Smooth Robust Multi-Horizon Forecasts

Andrew B. Martinez*, Jennifer L. Castle†, David F. Hendry‡

December 21, 2020

Abstract

We investigate whether smooth robust methods for forecasting can help mitigate pronounced and persistent failure across multiple forecast horizons. We demonstrate that naive predictors are interpretable as local estimators of the long-run relationship with the advantage of adapting quickly after a break, but at a cost of additional forecast error variance. Smoothing over naive estimates helps retain these advantages while reducing the costs, especially for longer forecast horizons. We derive the performance of these predictors after a location shift, and confirm the results using simulations. We apply smooth methods to forecasts of U.K. productivity and U.S. 10-year Treasury yields and show that they can dramatically reduce persistent forecast failure exhibited by forecasts from macroeconomic models and professional forecasters.

Keywords: Location Shifts; Long differencing; Productivity forecasts; Robust forecasts.

JEL classifications: C51, C53.

*Office of Macroeconomic Analysis, U.S. Department of the Treasury; H.O. Stekler Research Program on Forecasting, George Washington University; Climate Econometrics, Nuffield College. The views expressed here are those of the authors and not necessarily those of the Treasury Department or the U.S. Government. This research was supported in part by a grant from the Robertson Foundation (grant 9907422). Thanks to participants at the 22nd Dynamic Econometrics Conference and the 40th International Symposium on Forecasting, and to Allan Timmermann and an anonymous referee for helpful comments and suggestions. Contact: Andrew.Martinez@treasury.gov.

†Magdalen College, Climate Econometrics, and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Contact: Jennifer.Castle@magd.ox.ac.uk.

‡Nuffield College, Climate Econometrics, and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Contact: David.Hendry@nuffield.ox.ac.uk.

1 Introduction

Forecasts following structural breaks often experience bouts of large systematic forecast errors, i.e. forecast failure. This was first diagnosed by Smith (1929) and highlighted by Nelson (1972) and Cooper and Nelson (1975) who showed that 'naive' time-series models could outperform large macro-econometric systems. As is now well known, most econometric specifications are equilibrium correction and, as demonstrated by Clements and Hendry (1999), shifts in the mean of that term are the primary source of forecast failure. Consequently, adjusting for changes in that mean can be useful to offset failures that would otherwise occur. Intercept corrections can do so in some settings (see Hendry and Clements, 1994), as can differencing (see Hendry, 2006), both with a cost in forecast-error variances. Castle et al. (2015a) extended these ideas to averaging over a small number of periods after the shift to estimate the necessary corrections.

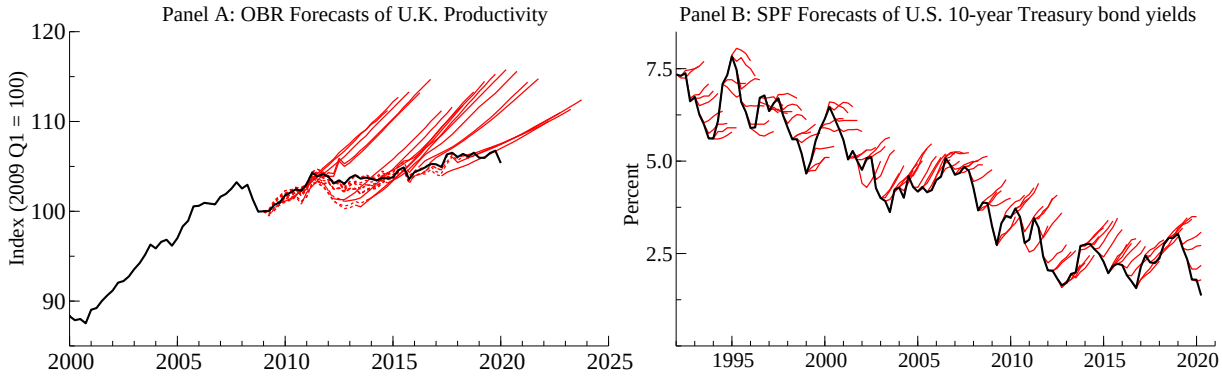


Figure 1.1: Examples of Persistent Forecast Failure

Despite methods available to robustify forecasts to avoid persistent forecast bias, prominent forecasters still exhibit bouts of pronounced and prolonged forecast failure. Panel A of Figure 1.1 shows that the Office of Budget Responsibility (OBR) five-year-ahead forecasts of U.K. productivity over the last decade were persistently above actual productivity as they returned to the pre-2008 trend. Panel B shows that Surveys of Professional Forecasters (SPF) have consistently over-predicted U.S. Treasury bond yields for the past 25 years. This shows that forecast failure is widespread.

In this paper we demonstrate that many of the most egregious and persistent forecast failures can be prevented after a break has occurred with a simple class of smooth robust methods. We start by showing that many of the existing robust predictors can be reinterpreted as local estimates of the long-run mean based on internal information available since the structural break. By smoothing

over these local post-break estimates, we can reduce the costs associated with these predictors at longer forecast horizons namely, by reducing forecast error variances. These methods also emulate the benefits of other techniques for forecasting after structural breaks; see Pesaran and Timmermann (2007), Pesaran et al. (2013), Giraitis et al. (2013) and Inoue et al. (2017).

The structure of the paper is as follows. Section 2 considers multi-step forecasting after a structural break with a misspecified autoregressive model that often does well in forecast comparisons. While we consider a univariate model in the derivations, the results generalize to multivariate models. Section 3 discusses commonly used robust methods, such as random walks, which are an example of differencing to convert a location shift to an impulse, and other robustified predictors. Section 4 illustrates how these predictors are akin to updating the long-run mean, while section 5 introduces the smoothed robust predictor. Section 6 simulates the new methods and section 7 considers the OBR's forecasts of U.K. productivity and the SPF's forecasts of U.S. long-term interest rates to demonstrate how substantial the improvements from these methods can be. Finally, section 8 concludes by discussing how these methods could be useful following the COVID-19 pandemic.

