when dynamics remain constant, and can even provide a reasonable approximation to the data generation process (DGP) when location shifts are induced by changes in an autoregressive parameter. Yet doing so can result in poorer forecasts than those obtained by ignoring the shifts. This phenomenon requires a modification of the taxonomy of forecast errors in Clements and Hendry (1998) from equilibrium-mean shifts at the forecast origin to include the discrepancy between the equilibrium mean and the data value at the forecast origin.

Fig. 1 illustrates the problem for forecasting monthly UK log(GDP) over 2021(1)–2022(6) by a first-order autoregression (AR(1)) that prompted the research reported below. Fitted values and one-step-ahead forecasts are reported without and with impulses determined by impulse-indicator saturation (IIS; see Hendry, Johansen, & Santos, 2008) and step indicators selected by SIS at 0.1% using Autometrics (Doornik, 2009). The estimation sample from 2019(1)–2020(12) included lockdowns to reduce the spread of Covid-19, leading to the largest ever fall in UK

☆ The results presented in this paper were reproduced by the Editor-in-Chief on 1 October 2023.

* Corresponding author at: Nuffield College, Oxford, UK.
E-mail address: david.hendry@nuffield.ox.ac.uk (D.F. Hendry).

https://doi.org/10.1016/j.ijforecast.2023.09.006
0169-2070/

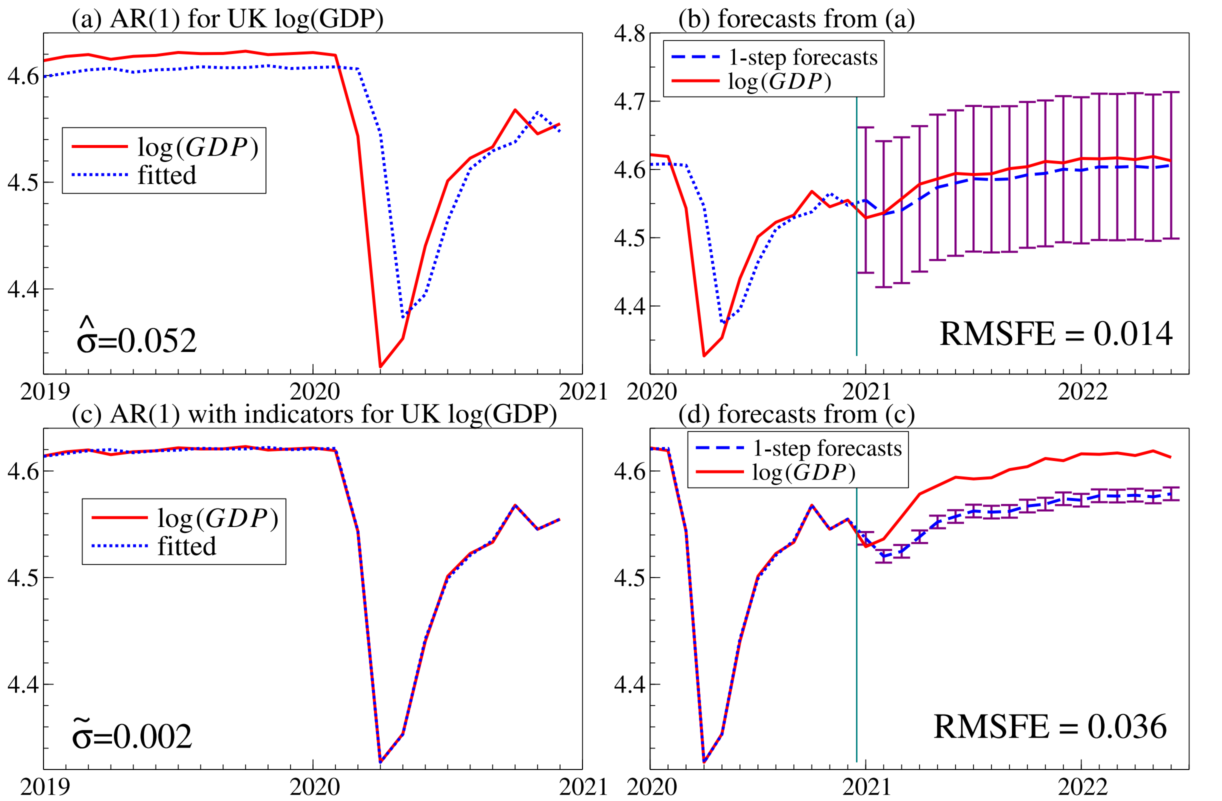

Fig. 1. (a) Actual and fitted values from an AR(1) for UK log(GDP); (b) one-step-ahead 'forecasts' from the AR(1); (c) actual and fitted values from an AR(1) with three impulse (months 4, 10, and 12 in 2020) and four step indicators (ending in months 2, 5, 7, and 10 in 2020); (d) one-step-ahead 'forecasts' from the AR(1) with selected indicators.

GDP. IIS+SIS selected three impulse and four step indicators: $\widehat{\sigma}$ and $\widetilde{\sigma}$ are the equation standard errors for the model without and with selected indicators, respectively, and RMSFE denotes the root mean square forecast error. Despite a very poor description of the in-sample data (see Fig. 1(a)) the simple AR(1) produces good forecasts, as shown in Fig. 1(b). The forecasts are far better than the in-sample fitted model. On the other hand, the model with indicators forecasts very badly relative to both its in-sample fit and to the AR(1) forecasts; see Fig. 1(d). Poor forecasts arise even though the indicators are selected at a tight significance level and the resulting equation characterises the exogenously created dramatic drop in GDP (Fig. 1(c)).

There are two distinct phenomena to be explained: why the AR(1) forecasts so well relative to its fit, and why the equation with indicators forecasts so badly. The latter has many possible explanations: overfitting (although $\widetilde{\sigma} = .2\%$ is small, it is close to the value found by fitting an AR(1) to the period 2012(8)–2018(12)), bad luck on starting the forecast on a slight downturn just before a strong recovery (but the AR(1) also did), and IIS+SIS describing the lockdown incorrectly by ignoring shifts in the dynamics. To uncover whether a location shift in-sample is direct, from a shift in an intercept given constant dynamics, or induced by changing dynamics given a constant intercept, we multiply the step indicators selected using SIS by the lagged regressand and add these interaction variables to the model. We show that this procedure can reveal whether there were shifts in the dynamics, and if so, can deliver a final selection closer to the DGP with improved forecasts.

The structure of the paper is as follows. The properties of forecasting following in-sample shifts are considered in Sections 2 and 2.1, and this leads to a taxonomy of forecast errors after an induced equilibrium-mean shift in Section 2.2. Section 3 considers estimation and forecasting for the case where the break date is known (Section 3.1) and unknown (Section 3.2). Section 3.3 notes the likely impact when there are additional regressors, and Section 3.4 discusses the possibility that the process may have a unit root. Section 4 presents a simulation analysis, with the detailed results recorded in the Appendix. Section 5 applies the new tools to an expectations-augmented Phillips Curve along with a model of the UK unemployment rate in Section 6. We briefly return to UK GDP in Section 7, before concluding in Section 8.

2. Properties of forecasting after in-sample shifts

A shift in the equilibrium mean can not only arise from a changing intercept, but also be induced by a change in the dynamics of the process. The forecast-error taxonomy of Clements and Hendry (1998) considered the former, and is updated below for the latter.

The impact of induced equilibrium-mean shifts can be seen from their effects on a previously stationary first-order autoregressive DGP:

$$y_t = \gamma + \rho y_{t-1} + \epsilon_t, \quad (1)$$

where $|\rho| < 1$, and $\epsilon_t \sim \mathsf{IN}[0, \sigma^2_\epsilon]$ denotes an independently distributed normal random variable with a constant mean of zero and a constant variance of σ^2_ϵ in-sample. The equilibrium mean in (1) is

$$\mathsf{E}[y_t] = \gamma + \rho \mathsf{E}[y_{t-1}] = \frac{\gamma}{1-\rho} = \mu. \quad (2)$$

Reformulating (1) in deviations from the in-sample equilibrium mean μ,

$$y_t - \mu = \rho\left(y_{t-1} - \mu\right) + \epsilon_t, \quad (3)$$

where (3) is estimated from a sample of $t = 1, \ldots, T$ by

$$\widehat{y}_t = \widehat{\mu} + \widehat{\rho}(y_{t-1} - \widehat{\mu}). \quad (4)$$

In (4) '^' denotes estimates on parameters and forecasts on random variables, and the estimated equilibrium mean is $\widehat{\mu} = \widehat{\gamma}/(1-\widehat{\rho})$. The h-step-ahead forecasts from (4) given a forecast origin at time T for $h = 1, \ldots$ can be written as

$$\widehat{y}_{T+h|T} - \widehat{\mu} = \widehat{\rho}\left(\widehat{y}_{T+h-1|T} - \widehat{\mu}\right) = \cdots = \widehat{\rho}^h\left(y_T - \widehat{\mu}\right). \quad (5)$$

Since

$$y_{T+h} - \mu = \rho^h\left(y_T - \mu\right) + \epsilon_T\left(h, \rho\right), \quad (6)$$

where for notational convenience we introduce

$$\epsilon_T\left(h, \rho\right) = \sum_{i=1}^{h} \rho^{h-i}\epsilon_{T+i}, \quad (7)$$

which has expectation zero. Then the forecast errors $\widehat{\epsilon}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$ from (5) are

$$\widehat{\epsilon}_{T+h|T} = \left(1-\widehat{\rho}^h\right)\left(\mu - \widehat{\mu}\right) + \left(\rho^h - \widehat{\rho}^h\right)\left(y_T - \mu\right) + \epsilon_T\left(h, \rho\right), \quad (8)$$

with $\mathsf{E}[\widehat{\epsilon}_{T+h|T}] \simeq 0$.

2.1. In-sample shift

A break at the forecast origin or during a forecast period will later become an in-sample shift. When the system has adjusted to its previous equilibrium mean μ before an induced shift from ρ to $\rho_\diamond$ at time $\tau < T$ altering μ to $\mu_\diamond = (1-\rho_\diamond)^{-1}\gamma_\diamond$, where we allow for the possibility that the intercept may also have changed, then providing there are no forecast period breaks,

$$y_\tau = \gamma_\diamond + \rho_\diamond y_{\tau-1} + \epsilon_\tau = \mu_\diamond + \rho_\diamond\left(y_{\tau-1} - \mu_\diamond\right) + \epsilon_\tau. \quad (9)$$

In (9), a *direct shift* is when $\rho = \rho_\diamond$ but $\mu \neq \mu_\diamond$ because γ has changed; an *induced shift* is $\mu \neq \mu_\diamond$ because ρ has changed with $\gamma = \gamma_\diamond$ constant. Combinations are clearly possible. Then for $s = 1, \ldots, T - \tau + h$:

$$y_{\tau+s} = \mu_\diamond + \rho^s_\diamond\left(y_\tau - \mu_\diamond\right) + \epsilon_\tau\left(s, \rho_\diamond\right). \quad (10)$$

Taking expectations of (9) at τ when the distribution of $\{\epsilon_t\}$ is unchanged, as $\mathsf{E}[y_{\tau-1}] = \mu$,

$$\mathsf{E}\left[y_\tau\right] = \mu_\diamond - \rho_\diamond\left(\mu_\diamond - \mu\right),$$

and as $y_{\tau+s}$ is y_T when $s = T - \tau$,

$$\mathsf{E}\left[y_T\right] = \mu_\diamond - \rho_\diamond^{\,T-\tau+1}\left(\mu_\diamond - \mu\right). \quad (11)$$

Consequently, depending on the recency of the last shift at τ, the magnitudes of the shifts in γ and ρ and the absolute magnitude of $\rho_\diamond$, $\mathsf{E}[y_T]$ could be far from the new equilibrium mean $\mu_\diamond$. The forecast errors from using (5) when future data are generated by

$$y_{T+h} = \mu_\diamond + \rho^h_\diamond\left(y_T - \mu_\diamond\right) + \epsilon_T\left(h, \rho_\diamond\right), \quad (12)$$

are

$$\begin{aligned}\widehat{\epsilon}_{T+h|T} &= \left(\mu_\diamond - \widehat{\mu}\right) + \rho^h_\diamond\left(y_T - \mu_\diamond\right) \\ &\quad - \widehat{\rho}^h\left(y_T - \widehat{\mu}\right) + \epsilon_T\left(h, \rho_\diamond\right) \\ &= \left(1-\widehat{\rho}^h\right)\left(\mu_\diamond - \widehat{\mu}\right) \\ &\quad + \left(\rho^h_\diamond - \widehat{\rho}^h\right)\left(y_T - \mu_\diamond\right) + \epsilon_T\left(h, \rho_\diamond\right). \quad (13)\end{aligned}$$