2 Forecasting after a break with a misspecified model

We start by considering how typical dynamic forecast models perform after a break if they fail to account for that break. We separate the forecast horizon 'h' from the length of time that has passed since the break 'b' at time T. This allows us to consider multi-horizon forecasts while also understanding how moving past the break date matters. We assume that the data generation process (DGP) for $t = \{1, \dots, T-1\}$ follows an autoregressive process with one lag, i.e an AR(1)

$$x_t = \mu + \rho\left(x_{t-1} - \mu\right) + \epsilon_t, \tag{2.1}$$

where μ represents the long-run or equilibrium mean, ρ represents the dynamics, which we assume are $|\rho| < 1$ unless otherwise specified to ensure that we have an I(0) process, and ϵ_t represents the independent and identically distributed innovations. This setup and the following analysis generalizes to a multivariate setting using a vector autoregression. A shift in the DGP occurs so

that for $t \geq T$ it becomes

$$x_t = \mu* + \rho* (x_{t-1} - \mu*) + \epsilon_t, \tag{2.2}$$

where we assume $|\rho*| < 1$. The multi-horizon iterated AR(1) forecast with pre-break parameters is

$$\hat{x}_{T+b+h|T+b} = \mu + \rho \left(\hat{x}_{T+b+h-1|T+b} - \mu \right) = \mu + \rho^h \left(x_{T+b} - \mu \right), \tag{2.3}$$

where for simplicity we ignore parameter estimation uncertainty and any other $O_p(1/T)$ terms such as forecast origin estimation in order to focus on the impact of the break. Subtracting the forecast (2.3) from the DGP (2.2) at period $T + b + h$ gives the forecast error

$$\begin{aligned} \hat{e}_{T+b+h|T+b} =& x_{T+b+h} - \hat{x}_{T+b+h|T+b} \\ =& \mu* + (\rho*)^h (x_{T+b} - \mu*) - \mu - \rho^h (x_{T+b} - \mu) + \sum_{j=0}^{h-1} \rho*^j \epsilon_{T+b+h-j} \\ =& \left(1 - \rho^h\right) \nabla\mu* + \left(\rho*^h - \rho^h\right) (x_{T+b} - \mu*) + \sum_{j=0}^{h-1} \rho*^j \epsilon_{T+b+h-j}, \end{aligned} \tag{2.4}$$

where $\nabla\mu* = \mu* - \mu$ is the change in the long-run mean. The expected value of $\hat{\epsilon}_{T+b+h|T+b}$ is

$$\begin{aligned} \mathbb{E}\left[\hat{\epsilon}_{T+b+h|T+b}\right] &= \left(1 - \rho^h\right) \nabla\mu* + \left(\rho*^h - \rho^h\right) \mathbb{E}\left[x_{T+b} - \mu*\right] \\ &= \left[\left(1 - \rho^h\right) - \rho*^b \left(\rho*^h - \rho^h\right)\right] \nabla\mu*. \end{aligned} \tag{2.5}$$

This simplifies to $\left(1 - \rho^h\right) \nabla\mu*$ when $\rho = \rho*$ and at longer horizons (2.5) converges to

$$\mathbb{E}\left[\hat{\epsilon}_{T+b+h|T+b}\right] \underset{h \to \infty}{\to} \nabla\mu*. \tag{2.6}$$

When $\rho = \rho*$, the mean square forecast error (MSFE) is

$$\begin{aligned} \mathbb{E}\left[\hat{\epsilon}^2_{T+b+h|T+b}\right] &= \left(\mathbb{E}\left[\hat{\epsilon}_{T+b+h|T+b}\right]\right)^2 + \left(\sum_{j=0}^{h-1} \rho^{2j}\right) \sigma^2_\epsilon \\ &= \left(1 - \rho^h\right)^2 (\nabla\mu*)^2 + \left(\sum_{j=0}^{h-1} \rho^{2j}\right) \sigma^2_\epsilon, \end{aligned} \tag{2.7}$$

which at longer horizons converges to

$$\mathbb{E}\left[\hat{\epsilon}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} (\nabla\mu*)^2 + \frac{\sigma^2_\epsilon}{1-\rho^2}. \tag{2.8}$$

Thus, the pre-break AR(1) model exhibits a persistent forecast bias across all horizons. Since this bias persists across all horizons, differencing the data and the forecasts can remove the bias at longer horizons. However, if the object of interest is the data in levels, then simple differenced forecasts are not beneficial. Thus, it is necessary to consider alternative approaches.

3 Naive Robust Forecasts

We now summarize how multi-horizon forecasts from naive predictors perform after a break.

3.1 A Random Walk

A random walk is one of the simplest and most commonly used naive forecasting procedures. It predicts that the process will be unchanged from the last observable period

$$\bar{x}_{T+b+h|T+b} = x_{T+b}, \tag{3.1}$$

so the forecast error after the break is

$$\bar{e}_{T+b+h|T+b} = \left(\rho *^h - 1\right)(x_{T+b} - \mu^*) + \sum_{j=0}^{h-1} \rho *^j \epsilon_{T+b+h-j}, \tag{3.2}$$

and the expected forecast error (i.e. bias) is

$$\mathbb{E}\left[\bar{e}_{T+b+h|T+b}\right] = \left(\rho *^h - 1\right)\mathbb{E}\left[(x_{T+b} - \mu^*)\right] \approx \left(1 - \rho*^h\right)\rho *^b \nabla\mu * . \tag{3.3}$$

If $|\rho*| < 1$, then the bias is in the same direction as $\nabla\mu*$. However, as the forecast origin moves away from the break date, the bias disappears as the forecast adapts to the new long-run mean

$$\mathbb{E}\left[\bar{e}_{T+b+h|T+b}\right] \underset{b\to\infty}{\to} 0. \tag{3.4}$$

When $\rho = \rho*$ the MSFE is

$$\mathbb{E}\left[\bar{e}^2_{T+b+h|T+b}\right] = \left(1-\rho^h\right)^2\left(\rho^{2b}\left(\nabla\mu*\right)^2 + \frac{\sigma^2_\epsilon}{1-\rho^2}\right) + \left(\sum_{j=0}^{h-1}\rho^{2j}\right)\sigma^2_\epsilon, \tag{3.5}$$

which at longer horizons converges to

$$\mathbb{E}\left[\bar{e}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} \rho^{2b}\left(\nabla\mu*\right)^2 + \frac{2\sigma^2_\epsilon}{1-\rho^2}. \tag{3.6}$$

The random walk forecast always outperforms the AR(1) model when the size of the break satisfies $\left(\frac{\nabla\mu*}{\sigma_\epsilon}\right)^2 > \frac{1}{(1-\rho^{2b})(1-\rho^2)}$. The larger the break, the more persistence there is, and the further away from the break date, then the better the random walk performs relative to the AR(1).

3.2 A Robust Predictor

An alternative approach is a robust predictor, which Castle et al. (2015a) and Hendry (2018) show works very well for 1-step-ahead forecasts during and after a structural break. The robust predictor has the advantage of rapidly adjusting to the new long-run mean like the random walk while also retaining the economic content of the persistence parameter despite the fact that it may have shifted. The formulation of the robust predictor is model specific but the idea behind it is to quasi-difference the forecast model so that the long-run mean is removed. In the context of the AR(1) model, the robustified 1-step-ahead forecasts become

$$\tilde{x}_{T+b+1|T+b} = x_{T+b} + \rho\Delta x_{T+b}, \tag{3.7}$$

which completely removes the long-run run mean but keeps the persistence parameter. When $\rho = 0$ (3.7) collapses to the random walk forecast and when $\rho = 1$ it is the double difference device; see Hendry (2006). The iterated forecasts from this predictor at multiple-horizons are

$$\begin{aligned}\tilde{x}_{T+b+h|T+b} =& \tilde{x}_{T+b+h-1|T+b} + \rho\Delta\tilde{x}_{T+b+h-1|T+b} \\ =& x_{T+b} + \sum_{j=1}^{h}\rho^j\Delta x_{T+b} = x_{T+b} + \left(1-\rho^h\right)\frac{\rho}{1-\rho}\Delta x_{T+b},\end{aligned} \tag{3.8}$$

where the last equality holds because $|\rho| \neq 1$. The forecast errors are

$$\begin{aligned}\tilde{\hat{e}}_{T+b+h|T+b} &= \left(\rho^h - 1\right) \frac{\rho}{1-\rho} \Delta x_{T+b} + \left(\rho *^h - 1\right) \left(x_{T+b} - \mu*\right) + \sum_{j=0}^{h-1} \rho *^j \epsilon_{T+b+h-j} \\ &= \left(\rho *^h - 1\right) \left(\frac{\epsilon_{T+b}}{1-\rho*}\right) + \left[\left(\frac{\rho^h - 1}{1-\rho}\right) \rho - \rho * \left(\frac{\rho *^h - 1}{1-\rho*}\right)\right] \Delta x_{T+b} + \sum_{j=0}^{h-1} \rho *^j \epsilon_{T+b+h-j},\end{aligned} \tag{3.9}$$

where we use the fact that $\Delta x_{T+b} = (\rho * - 1)\left(x_{T+b-1} - \mu*\right) + \epsilon_{T+b}$. The expected forecast error is

$$\mathbb{E}\left[\tilde{\hat{e}}_{T+b+h|T+b}\right] = \left[1 - \frac{1-\rho^h}{1-\rho*^h}\left(\frac{1-\rho*}{1-\rho}\right)\frac{\rho}{\rho*}\right]\left(1 - \rho*^h\right) \rho *^b \nabla \mu* \tag{3.10}$$

where the direction depends on $\rho/\rho*$. When this ratio is unity, then there is no forecast bias.

The MSFE of this robust predictor is approximately

$$\begin{aligned}\mathbb{E}\left[\tilde{\hat{\epsilon}}^2_{T+b+h|T+b}\right] \approx &\left[1 - \frac{1-\rho^h}{1-\rho*^h}\left(\frac{1-\rho*}{1-\rho}\right)\frac{\rho}{\rho*}\right]^2 \left(1 - \rho*^h\right)^2 \left(\rho *^{2b} \left(\nabla \mu*\right)^2 + \frac{2\rho *^2 \sigma^2_\epsilon}{(1-\rho*)(1-\rho*^2)}\right) \\ &+ \left\{2\left[\frac{1-\rho^h}{1-\rho*^h}\left(\frac{1-\rho*}{1-\rho}\right)\rho - \rho*\right]\left(\frac{1-\rho*^h}{1-\rho*}\right)^2 + \left(\sum_{j=0}^{h-1} \rho *^{2j} + \left(\frac{1-\rho*^h}{1-\rho*}\right)^2\right)\right\} \sigma^2_\epsilon,\end{aligned} \tag{3.11}$$

which at longer horizons converges to

$$\begin{aligned}\mathbb{E}\left[\tilde{\hat{\epsilon}}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} &\left[1 - \frac{(1-\rho*)}{(1-\rho)}\frac{\rho}{\rho*}\right]^2 \left(\rho *^{2b} \left(\nabla \mu*\right)^2 + \frac{2\rho *^2 \sigma^2_\epsilon}{(1-\rho*)(1-\rho*^2)}\right) \\ &+ \left(2\rho * \left[\frac{1-\rho*}{1-\rho}\frac{\rho}{\rho*} - 1\right] + 1\right) \frac{\sigma^2_\epsilon}{(1-\rho*)^2} + \frac{\sigma^2_\epsilon}{1-\rho*^2}.\end{aligned} \tag{3.12}$$

This illustrates that the forecast-error variance explodes and the bias remains when $\rho \neq \rho*$. The variance remains inflated even as the forecast origin moves away from the break date

$$\begin{aligned}\mathbb{E}\left[\tilde{\hat{\epsilon}}^2_{T+b+h|T+b}\right] \underset{h,b\to\infty}{\to} &\left[1 - \frac{(1-\rho*)}{(1-\rho)}\frac{\rho}{\rho*}\right]^2 \left(\frac{2\rho *^2 \sigma^2_\epsilon}{(1-\rho*)(1-\rho*^2)}\right) \\ &+ 2\rho * \left[\frac{1-\rho*}{1-\rho}\frac{\rho}{\rho*} - 1\right] \frac{\sigma^2_\epsilon}{(1-\rho*)^2} + \frac{\sigma^2_\epsilon}{1-\rho*^2} + \frac{\sigma^2_\epsilon}{(1-\rho*)^2},\end{aligned} \tag{3.13}$$

which is much larger than the random walk.