As the induced in-sample shift has made the forecast origin depend on $\rho_\diamond$ and $\mu_\diamond$, we use (11) to write $y_T - \mu_\diamond$ as

$$\begin{aligned}y_T - \mu_\diamond &= \left(\mathsf{E}[y_T] - \mu_\diamond\right) + \left(y_T - \mathsf{E}\left[y_T\right]\right) \\ &= -\rho_\diamond^{\,T-\tau+1}\left(\mu_\diamond - \mu\right) + \left(y_T - \mathsf{E}\left[y_T\right]\right), \quad (14)\end{aligned}$$

and we rewrite the forecast errors in (13) as

$$\begin{aligned}\widehat{\epsilon}_{T+h|T} &= \left[\left(1-\rho^h_\diamond\right) + \left(\rho^h_\diamond - \widehat{\rho}^h\right)\left(1 - \rho_\diamond^{\,T-\tau+1}\right)\right]\left(\mu_\diamond - \mu\right) \\ &\quad + \left(1-\widehat{\rho}^h\right)\left(\mu - \widehat{\mu}\right) \\ &\quad + \left(\rho^h_\diamond - \widehat{\rho}^h\right)\left(y_T - \mathsf{E}\left[y_T\right]\right) + \epsilon_T\left(h, \rho_\diamond\right). \quad (15)\end{aligned}$$

Full-sample expected values of parameter estimates are denoted by a subscript $_p$, so $\mathsf{E}[\widehat{\mu}] = \mu_p$, $\mathsf{E}[\widehat{\rho}] = \rho_p$, and we approximate by setting $\mathsf{E}[\widehat{\rho}^h] = \rho^h_p$. Then, taking expectations of the multi-step forecast error in (15) leads to the following[1]:

$$\begin{aligned}\mathsf{E}\left[\widehat{\epsilon}_{T+h|T}\right] &\simeq \left(1-\rho^h_p\right)\left(\mu_\diamond - \mu_p\right) \\ &\quad - \rho_\diamond^{\,T-\tau+1}\left(\rho^h_\diamond - \rho^h_p\right)\left(\mu_\diamond - \mu\right). \quad (16)\end{aligned}$$

Thus, (16) shows two sources of systematic non-zero forecast errors from induced shifts due to $\mu_\diamond \neq \mu$ and $\mu_\diamond \neq \mu_p$. The magnitude of the first right-hand-side term in (16) depends on how close to unity the estimate of ρ is, and the second on the time since the shift and the deviation $\rho^h_\diamond - \rho^h_p$.

2.2. A forecast-error taxonomy after an in-sample shift

The terms in (13) can be separated to isolate their effects, and (19) puts these into a taxonomy. First,

$$\left(\mu_\diamond - \widehat{\mu}\right) = \left(\mu_\diamond - \mu\right) + \left(\mu - \mu_p\right) + \left(\mu_p - \widehat{\mu}\right), \quad (17)$$

corresponding to the equilibrium-mean shift, the bias in estimating the previous equilibrium mean, and sampling variability. Similarly,

$$\left(\rho^h_\diamond - \widehat{\rho}^h\right) = \left(\rho^h_\diamond - \rho^h\right) + \left(\rho^h - \rho^h_p\right) + \left(\rho^h_p - \widehat{\rho}^h\right), \quad (18)$$

and

$$\left(1-\widehat{\rho}^h\right) = \left(1-\rho^h_p\right) + \left(\rho^h_p - \widehat{\rho}^h\right).$$

[1] In their taxonomy of forecast-origin shifts, Clements and Hendry (1998) used the approximation $\widehat{\rho}^h = (\rho_p + (\widehat{\rho} - \rho_p))^h \simeq \rho^h_p + C_h$ so $C_h(y_T - \mu_p) = F_h(\widehat{\rho} - \rho_p)$.

Using (17), ((18), and (14), we rewrite the forecast errors as

$$\begin{aligned}\widehat{\epsilon}_{T+h|T} = &\ (\mu_\diamond - \mu) + (\mu - \mu_p) + (\mu_p - \widehat{\mu}) \\ &+ \left(\left(\rho_\diamond^h - \rho^h\right) + \left(\rho^h - \rho_p^h\right) + \left(\rho_p^h - \widehat{\rho}^h\right)\right) \\ &\times \left(\left(\mathsf{E}[y_T] - \mu_\diamond\right) + \left(y_T - \mathsf{E}[y_T]\right)\right) \\ &- \widehat{\rho}^h\left((\mu_\diamond - \mu) + (\mu - \mu_p) + (\mu_p - \widehat{\mu})\right) \\ &+ \epsilon_T(h, \rho_\diamond),\end{aligned}$$

which can be reorganised as

$$\begin{aligned}\widehat{\epsilon}_{T+h|T} = &\ \left(\rho_\diamond^h - \rho^h\right)(y_T - \mathsf{E}[y_T]) + \left(1 - \rho_p^h\right)(\mu_\diamond - \mu) \\ &+ \left(\rho^h - \rho_p^h\right)(y_T - \mathsf{E}[y_T]) \\ &+ \left(1 - \widehat{\rho}^h\right)(\mu - \mu_p) + \left(\rho_p^h - \widehat{\rho}^h\right)(y_T - \mu) \\ &+ \left(1 - \rho_\diamond^h\right)(\mu_p - \widehat{\mu}) \\ &+ \left(\rho_\diamond^h - \rho_p^h\right)(\mathsf{E}[y_T] - \mu_\diamond) \\ &+ \left(\rho_\diamond^h - \widehat{\rho}^h\right)(\mu_p - \widehat{\mu}) + \epsilon_T(h, \rho_\diamond),\end{aligned}$$

leading to the multi-step forecast-error taxonomy in (19).

Multi-step forecast-error taxonomy after in-sample parameter changes at $\tau < T$

$$\begin{aligned}\widehat{\epsilon}_{T+h|T} = &\ \left(\rho_\diamond^h - \rho^h\right)(y_T - \mathsf{E}[y_T]) \\ &\quad (i.a) \text{ slope change (expectation zero)} \\ &+ \left(1 - \rho_p^h\right)(\mu_\diamond - \mu) \\ &\quad (i.b) \text{ equilibrium-mean change} \\ &+ \left(\rho^h - \rho_p^h\right)(y_T - \mathsf{E}[y_T]) \\ &\quad (ii.a) \text{ slope mis-specification (expectation zero)} \\ &+ \left(1 - \widehat{\rho}^h\right)(\mu - \mu_p) \\ &\quad (ii.b) \text{ equilibrium-mean mis-specification} \\ &+ \left(\rho_p^h - \widehat{\rho}^h\right)(y_T - \mu) \\ &\quad (iii.a) \text{ slope estimation (expectation zero)} \\ &+ \left(1 - \rho_\diamond^h\right)(\mu_p - \widehat{\mu}) \\ &\quad (iii.b) \text{ equilibrium-mean estimation (expectation zero)} \\ &+ \left(\rho_\diamond^h - \rho_p^h\right)(\mathsf{E}[y_T] - \mu_\diamond) \\ &\quad (iv) \text{ forecast origin mis-specification} \\ &+ \left(\rho_\diamond^h - \widehat{\rho}^h\right)(\mu_p - \widehat{\mu}) \\ &\quad (v) \text{ estimation covariance} \\ &+ \epsilon_T(h, \rho_\diamond) \\ &\quad (vi) \text{ cumulative error (expectation zero).}\end{aligned} \tag{19}$$

Under the assumption of no shifts in-sample, $\rho_\diamond = \rho$, $\mu_\diamond = \mu$ and $\mu_p = \mu$ so all the terms in (19) have expectations of zero, with the exception of *(v)*, which is $O_p(T^{-1})$. Next, a direct equilibrium-mean shift has $\rho_\diamond = \rho$ but $\mu_\diamond \neq \mu \neq \mu_p$ and hence $\rho_p \neq \rho$, so *(i.b)*, *(ii.b)*, and *(iv)* are non-zero in expectation. However, when $\rho_\diamond \neq \rho$ but $\mu_\diamond = \mu$ so the equilibrium-mean is constant, then *(i.b)*, *(ii.b)*, and *(iv)* are zero in expectation, and no systematic forecast errors will result. This implication provides a test of (19).

With unmodelled in-sample induced shifts, *(i.b)*, *(ii.b)*, and *(iv)* are again non-zero in expectation. Thus, (19) explains why even when there are no forecast-period shifts, the additional terms of equilibrium-mean mis-specification and forecast origin mis-specification can lead to forecast failure in addition to the well-known impact of *(i.b)* (see Clements & Hendry, 1998). The terms *(iii.a)* and *(iii.b)* and the estimation covariance *(v)* will cause variability in outcomes, but not systematic forecast failure. Setting $(T-\tau) = b$, $\rho_p = \rho$, $\mu_p = \mu$ and replacing $_\diamond$ terms by * in Martinez et al. (2022) delivers their equation (2.5), which leads to their smooth robust forecasting approach. However, ρ_p is likely to deviate importantly from both ρ and $\rho_\diamond$, with $\widehat{\rho}$ being driven towards unity to minimise the unmodelled gap between $\mu_\diamond$ and μ. Hendry and Neale (1991) demonstrated that this effect badly distorts tests for unit roots in the dynamics, falsely suggesting their presence. Setting $\widehat{\rho} = 1$ eliminates $\mu - \mu_p$, suggesting that forecasts from a first-order autoregression may not be greatly distorted by induced in-sample shifts that occur well before the forecast origin, despite the shift in dynamics not being modelled. This explains the outcomes in the top row of Fig. 1, where the RMSFE is far smaller than the in-sample $\widehat{\sigma}$. Conversely, while SIS led to a much better in-sample data description in Fig. 1, the forecasts were very poor, an issue we address in the next section.

3. Estimation in the presence of an induced equilibrium-mean shift

We first consider estimation and forecasting when the date τ of the induced shift is known. We then extend the analysis to the case where the timings of the (possible) breaks are unknown.

3.1. Induced equilibrium-mean shift at a known date

Let S_τ be the step indicator (SI) capturing the timing of the shift, so $S_\tau = 1$ until $t = \tau$, and 0 thereafter. When (9) becomes the DGP at $t = \tau$, the whole sample DGP can be written as

$$\begin{aligned}y_t &= \mu_\diamond - (1-\rho)(\mu_\diamond - \mu) S_\tau + \rho_\diamond (y_{t-1} - \mu_\diamond) \\ &\quad - (\rho_\diamond - \rho) S_\tau (y_{t-1} - \mu_\diamond) + \epsilon_t \\ &= \begin{cases} \mu + \rho (y_{t-1} - \mu) + \epsilon_t & \text{for } t \leq \tau \text{ when } S_\tau = 1, \\ \mu_\diamond + \rho_\diamond (y_{t-1} - \mu_\diamond) + \epsilon_t & \text{for } t > \tau \text{ when } S_\tau = 0. \end{cases}\end{aligned} \tag{20}$$

This clarifies that there is both a step shift and an interaction between that step indicator and the lagged regressand. Because the equilibrium-mean shift is induced by the change in dynamics, the timing of S_τ is the same for the interaction.

Consider adding just S_τ to the in-sample estimated model:

$$\widetilde{y}_t = \widetilde{\mu}_\diamond + \widetilde{\rho}_\diamond (y_{t-1} - \widetilde{\mu}_\diamond) + \widetilde{\phi} S_\tau, \tag{21}$$

denoting estimates by $\tilde{}$ and $\phi = -(1-\rho)(\mu_\diamond - \mu)$ with the forecasting equation

$$\widetilde{y}_{T+h|T} = \widetilde{\mu}_\diamond + \widetilde{\rho}_\diamond^h (y_T - \widetilde{\mu}_\diamond),$$

which leads to the following forecast errors:

$$\widetilde{\epsilon}_{T+h|T} = \mu_\diamond + \rho_\diamond^h (y_T - \mu_\diamond) - \widetilde{\mu}_\diamond - \widetilde{\rho}_\diamond^h (y_T - \widetilde{\mu}_\diamond) + \epsilon_T(h, \rho_\diamond). \tag{22}$$

Even if $\widetilde{\phi} \approx -(1-\widetilde{\rho})(\widetilde{\mu}_\diamond - \widetilde{\mu})$ in (21) that need not remove the forecast failure as $\mathsf{E}[\widetilde{\rho}_\diamond] \neq \rho_\diamond$ by omitting the interaction term. However, when $\rho_\diamond = \rho$, as in a direct shift, then $\mathsf{E}[\widetilde{\rho}] \approx \rho$ and $\mathsf{E}[\widetilde{\mu}_\diamond] \approx \mu_\diamond$. Hence, the expected values of the terms in (22) will be close to zero, highlighting the key difference from an induced shift.

Importantly, adding the interacted step indicator (called MSI for the multiplicative step indicator) to the DGP formulation yields (20) to represent the shift. Then, letting $\delta = -(\rho_\diamond - \rho)$, once S_τ is known or found, the estimated model (denoted by $\bar{\cdot}$) becomes

$$\overline{y}_t = \overline{\mu}_\diamond + \overline{\phi} S_{\{\tau\}} + \overline{\rho}_\diamond (y_{t-1} - \overline{\mu}_\diamond) + \overline{\delta} S_\tau (y_{t-1} - \overline{\mu}_\diamond) = \begin{cases} \overline{\mu} + \overline{\rho}(y_{t-1} - \overline{\mu}) & \text{for } S_\tau = 1, \\ \overline{\mu}_\diamond + \overline{\rho}_\diamond (y_{t-1} - \overline{\mu}_\diamond) & \text{for } S_\tau = 0. \end{cases}$$

Approximating by $\overline{\delta} \approx -(\overline{\rho}_\diamond - \overline{\rho})$ and $\overline{\phi} \approx -(1 - \overline{\rho})(\overline{\mu}_\diamond - \overline{\mu})$, which estimates the DGP in both periods with $\overline{\epsilon}_t = y_t - \overline{y}_t$,

$$\overline{\epsilon}_t = \begin{cases} (1-\overline{\rho})(\mu - \overline{\mu}) \\ \quad + (\rho - \overline{\rho})(y_{t-1} - \mu) + \epsilon_t & \text{for } S_\tau = 1, \\ \left(1 - \overline{\rho}_\diamond\right)(\mu_\diamond - \overline{\mu}_\diamond) \\ \quad + \left(\rho_\diamond - \overline{\rho}_\diamond\right)(y_{t-1} - \mu_\diamond) + \epsilon_t & \text{for } S_\tau = 0. \end{cases}$$

However, exactly matching the post-shift DGP need not fully remove forecast biases as the estimated parameters are from a changing equilibrium mean and dynamics. Nevertheless, $\mathsf{E}[\overline{\epsilon}_t]$ should be small, delivering the forecasting equation $\overline{y}_{T+h|T} = \overline{\mu}_\diamond + \overline{\rho}_\diamond^h (y_T - \overline{\mu}_\diamond)$, which should avoid forecast failure.

3.2. Estimation for unknown date

When it comes to practical model development, the locations, sizes, and types (direct or induced) of shifts are usually unknown. To handle this, we adopt a model selection approach combined with saturating the basic model with a full set of indicator variables. No degrees of freedom are left after saturation, so an adapted procedure is needed to discover which are the relevant indicators. As noted in the introduction, other approaches are feasible too.

A theoretical analysis of impulse-indicator saturation (IIS) under the null of no outliers was provided by Hendry et al. (2008) for the location-scale model, and extended in Johansen and Nielsen (2009) to both stationary and unit-root autoregressions.

Step-indicator saturation (SIS) saturates the model with the full set of $\{S_\tau\}$ for all τ, so it is targeted to discover direct breaks in the equilibrium mean. Importantly, SIS does not require knowledge of the locations of breaks, the maximum number of shifts, nor impose a minimum break length, so it allows shifts to occur at the start and/or end of the sample. Castle et al. (2015b) showed that detecting location shifts by SIS has the correct null retention frequency in constant simple models for a nominal selection size of α.

Saturation methods (such as IIS and SIS) can be feasibly estimated because software like Autometrics handles more candidate variables, N, than observations, T: a combination of expanding and contracting multiple block searches is used, with model selection applied within each block. The algorithm proceeds until it can 'learn' no more from the data; details can be found in Castle, Doornik, and Hendry (2021b).

The analysis of the previous section shows that SIS will have the power to detect both direct and induced equilibrium mean shifts, although it is not optimal for the latter. This leads us to propose a two-stage estimation procedure, denoted MSIS, for the AR(1) model:

1. Apply SIS to the null model at an adopted significance level α_1:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \{S_3, \ldots, S_T\}^{\alpha_1} + u_t,\ t = 1, \ldots, T, \tag{23}$$

   with Autometrics selection at significance level α_1 over the $T-2$ step indicators inside $\{\cdot\}$. In dynamic models we usually set α_1 tightly to concentrate on large breaks. We denote the selected set of step indicators by $\mathcal{S}_1$. If this set is empty, no breaks were detected, and the second stage is redundant.
2. Construct the interactions of the selected indicators and the lagged dependent variable, denoted as $\mathcal{S}_1 \times y_{t-1}$. Add both $\mathcal{S}_1$ and $\mathcal{S}_1 \times y_{t-1}$ to the null model, and apply model selection, possibly at a looser significance level α_2:

$$y_t = \beta_0 + \{y_{t-1}, \mathcal{S}_1, \mathcal{S}_1 \times y_{t-1}\}^{\alpha_2} + v_t. \tag{24}$$

   We could decide to always include the lagged dependent variable.

The derivations in Section 3.1 show that model specifications matching the DGP can be formulated for both direct and induced equilibrium mean shifts once the appropriate step shifts have been identified, so this should avoid forecast failure. The proposed estimation procedure employs SIS to discover the breaks, and then uses the fact that the additional term from an induced shift occurs at the same date. However, that requires that SIS accurately locates the relevant shifts so the correct interactive steps can be created. Then model selection is applied to the resulting specification retaining the relevant terms (an issue of potency) without keeping too many irrelevant variables (an issue of gauge).

The simulation study in the next section investigates the extent to which estimation succeeds and the resulting impact on forecasts. But first we show that the procedure can be extended to models with additional regressors.

3.3. Additional regressors

Including in (1) a set of variables $\mathbf{x}_t$ where $\mathsf{E}[\mathbf{x}_t] = \boldsymbol{\lambda}$ with parameters $\boldsymbol{\beta}$ yields

$$y_t = \gamma + \rho y_{t-1} + \boldsymbol{\beta}'\mathbf{x}_t + \epsilon_t = \mu + \rho(y_{t-1} - \mu) + \boldsymbol{\beta}'(\mathbf{x}_t - \boldsymbol{\lambda}) + \epsilon_t, \tag{25}$$

where $\mu = (\gamma + \boldsymbol{\beta}'\boldsymbol{\lambda})/(1-\rho)$. Then (25) entails that a location shift could be due to a change in $\boldsymbol{\beta}$ to $\boldsymbol{\beta}_\diamond$ at time $\tau < T$:

$$y_\tau = \gamma + \boldsymbol{\beta}_\diamond'\boldsymbol{\lambda} + \rho y_{\tau-1} + \boldsymbol{\beta}_\diamond'(\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t = \mu_\diamond + \rho(y_{\tau-1} - \mu_\diamond) + \boldsymbol{\beta}_\diamond'(\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t, \tag{26}$$

where $\mu_\diamond = (\gamma + \boldsymbol{\beta}_\diamond'\boldsymbol{\lambda})/(1-\rho)$. Consequently, direct, rather than induced, location shifts result when ρ is constant, so SIS should be able to detect substantive shifts in $\boldsymbol{\beta}_\diamond'\boldsymbol{\lambda}$, since

$$\begin{aligned} y_t &= \mu_\diamond - (1-\rho)(\mu_\diamond - \mu) S_\tau + \rho (y_{t-1} - \mu_\diamond) \\ &\quad + \boldsymbol{\beta}_\diamond'(\mathbf{x}_t - \boldsymbol{\lambda}) - \left(\boldsymbol{\beta}_\diamond - \boldsymbol{\beta}\right)' S_\tau (\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t \\ &= \begin{cases} \mu + \rho (y_{t-1} - \mu) + \boldsymbol{\beta}'(\mathbf{x}_t - \boldsymbol{\lambda}) + \epsilon_t & \text{for } S_\tau = 1, \\ \mu_\diamond + \rho (y_{t-1} - \mu_\diamond) + \boldsymbol{\beta}_\diamond'(\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t & \text{for } S_\tau = 0. \end{cases} \end{aligned}$$

If ρ also changes, then

$$\begin{aligned} y_\tau &= \gamma + \boldsymbol{\beta}_\diamond'\boldsymbol{\lambda} + \rho_\diamond y_{\tau-1} + \boldsymbol{\beta}_\diamond'(\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t \\ &= \mu_\diamond + \rho_\diamond (y_{\tau-1} - \mu_\diamond) + \boldsymbol{\beta}_\diamond'(\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t, \end{aligned}$$

where $\mu_\diamond = (\gamma + \boldsymbol{\beta}_\diamond'\boldsymbol{\lambda})/(1-\rho_\diamond)$, so the DGP can be written as follows:

$$\begin{aligned} y_t &= \mu_\diamond - (1-\rho)(\mu_\diamond - \mu) S_\tau + \rho_\diamond (y_{t-1} - \mu_\diamond) \\ &\quad - (\rho_\diamond - \rho) S_\tau (y_{t-1} - \mu_\diamond) \\ &\quad + \boldsymbol{\beta}_\diamond'(\mathbf{x}_t - \boldsymbol{\lambda}) - \left(\boldsymbol{\beta}_\diamond - \boldsymbol{\beta}\right)' S_{\{\tau\}} (\mathbf{x}_\tau - \boldsymbol{\lambda}) + \epsilon_t. \end{aligned} \quad (27)$$

Hence, the above MSIS approach for the simpler case of an autoregression could be implemented more generally.

Next we briefly consider the possibility that $\rho = 1$.

3.4. Unit-root processes

When $\rho = 1$, then

$$y_t = \gamma + y_{t-1} + \epsilon_t, \quad (28)$$

or

$$\Delta y_t = \gamma + \epsilon_t,$$

so γ becomes a growth rate and only direct shifts can occur, albeit with small changes. A stochastic unit root ρ either requires $\gamma = 0$ or changes in γ are somehow linked to the changes in ρ to avoid large changes in the signs and magnitudes of the implied growth.

However, if (28) is the marginal process for $\{y_t\}$, and its DGP is known to include n strongly exogenous cointegrating regressors $\{\mathbf{x}_t\}$, that is,

$$y_t = \gamma + \boldsymbol{\beta}'\mathbf{x}_t + \psi y_{t-1} + \epsilon_t, \quad (29)$$

then shifts in ψ with γ and $\boldsymbol{\beta}$ constant are induced, whereas changes in γ and $\boldsymbol{\beta}$ are direct. Despite unit roots in $\{\mathbf{x}_t\}$ making it non-stationary, (29) can be estimated by least squares (see Sims, Stock, & Watson, 1990) including with SIS. If step shifts are detected, MSIS can be applied to check if the shifts were induced, as in (27).

4. Simulation results

The simulations study the impact on estimation and forecasting of two breaks, induced or direct, for a relatively small sample size of 45 observations.

4.1. Design of the experiments

The simulation DGP has two breaks at $0 < \tau_1 < \tau_2 \le T + H$, where H denotes the forecast horizon. Data are generated for $t = -19, \ldots, 0, 1, \ldots, T+H$. Observations $t = -19, \ldots, -1$ are discarded, using $t = 1, \ldots, T$ for estimation with one lag of the dependent variable using observation zero:

$$\begin{aligned} y_t &= \gamma_1 S_{\tau_1} + \gamma_2 \left(S_{\tau_2} - S_{\tau_1}\right) + \gamma_3 \left(S_{T+H} - S_{\tau_2}\right) \\ &\quad + \rho_1 y_{t-1} S_{\tau_1} + \rho_2 y_{t-1} \left(S_{\tau_2} - S_{\tau_1}\right) \\ &\quad + \rho_3 y_{t-1} \left(S_{T+H} - S_{\tau_2 t}\right) + \epsilon_t \\ &= (\gamma_1 - \gamma_2) S_{\tau_1} + (\gamma_2 - \gamma_3) S_{\tau_2} + \gamma_3 \\ &\quad + \left[(\rho_1 - \rho_2) S_{\tau_1} + (\rho_2 - \rho_3) S_{\tau_2} + \rho_3\right] y_{t-1} + \epsilon_t \end{aligned} \quad (30)$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ with $\sigma_\epsilon^2 = 1$. DGP (30) can also be written as

$$y_t = \delta_1 S_{\tau_1} + \delta_2 S_{\tau_2} + \gamma_3 + \phi_1 S_{\tau_1} y_{t-1} + \phi_2 S_{\tau_2} y_{t-1} + \rho_3 y_{t-1} + \epsilon_t. \quad (31)$$

Three sets of parameter combinations are investigated in the experiments, keeping $\gamma_1 = \gamma_3 = 5$ fixed throughout:

[1] $\gamma_2 = 5$ and $\rho_1 = \rho_3 = 0.9$ but ρ_2 changes from 0.9 to 0.4 by steps of 0.1 (induced 'recession' break);
[2] when $\gamma_2 = 5$ and $\rho_2 = 0.9$ are constant whereas $\rho_1 = \rho_3$ changes from 0.8 to 0.4 by steps of 0.1 (induced 'boom' break); and
[3] when $\rho_1 = \rho_2 = \rho_3 = 0.9$ are constant but γ_2 changes from 4 to 0 by steps of 1.0 (direct break: shift in intercept).