However, when $\rho = \rho*$, then at longer horizons the MSFE converges to

$$\mathbb{E}\left[\tilde{\hat{\epsilon}}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} \left(\frac{2}{1-\rho}\right)\frac{\sigma^2_\epsilon}{1-\rho^2}, \tag{3.14}$$

which shows that although there is no bias, the variance is more than double that of the AR(1) forecast. It only outperforms the AR(1) when $\left(\frac{\nabla\mu*}{\sigma^2_\epsilon}\right)^2 > \frac{1}{(1-\rho)^2}$ or the random walk when $\left(\frac{\nabla\mu*}{\sigma^2_\epsilon}\right)^2 > \frac{2\rho}{\rho^{2b}(1-\rho)(1-\rho^2)}$. Thus, the break has to be large and this requirement increases the further away the break is. This illustrates the trade-off between removing/reducing forecast bias and increasing the forecast-error variance. While completely removing the bias is beneficial at short forecast horizons, it is less beneficial when it induces higher forecast-error variance at longer horizons.

4 Reinterpreting Robust Forecasts

Naive forecasts can also be reinterpreted as local estimators of the long-run mean. We start with the expected value of $\mathbb{E}[x_{T+b}]$ b periods after the break

$$\bar{\mu} = \mathbb{E}[x_{T+b}] = \mu* - \rho*^b \nabla\mu*, \tag{4.1}$$

which collapses to the new long-run mean as the forecast origin moves away from the break. A special case estimator of this long-run mean is to choose $\hat{\bar{\mu}} = x_{T+b}$. Replacing the long-run mean in the AR(1) forecast with this local estimate gives

$$\begin{aligned}\hat{\bar{x}}_{T+b+h|T+b} =& \hat{\bar{\mu}} + \rho\left(\hat{\bar{x}}_{T+b+h-1|T+b} - \hat{\bar{\mu}}\right) \\ =& x_{T+b} + \rho^h\left(x_{T+b} - x_{T+b}\right) = x_{T+b} = \bar{x}_{T+b+h|T+b},\end{aligned} \tag{4.2}$$

which is identical to the random walk forecast in (3.1). This illustrates why the random walk does so well following a structural break as it provides an immediate update of the long-run mean. However, the cost of doing so is an inflated error variance with just a gradual reduction in the bias.

An alternative estimate of the long-run mean is based on a transformation that removes any

bias immediately after the break. A transformation that satisfies this requirement is

$$\begin{aligned}\tilde{\mu} = \mathbb{E}\left[x_{T+b} + \frac{\rho}{1-\rho}\Delta x_{T+b}\right] =& \mu* - \rho*^{b}\nabla\mu* + \left(\frac{\rho}{1-\rho}\right)\rho*^{b-1}(1-\rho*)\nabla\mu* \\ =& \mu* - \rho*^{b}\nabla\mu*\left(1 - \frac{\rho}{\rho*}\left(\frac{1-\rho*}{1-\rho}\right)\right),\end{aligned} \tag{4.3}$$

which is identical to $\mu*$ when $\rho = \rho*$ or as $b \to \infty$. We can estimate this transformation as $\hat{\tilde{\mu}} = x_{T+b} + \frac{\rho}{1-\rho}\Delta x_{T+b} = \frac{1}{1-\rho}(x_{T+b} - \rho x_{T+b-1})$ and plug it into the AR(1) forecast to get

$$\begin{aligned}\hat{\tilde{x}}_{T+b+h|T+b} = \hat{\tilde{\mu}} + \rho\left(\hat{\tilde{x}}_{T+b+h-1|T+b} - \hat{\tilde{\mu}}\right) = \hat{\tilde{\mu}} + \rho^{h}\left(x_{T+b} - \hat{\tilde{\mu}}\right) \\ = x_{T+b} + \left(1-\rho^{h}\right)\frac{\rho}{1-\rho}\Delta x_{T+b} = \tilde{x}_{T+b+h|T+b},\end{aligned} \tag{4.4}$$

which is identical to the robust predictor in (3.8).

This illustrates that robust forecasts can be interpreted as alternative local estimators of the long-run mean. Furthermore, this interpretation is not limited to just the random walk or the robust predictor. For example, forecasts from the double difference device proposed by Hendry (2006) are obtained when $\hat{\tilde{\mu}}_h = x_{T+b} + \frac{h}{1-\rho^h}\Delta x_{T+b}$. This suggests, that there is a broad class of predictors that can be embedded into policy or economic models to help improve their forecasts.

Reinterpreting naive predictors as local estimates of the long-run mean also illustrates a way to improve the forecasts further. In particular, robust predictors generally reduce forecast bias at the expense of larger forecast error variances at longer horizons. This stems from the fact that the local estimates of the mean are very noisy because they are based on a single observation. This additional forecast error variance can be reduced by averaging over multiple observations.

5 Smooth Robust Forecasts

We can reduce the noise in local estimates of the long-run mean by taking a local average of observations since the break. This builds on smoothed methods (i.e. long differencing) to estimate equilibrium correction models proposed by Hendry (2006) and evaluated in Castle et al. (2015a), and relates to the choice of estimation windows for forecasting after breaks in Pesaran and Timmermann (2007) or the adaptive procedures by Giraitis et al. (2013).

5.1 A Smooth Random Walk

Consider a smooth random walk where the long-run mean with an estimation window of size n is estimated as

$$\bar{\bar{\mu}} = \frac{1}{n}\sum_{j=0}^{n-1} x_{T+b-j} = \mu* + \frac{\rho*}{n}\sum_{j=0}^{n-1}\left(x_{T+b-1-j} - \mu*\right) + \sum_{j=0}^{n-1}\epsilon_{T+b-j}, \tag{5.1}$$

which is similar to estimates of backward-looking inflation expectations; e.g. see Atkeson and Ohanian (2001) and Ball and Mazumder (2011). Plugging (5.1) into the AR(1) forecast produces a smooth quasi-random walk forecast[1]

$$\bar{\bar{x}}_{T+b+h|T+b} = \bar{\bar{\mu}} + \rho\left(\bar{\bar{x}}_{T+b+h-1|T+b} - \bar{\bar{\mu}}\right) = \bar{\bar{\mu}} + \rho^h\left(x_{T+b} - \bar{\bar{\mu}}\right). \tag{5.2}$$

The forecast errors are

$$\bar{\bar{e}}_{T+b+h|T+b} = \left(\rho^h - 1\right)\left[\frac{1}{n}\sum_{j=0}^{n-1}\left(x_{T+b-j} - \mu*\right)\right] + \left(\rho*^h - \rho^h\right)\left(x_{T+b} - \mu*\right) + \sum_{j=0}^{h-1}\rho*^j\,\epsilon_{T+b+h-j}. \tag{5.3}$$

which are driven by a local average of the last n observations as $h \to \infty$. The expected error is

$$\mathbb{E}\left[\bar{\bar{e}}_{T+b+h|T+b}\right] = \left\{\frac{\left(1-\rho^h\right)}{n}\sum_{j=0}^{n-1}\rho*^{-j} + \rho^h - \rho*^h\right\}\rho*^b\,\nabla\mu*, \tag{5.4}$$

which illustrates that the forecast bias remains even if the slope does not shift.