The first two experiments have $\gamma_1 = \gamma_2 = \gamma_3$, and so $\delta_1 = \delta_2 = 0$. The third experiment has $\phi_1 = \phi_2 = 0$; this also happens in the first, but only when $\rho_2 = 0.9$. All experiments use $T = 45$, with the two breaks at $\tau_1 = 30$ and $\tau_2 = 40$.

All simulation models (including the DGP specification) were estimated by ordinary least squares (OLS) and saturation estimation using Autometrics:

DGP (30), omitting intercept shifts if $\gamma_1 = \gamma_2 = \gamma_3$, and autoregressive shifts if $\rho_1 = \rho_2 = \rho_3$.

AR1 AR(1) model: $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$.

SIS (23), i.e. step-indicator saturation of the AR(1) at significance level $\alpha_1 = 0.001$.

MSIS The two-step procedure (23) followed by (24), with significance $\alpha_1 = 0.001$ in the first stage, and $\alpha_2 = 0.01$ in the second stage.

$M = 10{,}000$ replications are used and all experiments have the same initial seed. To facilitate the interpretation of the outcomes, we mainly provide graphical reports in this section, with detailed results in the Appendix.

4.2. Estimation and break detection

Fig. 2 plots the average equation standard errors $\widehat{\sigma}_\epsilon$ for the 'recession', 'boom', and direct location shift cases and shows that $\widehat{\sigma}_\epsilon$ for MSIS is close to that of the DGP, so there is no evidence of overfitting or mis-specification, with improvements over SIS when location shifts are induced.

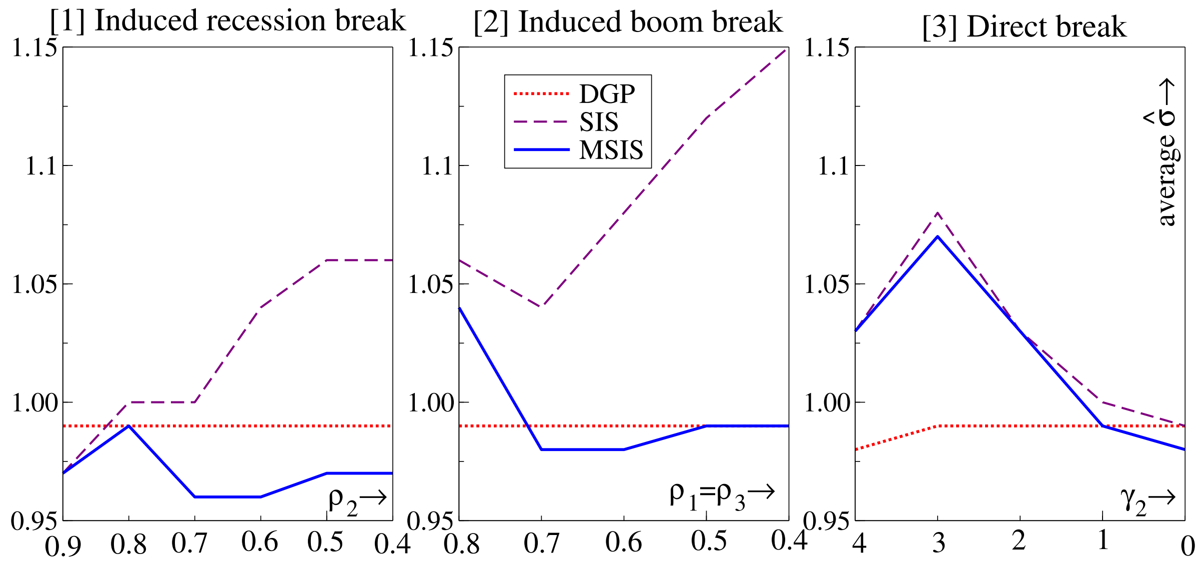

Fig. 2. Average equation standard errors, [1]: $\rho_1 = \rho_3 = 0.9$, $\rho_2 = 0.9$; [2]: $\rho_2 = 0.9$, $\gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9$, $\gamma_2 = 5$.

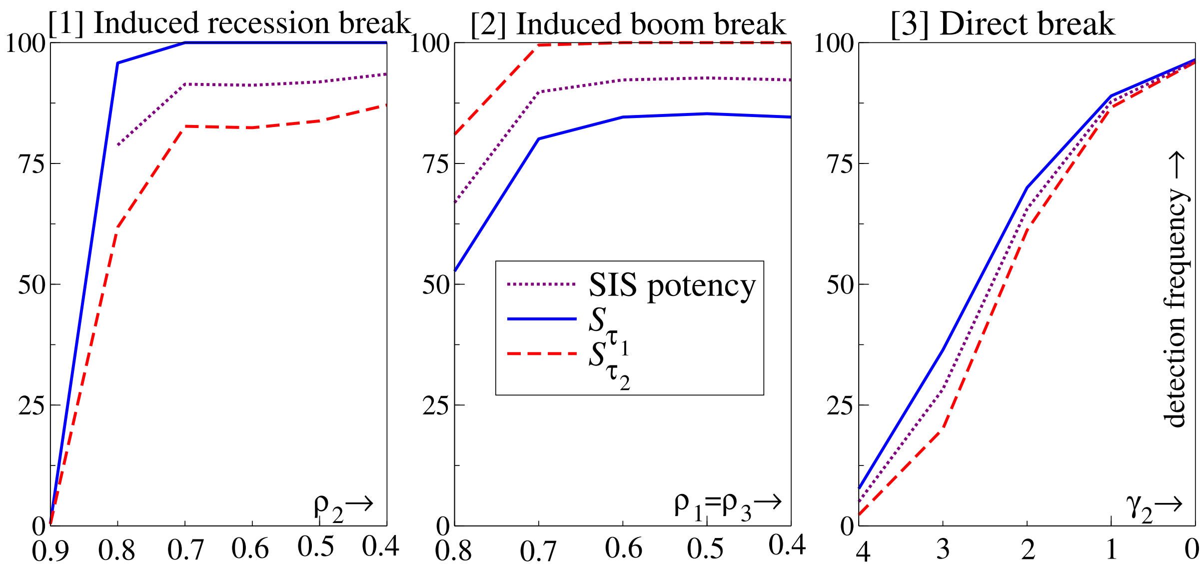

Fig. 3. Break detection frequencies and potencies for SIS, [1]: $\rho_1 = \rho_3 = 0.9$, $\rho_2 = 0.9$; [2]: $\rho_2 = 0.9$, $\gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9$, $\gamma_2 = 5$.

SIS is known to be somewhat over-gauged (retention of S_j, $j \neq \tau_1, \tau_2$) from earlier simulation studies, partly because two steps are needed to capture an outlier, but also here because the model is mis-specified when the location shift is induced by a change in dynamics.

Fig. 3 plots the break detection frequencies and potencies (retention of S_{τ_1}, S_{τ_2} when there is a break in the DGP) for SIS selecting at $\alpha_1 = 0.001$. Potency rises rapidly with changes in the autoregressive parameters, and lies between the break detection rates. In Panel [1], the initial point measures the gauge, then the break detection rate rises rapidly. However, the first point in Panel [2] already shows a high break detection rate from a small change in $\rho_1 = \rho_3$. Conversely, in Panel [3] substantial changes in γ_2 are required for detection. The explanation is that large changes in the equilibrium mean result from changing dynamics in the first two cases, versus only a small change from the direct impact.

For MSIS selecting at $\alpha_1 = 0.001$, the gauges and potencies are conditional on extending the SIS selection to MSIS. The 'gauge' is how often a step indicator falsely retained by SIS is also retained by MSIS, shown dotted in Fig. 4 as MSIS:τ_1, MSIS:τ_2. The values seem large, up to 50%, but as the overall retention rates are the product with the SIS gauge, they are low at less than 0.7%. There are two sets of break detection measures: first, how often MSIS retains a step indicator selected by SIS; and second, how often MSIS selects the interaction of the SIS indicator with y_{t-1}, shown as MSISy:τ_1 and MSISy:τ_2 (solid lines). Potency is how often a correctly detected step indicator is

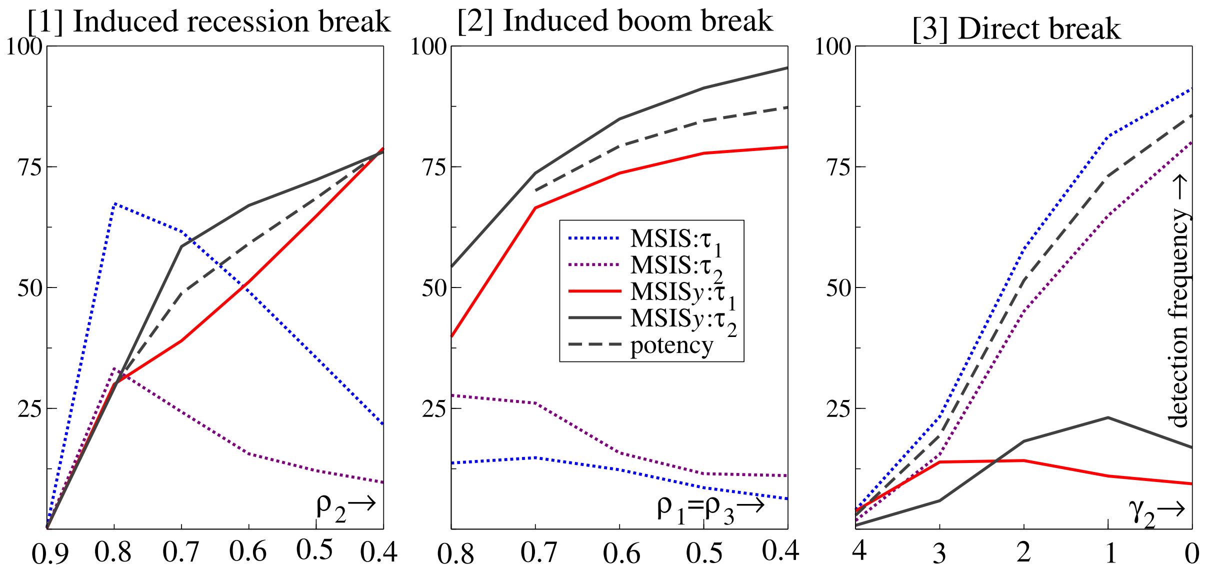

Fig. 4. Break detection frequencies for MSIS, [1]: $\rho_1 = \rho_3 = 0.9$, $\rho_2 = 0.9$; [2]: $\rho_2 = 0.9$, $\gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9$, $\gamma_2 = 5$.

also retained by MSIS, so it is the interaction in Panels [1] and [2], but the location shift in Panel [3], where potency and 'gauge' switch meanings. As Fig. 4 shows, potency again increases with the magnitudes of changes in the autoregressive parameters.

Table 1 gives the average unconditional root mean square errors (RMSEs) for estimating the parameters in the formulation in (31) in terms of the DGP and MSIS parameters δs and ϕs. Manifestly, infeasible knowledge of the DGP, including what shifts happened and when, makes a huge difference to RMSEs.

4.3. Impact on forecasts

Turning to the impact on forecasts of detecting (or not) direct and induced location shifts, four forecasting methods were considered for each experimental design:

(i) One-step -ahead (static) forecasts;

Breaks do not extend beyond the end of the estimation sample, so for a one-step-ahead forecast in the DGP (31),

$$\widehat{y}_{T+1} = \widehat{\gamma}_3 + \widehat{\rho}_3 y_T.$$

(ii) Robust one-step forecasts, as implemented in PcGive.

Forecasts are obtained from the differenced model using parameter estimates from the original (undifferenced) model, and then reintegrated. For the MSIS model this works as follows, assuming that SIS found one step indicator at $\widehat{\tau}$, and MSIS then retained the interaction and lagged dependent variable. The estimated MSIS model is

$$\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 y_{t-1} + \widehat{\beta}_2 S_{\widehat{\tau}} + \widehat{\beta}_3 S_{\widehat{\tau}} y_{t-1},$$

Then after differencing, the first-differenced forecast is

$$\Delta\widehat{y}_{T+1|T} = \widehat{\beta}_1 \Delta y_T,$$

because the step indicators are zero after $t = \widehat{\tau} \leq T$. So the robust one-step forecast is $y_T + \widehat{\beta}_1 \Delta y_T$.

(iii) Five-step dynamic (ex ante) forecasts, up to five periods ahead.

(iv) Cardt five-step dynamic forecasts from an automatic procedure (so they do not depend on the model).

Cardt is an extended version of the automated forecasting procedure that we designed for the M4 competition; see Doornik, Castle, and Hendry (2020) for the original, and Castle, Doornik, and Hendry (2021a) for the improvements. The procedure takes decisions about the use of logarithms and seasonality, neither important in the simulations, as well as dynamics in the forecasting model. 'Cardt' stands for the calibrated average of delta, rho, and THIMA: it averages over (1) a model in first differences ('delta'), with removal of the largest values and additional dampening; (2) a simple autoregressive model ('rho'), with protection against explosive roots and possibly in first differences with dampened mean; and (3) a trend-halved integrated moving average model (THIMA). The final calibration step treats the averaged forecasts as observations to re-estimate a model that includes the forecast period. The calibration model is a richer version of the first-stage autoregressive model, and the fitted values over the forecast period (now pseudo-in-sample) are the new forecasts. At this stage there is no issue with overfitting or explosive roots, because no further extrapolation is made. A small robustification adjustment is made to the first two forecasts.

Cardt is a useful benchmark method, which is fully documented, fast, and performs well on the M3 and M4 data sets.

Table 1
Average (unconditional) RMSE for $M = 10\,000, T = 45, \tau_1 = 30, \tau_2 = 40, \alpha_1 = 0.001, \alpha_2 = 0.01, \gamma_1 = \gamma_3 = 5$. All experiments use the same initial seed.

γ_2	ρ_1=ρ_3	ρ_2	DGP				MSIS			
			δ_1	δ_2	ϕ_1	ϕ_2	δ_1	δ_2	ϕ_1	ϕ_2
[1] Induced recession break $\delta_1 = \delta_2 = 0$										
5	0.9	0.9					0.29	1.00	0.01	0.02
5	0.9	0.8			0.01	0.02	5.42	2.62	0.11	0.10
5	0.9	0.7			0.02	0.02	8.29	2.37	0.17	0.13
5	0.9	0.6			0.02	0.03	10.66	2.11	0.21	0.17
5	0.9	0.5			0.02	0.03	11.94	1.84	0.24	0.21
5	0.9	0.4			0.02	0.03	11.45	2.87	0.23	0.40
[2] Induced boom break $\delta_1 = \delta_2 = 0$										
5	0.8	0.9			0.02	0.02	1.73	2.85	0.09	0.09
5	0.7	0.9			0.03	0.02	1.83	3.32	0.13	0.11
5	0.6	0.9			0.04	0.02	1.70	3.71	0.17	0.12
5	0.5	0.9			0.05	0.02	1.45	3.84	0.20	0.12
5	0.4	0.9			0.06	0.03	1.26	3.63	0.24	0.12
[3] Direct break $\phi_1 = \phi_2 = 0$										
4	0.9	0.9	0.71	0.66			1.31	2.51	0.02	0.05
3	0.9	0.9	0.71	0.63			2.69	4.39	0.04	0.10
2	0.9	0.9	0.69	0.63			4.19	4.18	0.08	0.11
1	0.9	0.9	0.68	0.63			4.99	2.92	0.10	0.09
0	0.9	0.9	0.67	0.63			5.35	2.42	0.10	0.09

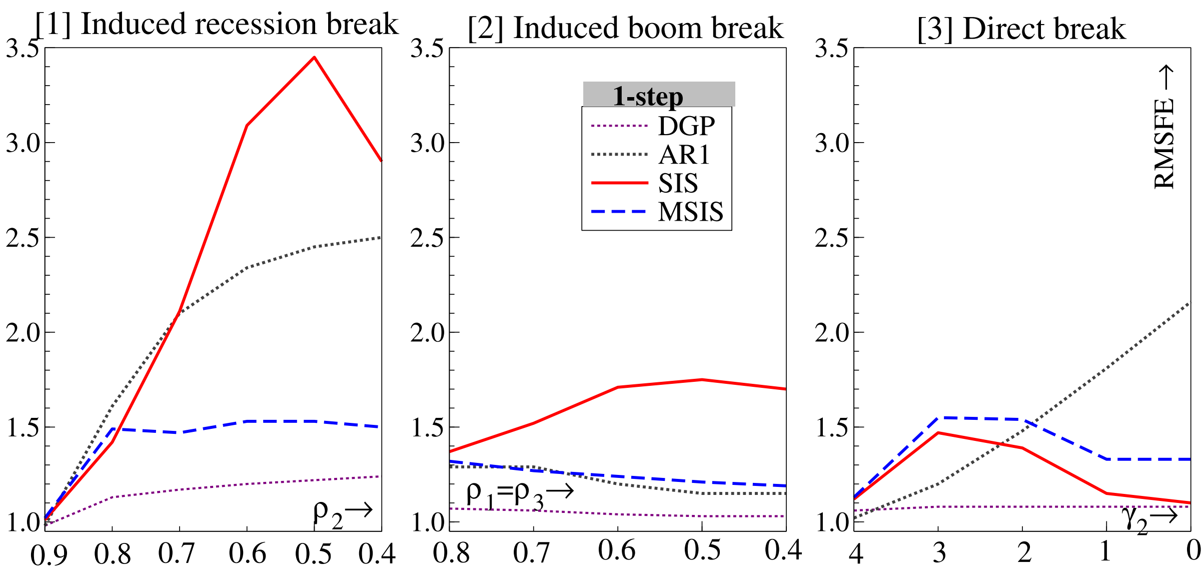

Fig. 5. One-step-ahead forecast RMSFEs, [1]: $\rho_1 = \rho_3 = 0.9, \rho_2 = 0.9$; [2]: $\rho_2 = 0.9, \gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9, \gamma_2 = 5$.

Figs. 5–7 respectively show the Monte Carlo averages of the root mean square forecast errors (RMSFEs) for three of the forecasting methods: one-step ($h = 1$), robust forecasts for $h = 1$, and dynamic forecasts for $h = 1, \ldots, 5$.

For one step-ahead, AR1 (a regular comparator for forecasting methods) can be best here, but can also be worst. SIS does well for direct shifts, but is usually worst for induced shifts, whereas MSIS is generally reliable, with little additional cost for direct shifts.

The RMSFEs of the four robust one-step-ahead forecasts in Fig. 6 are relatively close and much smaller than the corresponding conventional one-step forecasts about half the time. Although the robust AR1 is always the worst, its RMSFE is an improvement on the cases where it was worst in Fig. 5. Robust SIS does well for direct shifts, and for shifts induced by ρ_2 changing. Generally, robust MSIS provides a low-risk choice, usually close to the robust version of the estimated DGP. But the disadvantages of robustifying unnecessarily are shown by the robust DGP often no longer being the best.

The multi-step (dynamic) forecasts from MSIS are one of the lowest-risk methods, whereas AR1 is often the best for one step ahead. Cardt is sometimes the best for multi-step forecasts despite using only the data properties of the time series.

Overall, the formulation that is closest to the DGP specification does best or nearly best. Lacking knowledge of the underlying DGP, although no approach then dominates (a common finding: see Castle et al., 2021a), the

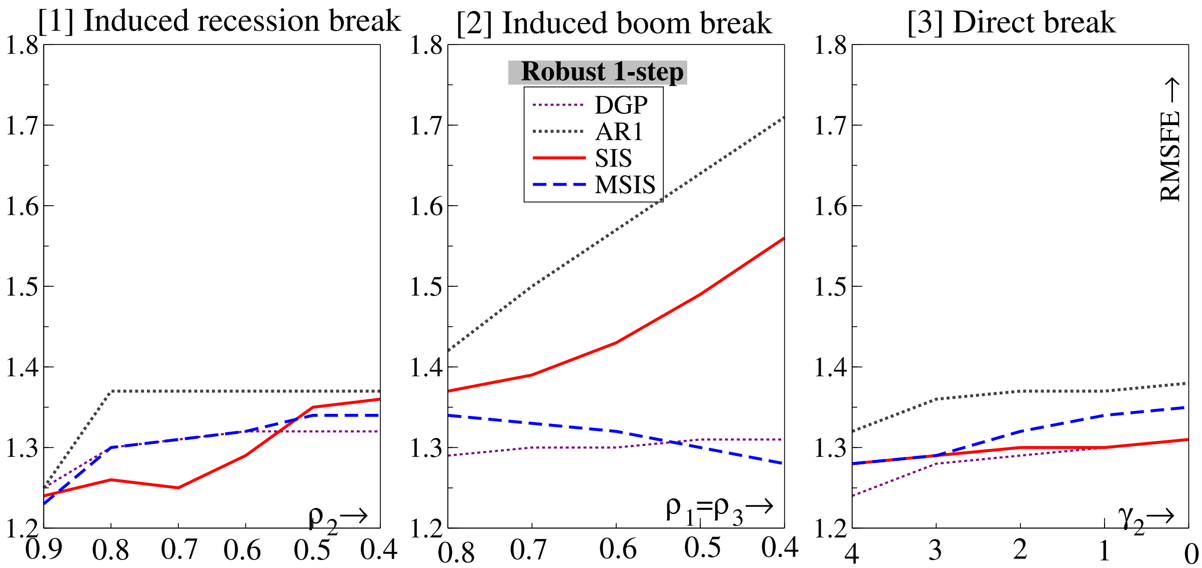

Fig. 6. Robust one-step-ahead forecast RMSFEs, [1]: $\rho_1 = \rho_3 = 0.9$, $\rho_2 = 0.9$; [2]: $\rho_2 = 0.9$, $\gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9$, $\gamma_2 = 5$.

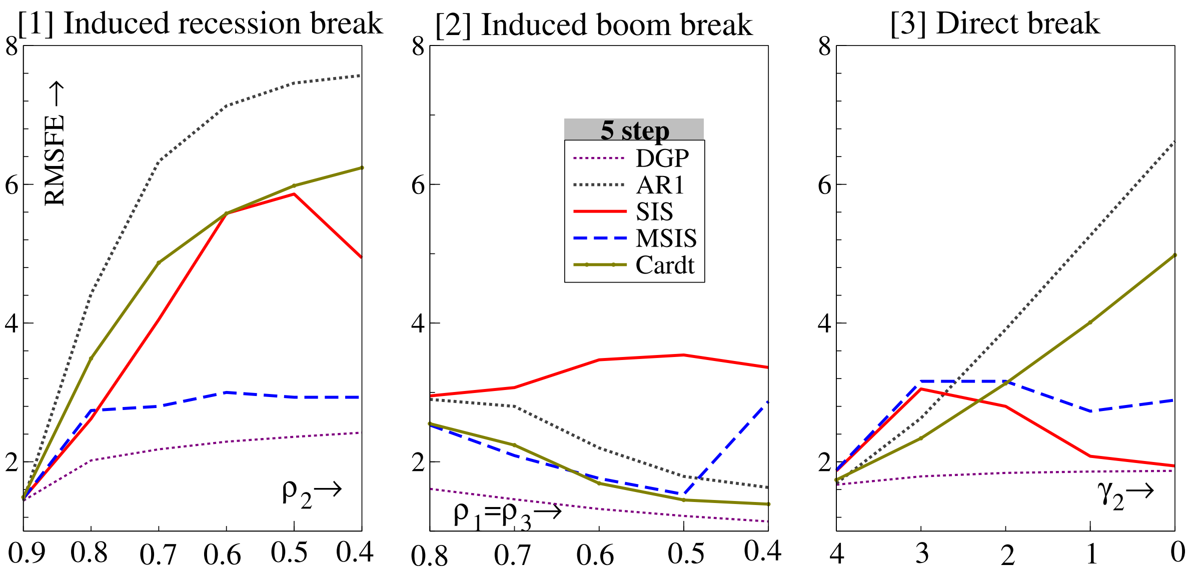

Fig. 7. Five-step-ahead forecast RMSFEs, [1]: $\rho_1 = \rho_3 = 0.9$, $\rho_2 = 0.9$; [2]: $\rho_2 = 0.9$, $\gamma_2 = 5$; [3]: $\rho_1 = \rho_3 = 0.9$, $\gamma_2 = 5$.

robust version of MSIS provides a low-risk choice, usually close to the robust version of the estimated DGP and never the worst.

5. An expectations-augmented phillips curve

We consider the expectations-augmented Phillips curve examined by Bai and Perron (2003):

$$\Delta w_t = \beta_0 + \beta_1 \mathsf{E}_{t-1}(\Delta p_t | I_{t-1}) + \beta_2 \Delta u_t + \beta_3 u_{t-1} + \epsilon_t \quad (32)$$

for $t = 1, \ldots, T$, and $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$. w_t is the log of nominal wages, p_t is the log of the GDP deflator, and u_t is the unemployment rate. Alogoskoufis and Smith (1991) assume that inflation follows an AR(1) process, such that

$$\mathsf{E}_{t-1}(\Delta p_t | I_{t-1}) = \gamma_0 + \gamma_1 \Delta p_{t-1} \quad (33)$$

which, substituting into (32) results in the Phillips curve,

$$\Delta w_t = \lambda_0 + \lambda_1 \Delta p_{t-1} + \beta_2 \Delta u_t + \beta_3 u_{t-1} + \epsilon_t \quad (34)$$

where $\lambda_0 = \beta_0 + \beta_1 \gamma_0$ and $\lambda_1 = \beta_1 \gamma_1$.