At long horizons the approximate MSFE converges to

$$\mathbb{E}\left[\bar{\bar{e}}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} \left\{\frac{1}{n}\sum_{j=0}^{n-1}\rho*^{b-j}\right\}^2 (\nabla\mu*)^2 + \frac{\sigma^2_\epsilon}{1-\rho*^2} + \mathbb{V}\left[\frac{1}{n}\sum_{j=0}^{n-1} x_{T+b-j}\right]. \tag{5.5}$$

This is insensitive to changes in the slope but with the cost of an inflated bias and variance. However, (5.5) shrinks to the smallest possible MSFE as the estimation window n grows[2]

$$\mathbb{E}\left[\bar{\bar{e}}^2_{T+b+h|T+b}\right] \underset{h,n\to\infty}{\to} \frac{\sigma^2_\epsilon}{1-\rho*^2}. \tag{5.6}$$

[1] It is not identical to a smooth random walk but does converge to one at longer horizons when $|\rho| < 1$.

[2] This requires that $\mathbb{V}\left[\frac{1}{n}\sum_{j=0}^{n-1} x_{T+b-j}\right] \underset{n\to\infty}{\to} 0$ which holds when $|\rho| < 1$, $|\rho*| < 1$ and $n < b$.

The smoothed local estimate of the long-run mean allows the misspecified forecast to converge back to the correct mean. This helps explain why Atkeson and Ohanian (2001) and Ball and Mazumder (2011) find that backward-looking inflation expectations were useful for forecasting U.S. inflation over the volatile period of the 1970's and 1980's.

5.2 A Smooth Robust Predictor

Now consider a smooth estimate of the robust predictor in (4.3) with a window of size n

$$\begin{aligned}\tilde{\tilde{\mu}} =&\frac{1}{n}\sum_{j=0}^{n-1}\left(x_{T+b-j}+\frac{\rho}{1-\rho}\Delta x_{T+b-j}\right)=\frac{1}{n\left(1-\rho\right)}\sum_{j=0}^{n-1}\left(x_{T+b-j}-\rho x_{T+b-j-1}\right)\\ =&\mu*+\frac{1}{n\left(1-\rho*\right)}\sum_{j=0}^{n-1}\epsilon_{T+b-j}+\left(\frac{\rho}{1-\rho}-\frac{\rho*}{1-\rho*}\right)\frac{1}{n}\sum_{j=0}^{n-1}\Delta x_{T+b-j}.\end{aligned} \tag{5.7}$$

Plugging this into the AR(1) produces a quasi-smooth robust predictor

$$\tilde{\tilde{x}}_{T+b+h|T+b}=\tilde{\tilde{\mu}}+\rho\left(\tilde{\tilde{x}}_{T+b+h-1|T+b}-\tilde{\tilde{\mu}}\right)=\tilde{\tilde{\mu}}+\rho^{h}\left(x_{T+b}-\tilde{\tilde{\mu}}\right). \tag{5.8}$$

At long horizons the approximate MSFE of this forecast converges to

$$\begin{aligned}\mathbb{E}\left[\tilde{\tilde{e}}_{T+b+h|T+b}^{2}\right]\underset{h\to\infty}{\to}&\left[1-\left(\frac{1-\rho*}{1-\rho}\right)\frac{\rho}{\rho*}\right]^{2}\left\{\frac{1}{n}\sum_{j=0}^{n-1}\rho*^{b-j}\nabla\mu*\right\}^{2}+\frac{\sigma_{\epsilon}^{2}}{1-\rho*^{2}}\\ &+\left(\frac{1}{1-\rho*}\right)^{2}\left\{\left[\left(\frac{1-\rho*}{1-\rho}\right)\rho-\rho*\right]^{2}\mathbb{V}\left[\frac{1}{n}\sum_{j=0}^{n-1}\Delta x_{T+b-j}\right]+\frac{\sigma_{\epsilon}^{2}}{n}\right\}\\ &+\frac{2}{\left(1-\rho*\right)^{2}}\left(\left(\frac{1-\rho*}{1-\rho}\right)\rho-\rho*\right)\frac{\sigma_{\epsilon}^{2}}{n^{2}}\left(1+\left(\rho*-1\right)\sum_{j=0}^{n-1}\sum_{i=0}^{n-1-j}\rho^{*i}\right),\end{aligned} \tag{5.9}$$

which indicates that it suffers from both bias and an inflated variance at all horizons. However, (5.9) shrinks as the estimation window n grows, and converges to the smallest possible MSFE[3]

$$\mathbb{E}\left[\tilde{\tilde{e}}_{T+b+h|T+b}^{2}\right]\underset{h,n\to\infty}{\to}\frac{\sigma_{\epsilon}^{2}}{1-\rho*^{2}}. \tag{5.10}$$

[3]The requires that $\mathbb{V}\left[\frac{1}{n}\sum_{j=0}^{n-1}\Delta x_{T+b-j}\right]\underset{n\to\infty}{\to}0$ which holds if Δx_{T+b} is stationary after the break and $n\leq b$.

When the slope parameter shifts, the smooth robust predictor becomes more accurate the further away from the break. The MSFE converges to the correctly specified model forecast even though it uses the incorrect slope parameter values to transform the long-run mean.