Estimating (34) on UK data over 1862–2009, with 2010–2018 withheld for nine conditional forecasts,

$$\widehat{\Delta w_t} = \underset{(0.006)}{0.027} + \underset{(0.060)}{0.833} \Delta p_{t-1} - \underset{(0.195)}{1.182} \Delta u_t - \underset{(0.084)}{0.160} u_{t-1}$$

$$\widehat{\sigma} = 3.8\%, \quad \mathsf{R}^2 = 0.615, \quad \mathsf{F}_{\mathsf{ar}}(2, 142) = 1.6,$$

$$\mathsf{F}_{\mathsf{arch}}(1, 146) = 10.6^{**}, \quad \chi^2_{\mathsf{nd}}(2) = 98.0^{**},$$

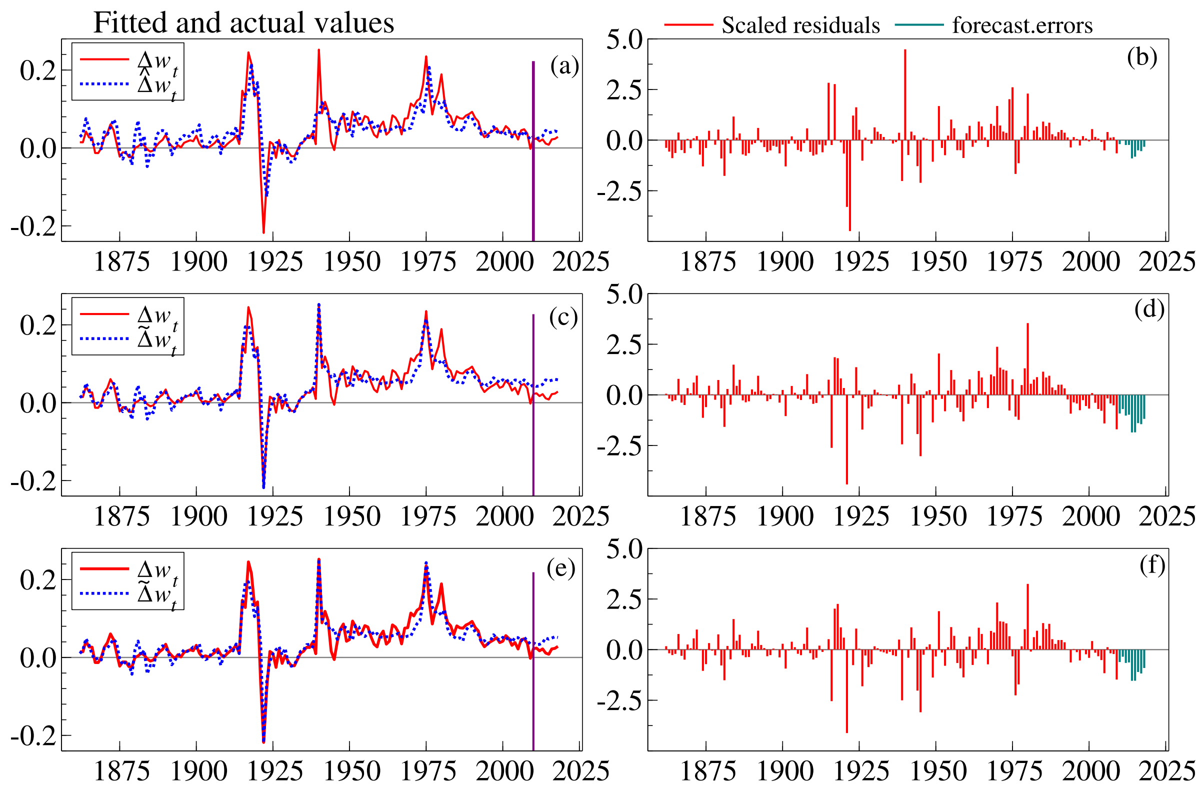

Fig. 8. (a) Model fit and forecasts, (b) residuals (35), (c) model fit and forecasts, (d) residuals for (36), (e) model fit and forecasts, and (f) residuals for (37).

$$\mathsf{F}_{\text{het}}(6, 141) = 4.0^{**}, \ \mathsf{F}_{\text{reset}}(2, 142) = 1.8,$$
$$\mathsf{F}_{\text{Chow}}(9, 144) = 0.22, \ T = 1862{-}2009. \qquad (35)$$

Coefficient standard errors are in parentheses, $\widehat{\sigma}$ is the residual standard deviation, R^2 is the coefficient of multiple correlation, F_{ar} tests residual autocorrelation (see Godfrey, 1978), F_{arch} tests autoregressive conditional heteroskedasticity (see Engle, 1982), F_{het} tests residual heteroskedasticity (see White, 1980), $\chi^2_{\text{nd}}(2)$ tests non-normality (see Doornik & Hansen, 2008), F_{Chow} is a parameter constancy forecast test over the last nine data points (see Chow, 1960), and $\mathsf{F}_{\text{reset}}$ tests non-linearity (see Ramsey, 1969). The model is badly mis-specified with significant outliers, as seen in Fig. 8(b), although the forecast Chow test does not indicate forecast failure.

We next apply SIS at a selection significance level of 0.0001, holding all regressors in (35) fixed. This results in four step shifts being retained for 1920, 1922, 1939, and 1940. The pairs of step shifts have almost equal but opposite signs, and so these were replaced by impulse indicators for 1922 and 1940. Including these impulse indicators in the GUM as fixed regressors, SIS was again applied but this time at the slightly looser level of 0.001. The resulting model retained four additional step indicators. The model is reported in (36):

$$\widehat{\Delta w_t} = \underset{(0.009)}{0.054} + \underset{(0.051)}{0.573}\Delta p_{t-1} - \underset{(0.136)}{0.943}\Delta u_t - \underset{(0.068)}{0.232}u_{t-1}$$
$$- \underset{(0.016)}{0.101}S_{1914} + \underset{(0.015)}{0.080}S_{1917} - \underset{(0.019)}{0.104}S_{1973}$$
$$+ \underset{(0.019)}{0.090}S_{1975} - \underset{(0.026)}{0.192}I_{1922} + \underset{(0.026)}{0.177}I_{1940}$$
$$\widehat{\sigma} = 2.6\%, \quad R^2 = 0.828, \quad \mathsf{F}_{\text{ar}}(2, 136) = 4.5^{*},$$
$$\mathsf{F}_{\text{arch}}(1, 146) = 0.1, \ \chi^2_{\text{nd}}(2) = 42.6^{**},$$
$$\mathsf{F}_{\text{het}}(10, 135) = 13.1^{**}, \ \mathsf{F}_{\text{reset}}(2, 136) = 2.5,$$
$$\mathsf{F}_{\text{Chow}}(9, 138) = 1.3, \ T = 1862{-}2009. \qquad (36)$$

The model is still badly mis-specified, failing normality, heteroskedasticity, and autocorrelation, which is unsurprising given the simplistic nature of (32). Fig. 8(c) and Fig. 8(d) show the forecasts are poorer with saturation applied, compared to ignoring shifts and outliers, although the forecast Chow test does not indicate forecast failure, despite the nine forecasts lying substantially above the outturns. These one-step forecasts condition on known Δu at t.

Finally, we create four interaction terms by interacting the retained step indicators with Δp_{t-1} and add these to the GUM. The economic variables, intercept, and two impulse indicators are fixed in the model, and the eight regressors comprising the four step indicators and four multiplicative indicators are selected over using a selection significance level of 0.01. The final selected model is

$$\widehat{\Delta w_t} = \underset{(0.006)}{0.045} + \underset{(0.066)}{0.736}\Delta p_{t-1} - \underset{(0.134)}{0.966}\Delta u_t$$
$$- \underset{(0.067)}{0.252}u_{t-1} - \underset{(0.026)}{0.200}I_{1922} + \underset{(0.026)}{0.175}I_{1940}$$

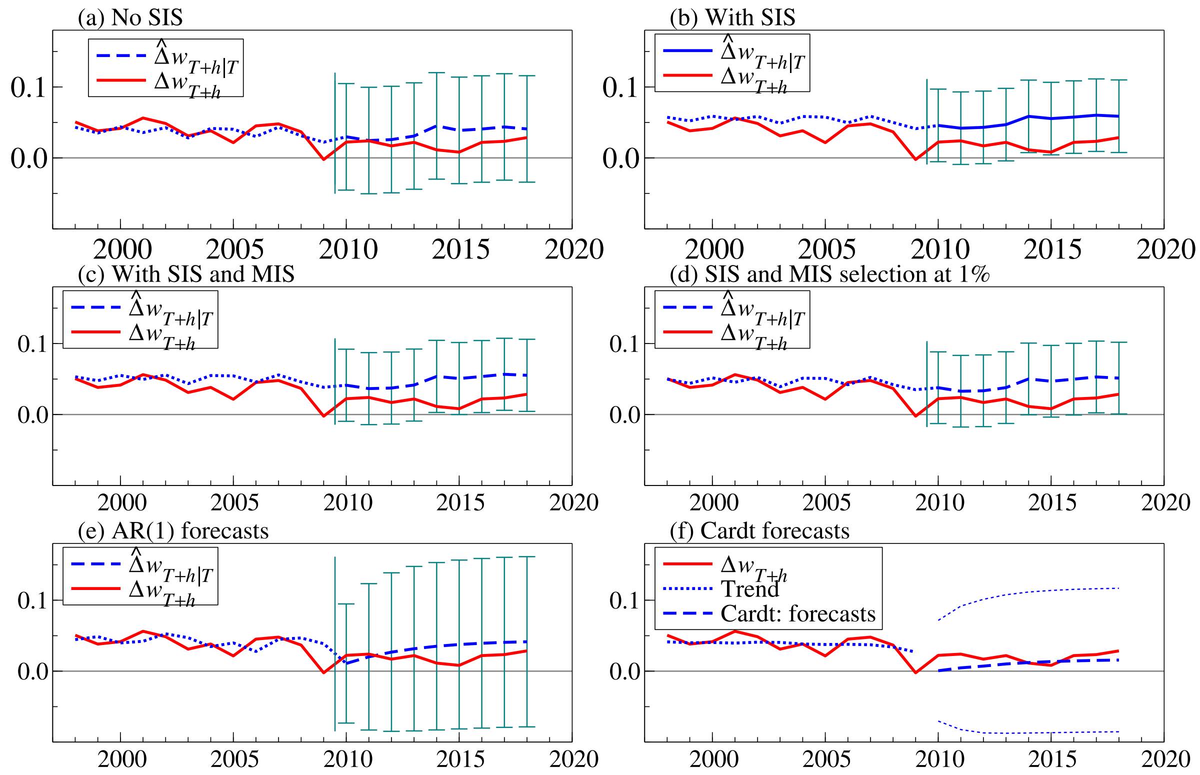

Fig. 9. Dynamic conditional forecasts for the Phillips curve.

$$- \underset{(0.016)}{0.105} S_{1914} + \underset{(0.016)}{0.081} S_{1917}$$
$$- \underset{(0.153)}{0.892} S_{1973} \times \Delta p_{t-1} + \underset{(0.159)}{0.664} S_{1975} \times \Delta p_{t-1}$$
$$\widehat{\sigma} = 2.5\%, \quad R^2 = 0.832, \quad \mathsf{F}_{\mathrm{ar}}(2, 136) = 5.6^{**},$$
$$\mathsf{F}_{\mathrm{arch}}(1, 146) = 0.6, \quad \chi^2_{\mathrm{nd}}(2) = 30.9^{**},$$
$$\mathsf{F}_{\mathrm{het}}(12, 133) = 11.5^{**}, \quad \mathsf{F}_{\mathrm{reset}}(2, 136) = 4.1^{*},$$
$$\mathsf{F}_{\mathrm{Chow}}(9, 138) = 0.91, \quad T = 1862\text{–}2009. \tag{37}$$

The step indicators for WWI are retained rather than the interaction terms, indicating a direct shift. These effects are almost equal and offsetting. However, for the oil crises the interaction indicators are retained rather than the step shifts, suggesting that the oil crises changed inflation dynamics which feed into wage inflation. However, we cannot conclude from this analysis that the formation of inflation expectations in the Phillips curve shifted in the oil crisis, as the model is mis-specified, and omitted variables will be influencing the parameter estimates. Fig. 8(e) and Fig. 8(f) record the outcome.

Finally, we test if we can replace S_{1914} and S_{1917} with a dummy variable that takes a value of 1 for 1914–1917 and 0 otherwise. The encompassing test rejects this restriction, $\mathsf{F}(1, 138) = 34.4^{**}$. The restriction to replace $S_{1973} \times \Delta p_{t-1}$ and $S_{1975} \times \Delta p_{t-1}$ with a dummy variable taking a value of 1 over 1973–1975 interacted with lagged inflation, and 0 otherwise, is also rejected, $\mathsf{F}(1, 138) = 10.6^{**}$.

Table 2 reports the RMSFE and mean absolute percentage error (MAPE) over 2010–2018 for the models, along with two benchmarks—an AR(1) for nominal wage inflation and Cardt— using the same sample. The benchmarks provide much more accurate forecasts, although with wider error bands. Of the Phillips curve models, not accounting for breaks improves the forecast performance, as with Fig. 1. There is a benefit to distinguishing between direct and induced breaks in terms of forecast performance, despite these breaks occurring decades before the forecast period (see Fig. 9).

Table 2
Forecast results.

	RMSFE %	MAPE
No SIS (35)	1.89	113.0
With SIS (36)	3.36	207.5
With SIS and MIS	2.94	180.3
SIS and MIS selection (37)	2.59	158.6
AR(1)	1.68	104.3
Cardt	1.25	53.0

6. An unemployment model

The long-run behaviour of the UK unemployment rate $(U_{r,t})$ provides a second case study of whether location shifts were direct or induced. The data are from Castle, Hendry, and Martinez (2022) based on the model in Hendry (2001) for the long-run annual time series where $R_{r,t} = (R_L - \Delta p - \Delta y)_t$ is a proxy for the profitability of hiring, R_L is the nominal long-term interest rate, Δp_t is the inflation rate, and Δy_t is the growth rate of GDP.

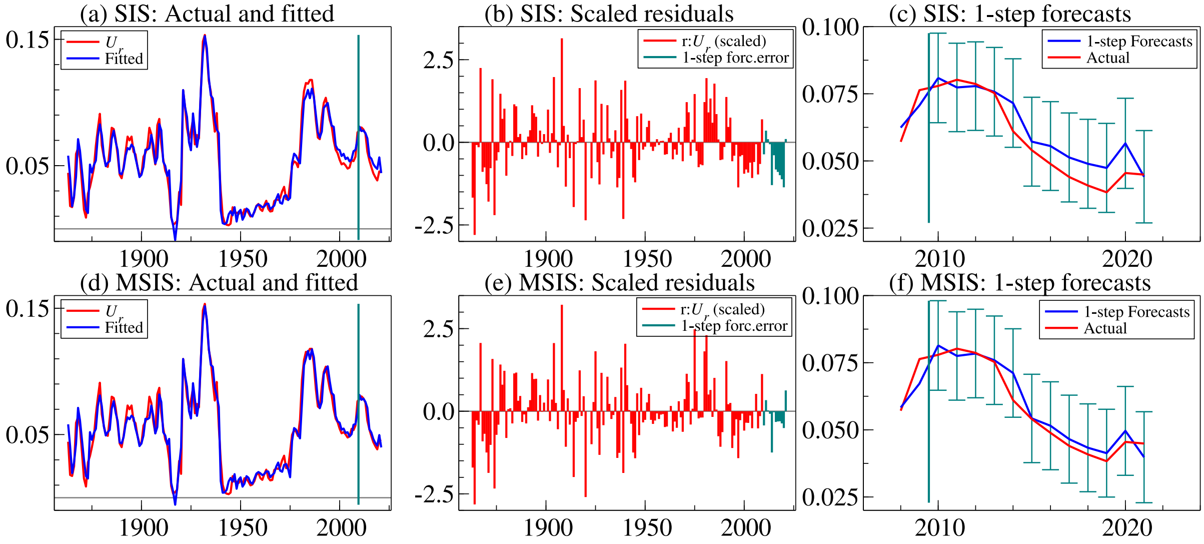

Fig. 10. Models of UK unemployment: SIS in (a), (b), and (c); MSIS in (d), (e), and (f).

Estimates of the model without any indicators delivers

$$\widehat{U}_{r,t} = \underset{(0.08)}{1.02}\ U_{r,t-1} - \underset{(0.07)}{0.15}\ U_{r,t-2} + \underset{(0.002)}{0.007} + \underset{(0.02)}{0.21}\ R_{r,t} - \underset{(0.03)}{0.11}\ R_{r,t-1} \quad (38)$$

$$\widehat{\sigma} = 1.06\% \quad \mathsf{R}^2 = 0.90 \quad \mathsf{F}_{\mathrm{ar}}(2, 140) = 0.38$$
$$\mathsf{F}_{\mathrm{arch}}(1, 145) = 8.86^{**} \quad \mathsf{F}_{\mathrm{Het}}(8, 138) = 11.3^{**}$$
$$\chi^2_{\mathrm{nd}}(2) = 17.7^{**} \quad \mathsf{F}_{\mathrm{Reset}}(2, 140) = 4.41^{*}$$
$$\mathsf{F}_{\mathrm{Chow}}(12, 142) = 0.37$$
$$T = 1862 - 2009 \quad \mathrm{RMSFE} = 0.67\%$$

These are similar to the estimates in Castle et al. (2022). Selecting step indicators by SIS at $\alpha = 0.1\%$ yields

$$\widehat{U}_{r,t} = \underset{(0.07)}{1.06}\ U_{r,t-1} - \underset{(0.06)}{0.34}\ U_{r,t-2} + \underset{(0.003)}{0.021} + \underset{(0.02)}{0.15}\ R_{r,t} - \underset{(0.021)}{0.044}\ R_{r,t-1} - \underset{(0.010)}{0.051}\ S_{1920} + \underset{(0.014)}{0.096}\ S_{1921} - \underset{(0.010)}{0.048}\ S_{1922} - \underset{(0.009)}{0.036}\ S_{1929} + \underset{(0.009)}{0.024}\ S_{1930} + \underset{(0.005)}{0.023}\ S_{1938} - \underset{(0.003)}{0.015}\ S_{1974} \quad (39)$$

$$\widehat{\sigma} = 0.81\% \quad \mathsf{R}^2 = 0.95 \quad \mathsf{F}_{\mathrm{ar}}(2, 133) = 2.35$$
$$\mathsf{F}_{\mathrm{arch}}(1, 145) = 0.43 \quad \mathsf{F}_{\mathrm{Het}}(12, 131) = 1.21$$
$$\chi^2_{\mathrm{nd}}(2) = 4.84 \quad \mathsf{F}_{\mathrm{Reset}}(2, 133) = 4.43^{*}$$
$$\mathsf{F}_{\mathrm{Chow}}(12, 135) = 0.54$$
$$T = 1862 - 2009 \quad \mathrm{RMSFE} = 0.65\%$$

Fig. 10(a–c) record the graphical outcomes.

Although seven step indicators are significant, the group for 1920–1922 is offsetting, and the pair for 1929–1930 substantially cancels. As 1974 is the closest step indicator to the forecast period, we added the mutiplicative indicator for $S_{1974} \times U_{r,t-1}$ to (39) and selected at 5%, which yielded

$$\widehat{U}_{r,t} = \underset{(0.07)}{1.24}\ U_{r,t-1} - \underset{(0.06)}{0.35}\ U_{r,t-2} + \underset{(0.002)}{0.008} + \underset{(0.02)}{0.14}\ R_{r,t} - \underset{(0.02)}{0.036}\ R_{r,t-1} - \underset{(0.009)}{0.051}\ S_{1920} + \underset{(0.014)}{0.094}\ S_{1921} - \underset{(0.010)}{0.047}\ S_{1922} - \underset{(0.009)}{0.036}\ S_{1929} + \underset{(0.009)}{0.021}\ S_{1930} + \underset{(0.005)}{0.029}\ S_{1938} - \underset{(0.041)}{0.23}\ S_{1974} \times U_{r,t-1} \quad (40)$$

$$\widehat{\sigma} = 0.81\% \quad \mathsf{R}^2 = 0.95 \quad \mathsf{F}_{\mathrm{ar}}(2, 133) = 2.38^{*}$$
$$\mathsf{F}_{\mathrm{arch}}(1, 145) = 0.26 \quad \mathsf{F}_{\mathrm{Het}}(13, 130) = 1.16$$
$$\chi^2_{\mathrm{nd}}(2) = 9.33^{*} \quad \mathsf{F}_{\mathrm{Reset}}(2, 133) = 0.25$$
$$\mathsf{F}_{\mathrm{Chow}}(12, 135) = 0.24$$
$$T = 1862 - 2009 \quad \mathrm{RMSFE} = 0.40\%$$

Thus, the multiplicative indicator was selected and the step eliminated, suggesting that the mid-1970s also saw a change in the dynamics of unemployment during the first oil crisis. The forecasts have an RMSFE roughly a third smaller than either previous specification, and Fig. 10(d–f) show how close to the outcomes these are, as well as having smaller interval forecasts. Despite ending in 1974 and so being decades before the forecast origin, the interaction term has a huge effect on forecast performance by substantively altering the estimated dynamics of unemployment.

7. Explaining the 'teaser' forecasts of UK DGP

The massive drop in UK GDP from mandated lockdowns at the start of the pandemic seems to be yet a different type of shock, and the solution described above

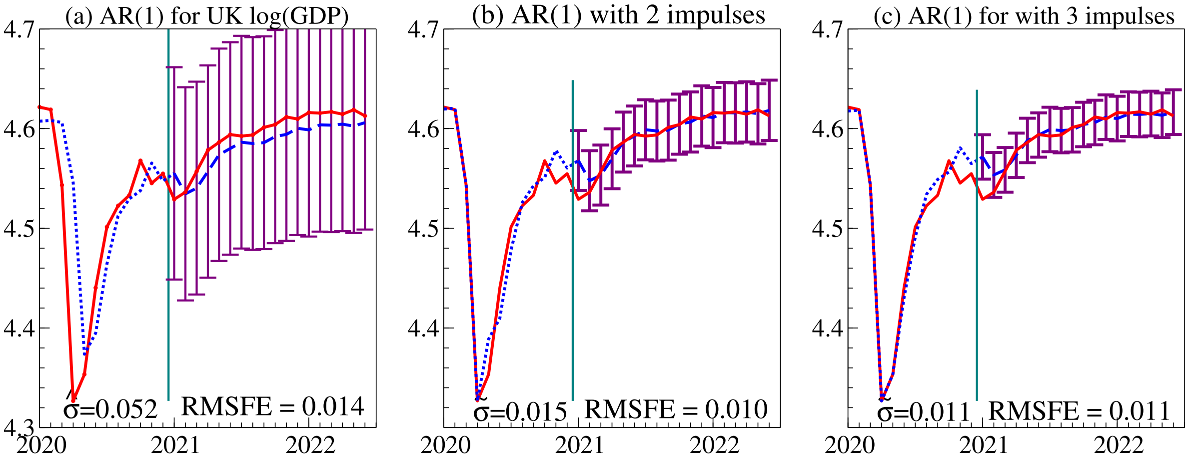

Fig. 11. Actual and fitted values (until the end of 2020) and actual and one-step forecasts of UK log(GDP).

does not correct the bad forecasts. However, dropping all the steps and including just impulse indicators for 2020(3) and 2020(4) (or also 2020(5)) based on a separate analysis of the impact of Covid-19 on UK GDP using quarterly data in Castle, Doornik, and Hendry (2023), delivered the models in (41) and (42) and the outcomes in Fig. 11(b) and (c).

The AR(1) estimates in Figs. 1(a) and (b), replicated in Fig. 11(a), for monthly UK DGP are

$$\log(\mathrm{GDP}) = \underset{(0.13)}{0.796} \log(\mathrm{GDP})_{t-1} + \underset{(0.58)}{0.931}.$$

The equation with the two impulse indicators for the pandemic crash is

$$\log(\mathrm{GDP}) = \underset{(0.036)}{0.786} \log(\mathrm{GDP})_{t-1} + \underset{(0.17)}{0.990} - \underset{(0.015)}{0.075}\ 1_{2020(3)} - \underset{(0.015)}{0.232}\ 1_{2020(4)}. \tag{41}$$

If the indicator for 2020(5) easing is also included,

$$\log(\mathrm{GDP}) = \underset{(0.034)}{0.701} \log(\mathrm{GDP})_{t-1} + \underset{(0.15)}{1.38} - \underset{(0.011)}{0.073}\ 1_{2020(3)} - \underset{(0.011)}{0.237}\ 1_{2020(4)} - \underset{(0.014)}{0.058}\ 1_{2020(5)}. \tag{42}$$

Although the very poor forecasts of UK GDP over the Covid-19 pandemic lockdowns after selection by IIS+SIS prompted the above analysis, the outcome transpires not to be from changed dynamics, and instead emphasises that empirical modelling can often require judgement.