In the special case where only the intercept shifts so that $\rho = \rho*$, then the MSFE is

$$\mathbb{E}\left[\tilde{\epsilon}^2_{T+b+h|T+b}\right] = \left(\frac{\rho^h - 1}{1-\rho}\right)^2 \frac{\sigma^2_\epsilon}{n} + \sigma^2_\epsilon \sum_{j=0}^{h-1} \rho^{2j}, \tag{5.11}$$

which has no bias and at longer horizons converges to

$$\mathbb{E}\left[\tilde{\epsilon}^2_{T+b+h|T+b}\right] \underset{h\to\infty}{\to} \left(1 + \frac{1}{n}\frac{1-\rho^2}{(1-\rho)^2}\right)\frac{\sigma^2_\epsilon}{1-\rho^2}. \tag{5.12}$$

Therefore, when the slope parameter does not shift, the bias is eliminated no matter how far away the break is. However, the forecast error variance remains inflated. When $n = 1$, this is the robust predictor MSFE in (3.14). As n increases the variance shrinks to the correctly specified forecast so that the smooth robust predictor outperforms the random walk and the AR(1) model exactly as the smooth random walk did in (5.6). Comparing against the smooth random walk for large h indicates that the smooth robust predictor performs best when

$$\left(\frac{\nabla\mu*}{\sigma_\epsilon}\right)^2 > \left(\frac{1}{(1-\rho)} - \frac{1}{n}\sum_{i=0}^{n-2}\sum_{j=0}^{n-2-i}\rho^j\right)\frac{2n\rho}{\left\{\sum_{j=0}^{n-1}\rho^{b-j}\right\}^2(1-\rho^2)}. \tag{5.13}$$

The smooth robust predictor dominates the smooth random walk for any value of n when $\rho < 0$. When $\rho > 0$ then the smallest break required for the smooth robust predictor to dominate the smooth random walk is greater than 2 standard deviations. However, simulations indicate that this requirement can be much larger depending on ρ. As a result, it may be valuable to combine both measures by averaging. There is a large literature which demonstrates the benefits of doing so; e.g. see Bates and Granger (1969), Hendry and Clements (2004), and Timmermann (2006).

The results for both the smooth predictors rely on the occurrence of a single break at time period $t = T$. If there is only one break then it is clearly optimal to choose an estimation window $n = b$ to maximize the available post-break information. However if there are additional breaks at $t > T$ then it may be better to set $n < b$ to account for the forecast-error bias-variance trade-off.

In practice, choosing $n = 4$ as suggested by Hendry (2006) for quarterly data, works reasonably well when there are multiple shifts. However, it is useful to extend this window when working with higher frequency data or using estimators that perform additional differencing.

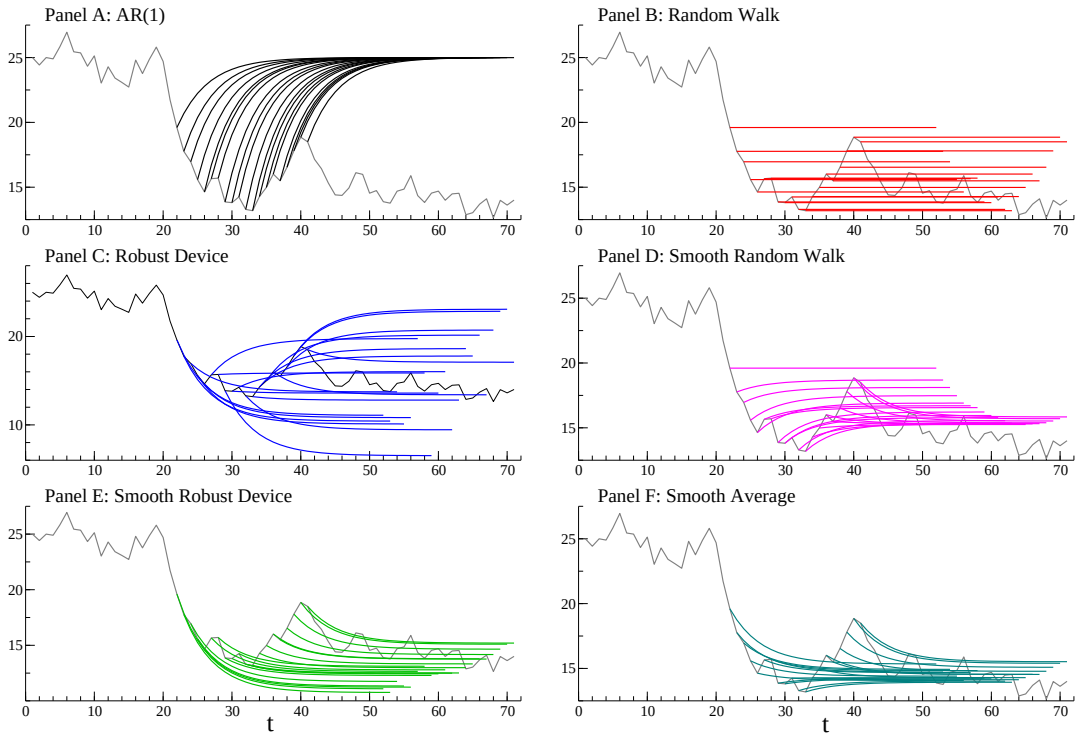
It may also be useful to extend the estimation window if there are suspected shifts in the variance. An increase in the residual variance reduces the signal around a shift in the mean. Therefore, a larger $\nabla\mu*$ would be required to satisfy (5.13). Thus, we would either favour the smooth random walk forecast or seek to extend the post-break estimation window.

It is possible to consider more sophisticated approaches for choosing an estimation window which take into account the length of the forecast horizon. For example, Pesaran and Timmermann (2007) show that pooling models and cross-validation work well when there are multiple unknown break dates. Alternatively, Giraitis et al. (2013) propose an adaptive procedure that gives weight to observations before and after a break based on model performance. Indicator saturation methods could also be used to simultaneously determine the break date and choose the estimation sample to utilize as much of the post-break information as possible; see Castle et al. (2015b) and Castle et al. (2018). These extensions are left for future research.

6 Simulation Evidence

In this section we explore the performance of the various forecasting methods considered above using simulations. We start by conducting a visual analysis of the forecast performance by plotting the forecasts after a change in the parameters. For the purpose of this analysis we choose the data generating process so that it coincides with (2.1) above where $\rho = 0.8$ and the long-run mean is $\mu = \frac{5}{1-\rho} = 25$. At $t = 20$ the model parameters shift so that the DGP becomes (2.2) where ρ changes to $\rho* = 0.66\overline{6}$ so that the long-run mean also shifts to $\mu* = 15$. This is an extremely large shift that any forecaster should be able to detect and adapt to. But it provides a clear illustration of the differences between the forecasting methods out through 30 forecast horizons.

Figure 6.1 presents the results of this exercise. Consistent with the mathematical derivations above, the AR(1) forecasts in panel A are biased after the break in parameters as they continually try to correct back to the pre-break long-run mean of the series. This illustrates why it is beneficial to have updated estimates of the long-run mean after a break.