8. Conclusion

Examining the properties of forecasting following in-sample shifts induced by changes in the dynamics revealed that the initial value of the dependent variable at the forecast origin could be far from the equilibrium mean in the data generation process (DGP) at that time, leading to very poor forecasts. The DGP could be better approximated by capturing the induced location shifts using SIS, but that did not improve forecasts, with asymmetrical errors being worse when the autoregressive parameter was reduced from a high to a low value and then back to the original high value than in the reverse order. We showed that, since the location shifts were induced by changes in dynamics, augmenting the model by the step indicators that interacted with the lagged regressand helped discriminate between direct and induced location shifts, improving both the model as an approximation to the DGP and its forecasts. The impact on inflation of the two World Wars was confirmed as being location shifts, whereas the oil price shocks in the 1970s induced changes in inflation dynamics which impacted on wage expectations and changed the dynamic behaviour of the unemployment rate.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: David Hendry reports financial support was provided by Nuffield College. Jurgen Doornik reports a relationship with Nuffield College that includes: employment. Royalties received from the software used.

Acknowledgments

Financial support from the Robertson Foundation, United States (award 9907422) and Nuffield College, UK is gratefully acknowledged, as are helpful comments from two referees and the editor, Andrew B. Martinez, and the participants of the International Symposium of Forecasting, the International Conference on Economic Modeling and Data Science, and the 25th Dynamic Econometrics conference.

Data and code sharing

All calculations and graphs use PcGive (Doornik & Hendry, 2021), Ox Professional (Doornik, 2021a) and OxMetrics (Doornik, 2021b). Data and code to replicate the

Table A.3
Average $\widehat{\sigma}$ for $M = 10\,000, T = 45, \tau_1 = 30, \tau_2 = 40, \alpha_1 = 0.001, \alpha_2 = 0.01, \gamma_1 = \gamma_3 = 5$.

γ_2	$\rho_1 = \rho_3$	ρ_2	DGP	AR1	SIS	MSIS
5	0.9	0.9	0.99	0.99	0.97	0.97
5	0.9	0.8	0.99	1.73	1.00	0.99
5	0.9	0.7	0.99	2.51	1.00	0.96
5	0.9	0.6	0.99	3.17	1.04	0.96
5	0.9	0.5	0.99	3.79	1.06	0.97
5	0.9	0.4	0.99	4.39	1.06	0.97
5	0.8	0.9	0.99	1.53	1.06	1.04
5	0.7	0.9	0.99	2.03	1.04	0.98
5	0.6	0.9	0.99	2.45	1.08	0.98
5	0.5	0.9	0.99	2.82	1.12	0.99
5	0.4	0.9	0.99	3.18	1.15	0.99
4	0.9	0.9	0.98	1.08	1.03	1.03
3	0.9	0.9	0.99	1.26	1.08	1.07
2	0.9	0.9	0.99	1.51	1.03	1.03
1	0.9	0.9	0.99	1.79	1.00	0.99
0	0.9	0.9	0.99	2.10	0.99	0.98

Table A.4
Selection for $M = 10\,000, T = 45, \tau_1 = 30, \tau_2 = 40, \alpha_1 = 0.001, \alpha_2 = 0.01, \gamma_1 = \gamma_3 = 5$.

γ_2	$\rho_1 = \rho_3$	ρ_2	SIS				MSIS					
			S_{τ_1}%	S_{τ_2}%	gauge%	pot%	S_{τ_1}%	S_{τ_2}%	yS_{τ_1}%	yS_{τ_2}%	gge%	pot%
5	0.9	0.9	0.5	0.4	0.5	–	0.3	0.2	0.2	0.3	0.3	–
5	0.9	0.8	95.8	61.8	0.9	78.8	67.4	33.2	30.0	29.2	50.3	29.6
5	0.9	0.7	100.0	82.7	0.8	91.4	61.6	24.3	39.0	58.5	43.0	48.8
5	0.9	0.6	100.0	82.4	1.5	91.2	49.2	15.6	51.2	67.0	32.4	59.1
5	0.9	0.5	100.0	83.8	2.8	91.9	35.5	12.1	64.8	72.3	23.8	68.5
5	0.9	0.4	100.0	87.1	4.3	93.5	21.6	9.7	78.9	78.1	15.6	78.5
5	0.8	0.9	52.7	81.0	0.9	66.9	13.7	27.7	39.8	54.3	20.7	47.0
5	0.7	0.9	80.1	99.5	1.0	89.8	14.8	26.1	66.5	73.7	20.4	70.1
5	0.6	0.9	84.6	100.0	1.9	92.3	12.3	15.8	73.7	84.9	14.0	79.3
5	0.5	0.9	85.3	100.0	3.0	92.7	8.6	11.5	77.8	91.3	10.1	84.5
5	0.4	0.9	84.6	100.0	3.8	92.3	6.3	11.1	79.1	95.5	8.7	87.3
4	0.9	0.9	7.7	2.3	0.7	5.0	4.0	1.8	3.8	0.8	2.3	2.9
3	0.9	0.9	36.4	20.1	1.0	28.3	23.3	15.5	13.9	5.9	9.9	19.4
2	0.9	0.9	70.0	61.2	1.0	65.6	58.0	45.1	14.2	18.2	16.2	51.5
1	0.9	0.9	89.0	86.6	0.6	87.8	81.3	64.9	11.0	23.1	17.0	73.1
0	0.9	0.9	96.5	96.0	0.4	96.2	91.2	80.2	9.4	16.9	13.1	85.7

S_{τ_1}% etc. retention of S_{τ_1} in final model
gauge% SIS gauge (retention of S_j, $j \neq \tau_1, \tau_2$ if break in model)
pot% SIS potency (retention of S_{τ_1}, S_{τ_2} if break in model)
gge% MSIS gauge (conditional on SIS gauge selection extended to MSIS)
pot% MSIS potency (conditional on SIS potency selection extended to MSIS).

Table A.5
Average RMSFE for $H = 5, M = 10\,000, T = 45, \tau_1 = 30, \tau_2 = 40, \alpha_1 = 0.001, \alpha_2 = 0.01, \gamma_1 = \gamma_3 = 5$.

γ_2	$\rho_1 = \rho_3$	ρ_2	DGP			AR1			SIS			MSIS			
			1-step	rob	dyn	1-step	rob	dyn	1-step	rob	dyn	1-step	rob	dyn	Cardt
5	0.9	0.9	0.98	1.25	1.44	*0.98*	1.25	**1.44**	1.01	1.24	1.47	1.02	1.23	1.47	1.49
5	0.9	0.8	1.13	1.30	2.02	1.61	1.37	4.42	*1.42*	1.26	**2.62**	1.49	1.30	2.74	3.49
5	0.9	0.7	1.17	1.31	2.18	2.10	1.37	6.33	2.11	1.25	4.05	*1.47*	1.31	**2.80**	4.87
5	0.9	0.6	1.20	1.32	2.29	2.34	1.37	7.13	3.09	1.29	5.58	*1.53*	1.32	**3.00**	5.58
5	0.9	0.5	1.22	1.32	2.36	2.45	1.37	7.46	3.45	1.35	5.86	*1.53*	1.34	**2.93**	5.98
5	0.9	0.4	1.24	1.32	2.42	2.50	1.37	7.57	2.90	1.36	4.94	*1.50*	1.34	**2.93**	6.24
5	0.8	0.9	1.07	1.29	1.61	*1.29*	1.42	2.90	1.37	1.37	2.95	1.32	1.34	**2.53**	2.55
5	0.7	0.9	1.06	1.30	1.46	*1.29*	1.50	2.80	1.52	1.39	3.07	1.27	1.33	**2.09**	2.24
5	0.6	0.9	1.04	1.30	1.32	*1.20*	1.57	2.20	1.71	1.43	3.47	1.24	1.32	1.76	**1.69**
5	0.5	0.9	1.03	1.31	1.22	*1.15*	1.64	1.79	1.75	1.49	3.54	1.21	1.30	1.53	**1.45**
5	0.4	0.9	1.03	1.31	1.14	*1.15*	1.71	1.63	1.70	1.56	3.36	1.19	1.28	2.87	**1.39**
4	0.9	0.9	1.06	1.24	1.67	*1.02*	1.32	**1.69**	1.12	1.28	1.87	1.13	1.28	1.88	1.74
3	0.9	0.9	1.08	1.28	1.79	*1.20*	1.36	2.63	1.47	1.29	3.05	1.55	1.29	3.16	**2.34**
2	0.9	0.9	1.08	1.29	1.84	1.48	1.37	3.91	*1.39*	1.30	**2.80**	1.54	1.32	3.16	3.13
1	0.9	0.9	1.08	1.30	1.86	1.81	1.37	5.26	*1.15*	1.30	**2.08**	1.33	1.34	2.73	4.01
0	0.9	0.9	1.08	1.31	1.87	2.16	1.38	6.62	*1.10*	1.31	**1.94**	1.33	1.35	2.89	4.98

'dyn' is dynamic (five-step) forecasts, and 'rob' is robust one-step forecasts.
Italics denote the smallest RMSFE for one-step forecasts and bold the smallest for dynamic (five-step) forecasts.

results can be downloaded from https://www.doornik.com/research.html.

Appendix. Tables with simulation results

See Tables A.3–A.5.

References

Alogoskoufis, G. S., & Smith, R. (1991). The Phillips curve, the persistence of inflation, and the Lucas critique: Evidence from exchange-rate regimes. *American Economic Review*, *81*, 1254–1275, https://www.jstor.org/stable/2006916.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*, 47–78. http://dx.doi.org/10.2307/2998540.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*, 1–22. http://dx.doi.org/10.1002/jae.659.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015a). Robust approaches to forecasting. *International Journal of Forecasting*, *31*, 99–112. http://dx.doi.org/10.1016/j.ijforecast.2014.11.002.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2021a). Forecasting principles from experience with forecasting competitions. *Forecasting*, *3*(1), 138–165, https://www.mdpi.com/2571-9394/3/1/10.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2021b). Robust discovery of regression models. *Econometrics and Statistics*, *26*, 31–51. http://dx.doi.org/10.1016/j.ecosta.2021.05.004.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2023). The impact of covid on economic models – the case of UK GDP. *Working paper: Economics Department*, Oxford University.

Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015b). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, *3*(2), 240–264. http://dx.doi.org/10.3390/econometrics3020240.

Castle, J., Hendry, D., & Martinez, A. (2022). *Historical role of energy in UK inflation and productivity and implications for price inflation in 2022*: *Discussion paper: Department of Economics*, University of Oxford, https://www.economics.ox.ac.uk/publication/1277963/hyrax.

Castle, J. L., & Shephard, N. (Eds.), (2009). *The methodology and practice of econometrics*. Oxford: Oxford University Press.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*, 591–605. http://dx.doi.org/10.2307/1910133.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.

Doornik, J. A. (2009). Autometrics. In Castle and Shephard (2009).

Doornik, J. A. (2021a). *Object-Oriented Matrix Programming using Ox (9th ed.)*. London: Timberlake Consultants Press.

Doornik, J. A. (2021b). *OxMetrics: An Interface to Empirical Modelling (9th ed.)*. London: Timberlake Consultants Press.

Doornik, J. A., Castle, J. L., & Hendry, D. F. (2020). Card forecasts for M4. *International Journal of Forecasting*, *36*, 129–134. http://dx.doi.org/10.1016/j.ijforecast.2019.03.012.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939. http://dx.doi.org/10.1111/j.1468-0084.2008.00537.x.

Doornik, J. A., & Hendry, D. F. (2021). *Empirical Econometric Modelling using PcGive: Volume I (9th ed.)*. London: Timberlake Consultants Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007. http://dx.doi.org/10.2307/1912773.

Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, *76*, 669–705. http://dx.doi.org/10.1111/j.1467-937X.2009.00545.x.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, *46*, 1303–1313. http://dx.doi.org/10.2307/1913830.

Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, *16*, 255–275. http://dx.doi.org/10.1002/jae.615.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, *135*, 399–426. http://dx.doi.org/10.1016/j.jeconom.2005.07.029.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, *23*, 317–335. http://dx.doi.org/10.1007/s00180-007-0054-z, Erratum, 337–339.

Hendry, D. F., & Neale, A. J. (1991). A Monte Carlo study of the effects of structural breaks on tests for unit roots. In P. Hackl, & A. H. Westlund (Eds.), *Economic structural change, analysis and forecasting* (pp. 95–119). Berlin: Springer-Verlag.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*, 55–67. http://dx.doi.org/10.1016/j.jeconom.2016.03.006.

Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. http://dx.doi.org/10.1093/acprof:oso/9780199237197.003.0001, In Castle & Shephard (2009).

Martinez, A. B., Castle, J. L., & Hendry, D. F. (2022). Smooth robust multi-horizon forecasts. *Advances in Econometrics*, *43A*, 143–165. http://dx.doi.org/10.1108/S0731-90532021000043A008.

Pesaran, M. H., Pick, A., & Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, *177*, 134–152. http://dx.doi.org/10.1016/j.jeconom.2013.04.002.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*, 134–161. http://dx.doi.org/10.1016/j.jeconom.2006.03.010.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, *31*, 350–371, https://www.jstor.org/stable/2984219.

Sims, C. A., Stock, J. H., & Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, *58*, 113–144.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838. http://dx.doi.org/10.2307/1912934.