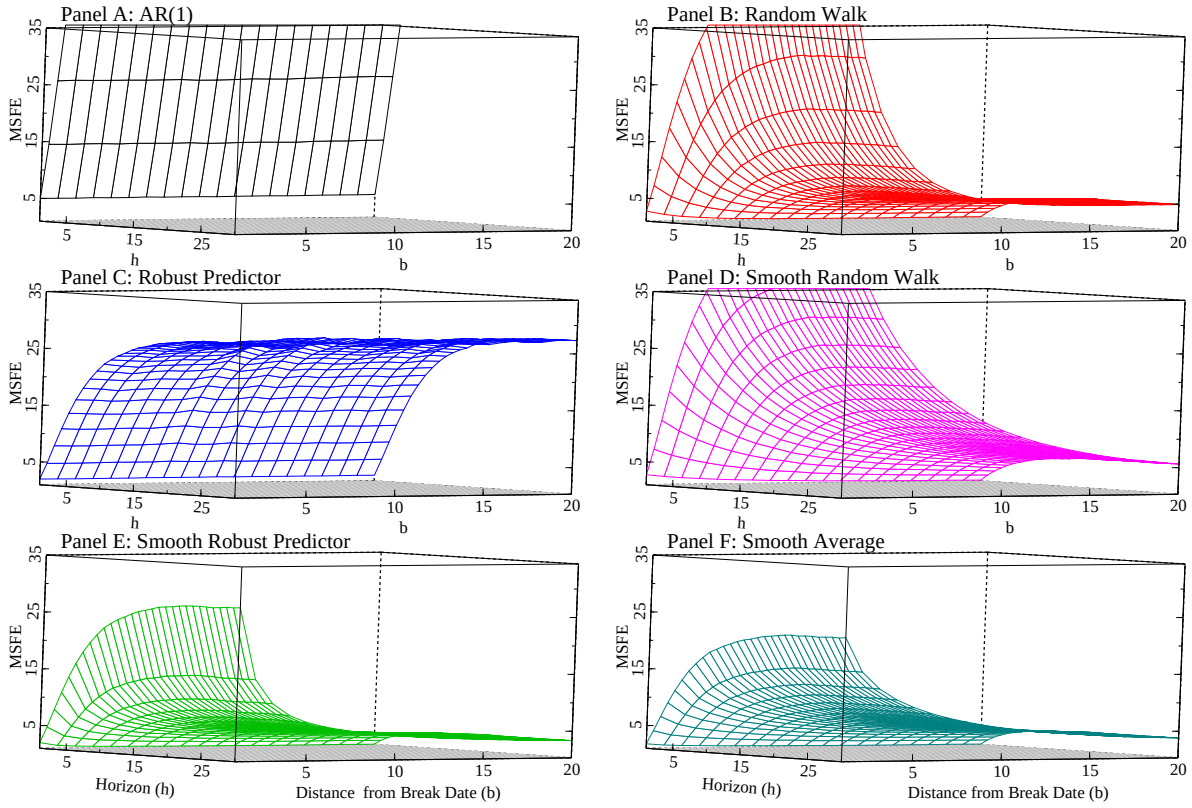


Notes: The break occurs at $t = 20$ where $\rho = 0.8$ changes to $\rho* = 0.666$ with an long-run mean of $\frac{5}{1-\rho}$ which becomes $\frac{5}{1-\rho*}$. The forecasts start at $t = 21$. The smooth forecasts use an expanding window of $n = b$ to utilize the entire post break sample.

Figure 6.1: Forecast Illustration: Break in Long-Run Mean with Slope Shift

The robust forecasts provide a way to reduce the forecast bias at the cost of greater volatility. The random walk in panel B reduces the bias by extrapolating from each forecast origin and thus gradually corrects to the new long-run mean but is volatile and exhibits an initial upward bias. The robust predictor in panel C remains centered around the new long-run mean on average but suffers from large oscillations in the forecasts as they extrapolate changes in the series. Therefore, any benefits from reductions in the bias are drowned out by the large variance.

The smooth robust forecasts exhibit less volatility than the robust forecasts with the cost of greater bias which can be offset by combining them. The smooth random walk in panel D slowly converges to the new long-run mean from above while the much less volatile smooth robust predictor in panel E converges to the new long-run mean from below. Given that these two forecasts are differentially misspecified so that they are biased in opposite directions, then there are large benefits to combining them; see Hendry and Clements (2004). The smooth average of the random walk and robust predictor in panel F reduces both the bias and the volatility in the forecasts and generates a forecast that closely follows the series many steps into the future.

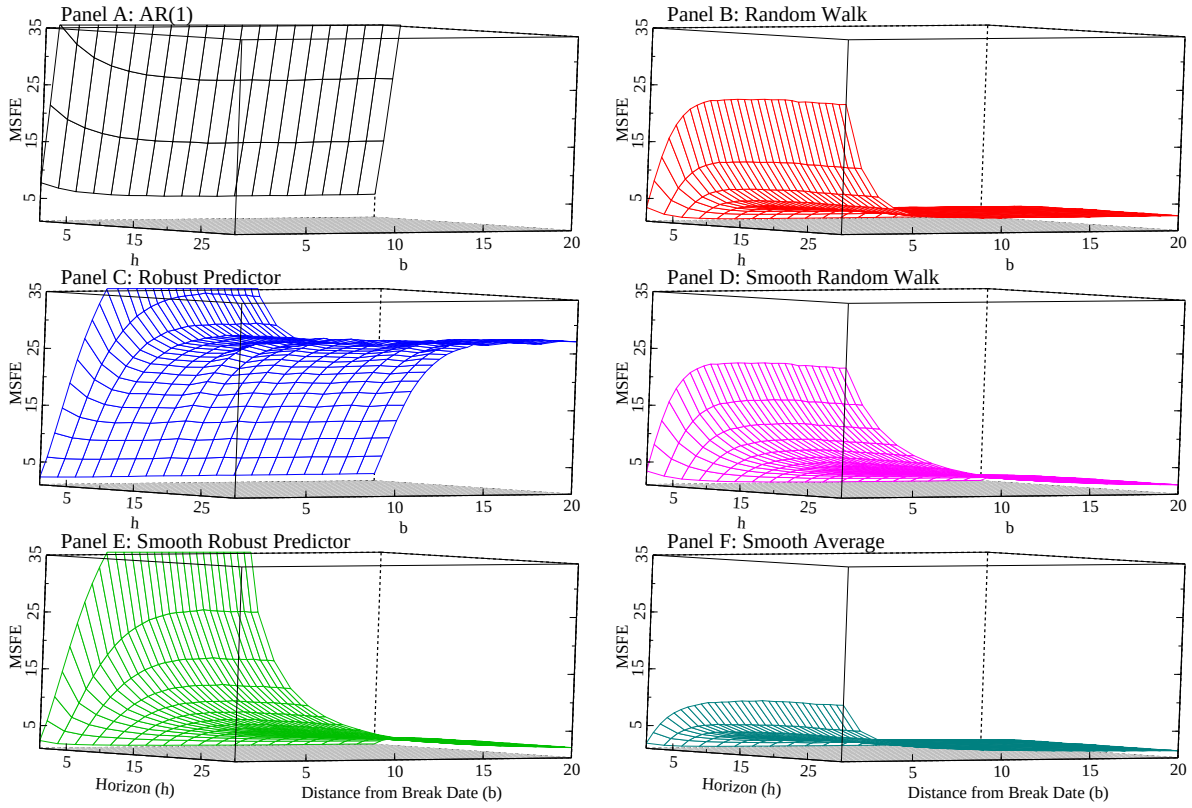


Notes: 20,000 replications.

Figure 6.2: Simulations: Break in Long-Run Mean without Slope Shift

Next we examine how forecast performance evolves across horizons and as the forecasts move away from the break, both when the slope parameter shifts and when it does not. We generate the forecasts across 20,000 draws of the data using the same parameterization as described above and compute the MSFE across draws at each horizon and for each forecast origin. The long-run mean shifts from 25 to 15 at time period $t = 20$ and the forecasts are generated starting from $t = 21$ or $b = 1$ through $b = 20$ across 30 forecast horizons. Figures 6.2 and 6.3 present the results for the case where $n = b$. The results for $n = 4$ are similar; see Appendix A.1.

Figure 6.2 presents the case of a break in the long-run mean without a slope shift. For comparison, the MSFEs of the AR(1) in panel A exceed 100 at the longest horizon. The random walk in Panel B has lower MSFEs at all horizons particularly as it moves away from the break. The robust predictor in panel C outperforms the random walk initially after the break, but does not benefit from moving away from the break as the forecast performance suffers from elevated forecast error variance. The smooth random walk in panel D performs similar to the random walk with a slower reduction in MSFE as b increases. The smooth robust predictor in panel E sees substantial declines



Notes: 20,000 replications.

Figure 6.3: Simulations: Break in Long-Run Mean due to Slope Shift

in the MSFEs relative to the non-smoothed version as it moves away from the break date. An expanding estimation window helps to reduce the forecast-error variance so that by 20 periods after the break the forecast performance converges to the smallest possible MSFE. The smooth average in panel F has the smallest MSFEs initially but sees more gradual declines relative to the smooth robust predictor moving away from the break date although the MSFEs are already small.

These results illustrate that when the slope parameters do not shift, then the differencing done by the robust predictor is correct and does not suffer from any bias. Thus, in this case it is more beneficial to use the smooth robust predictor as compared to the smooth random walk or even a combination of the two, particularly as the forecast moves away from the break date.

We compare the results in Figure 6.2 with Figure 6.3 for the case where the slope parameter shifts and so the differencing done by the robust predictor induces additional forecast misspecification. The AR(1) in panel A continues to have the highest MSFEs. The random walk in panel B sees a considerable improvement relative to Figure 6.2. This is in part because the slope parameter becomes less persistent and so the process corrects to the long-run mean more quickly. The robust

predictor in panel C has a higher initial MSFE as its transformation does not completely remove bias until the forecasts move away from the break. The smooth random walk in panel D again experiences a slower forecast improvement relative to the random walk. However, as it moves away from the break then its MSFEs fall below the random walk due to the lower forecast error variance. The smooth robust predictor in panel E suffers from the same initial misspecification as the robust predictor, but improves moving away from the break. The smooth average in panel F has the lowest MSFEs due to the differences in mis-specification between the random walk and robust predictor.

Taken together, the results in Figures 6.2 and 6.3 illustrate that the smooth robust forecasts can improve upon their non-smoothed counterparts. It is also possible to generate further improvements by combining the smooth forecasts when they are differentially misspecified. This illustrates that it is possible to see dramatic improvements in forecast performance immediately after a break occurs and across all horizons.

7 Empirical Examples

7.1 U.K. Productivity

In this section we examine whether it is possible to improve upon forecasts of productivity in the United Kingdom. The slowdown in U.K. productivity (as measured by output per hour worked) since the Global Financial Crisis has been a puzzling issue that has received much attention. The Office of Budget Responsibility (OBR) has repeatedly issued long-term forecasts of productivity that proved to be optimistic. We start by estimating a statistical model that performs well prior to 2008 but fails thereafter. We show that the OBR's forecast performance is well represented by a transformation of the model. Finally, we show that it is possible to dramatically improve these forecasts when using smooth robust methods.

We measure productivity in the U.K. as gross value added, excluding oil and gas, in millions of pounds divided by the total number of hours worked by all workers and normalized to be 100 in 2009 Q1. We construct a dataset of historical vintages of productivity based on the second and third releases of GDP; see Appendix A.2. We also obtain the latest available productivity data (as of June 30th, 2020) from the Office for National Statistics from 1997-Q1 through 2020-Q1 which we use to evaluate the forecasts. Our definitions and data sources are aligned with the OBR.

We start by estimating an AR(1) model with a constant and a linear trend from 1997-Q2 through to 2007-Q4 using the latest available data:

$$\hat{x}_t = \underset{(10.6)}{36.6} + \underset{(0.08)}{0.25}t + \underset{(0.13)}{0.54}x_{t-1}$$
$$\hat{\sigma}_x = 0.53 \;\; R^2 = 0.99,$$

where the coefficient standard errors are shown in parentheses. The long-run trend is 0.54% per quarter, or just over 2% annually, which is exactly the long-run trend that the OBR has used to guide its long-run forecasts. The inclusion of the trend in the model differs from the theoretical discussion and simulations above. This helps illustrate the limitations of the robust predictor but highlights the strengths of alternative estimators of the long-run.

To generate forecasts that align with the OBR forecasts, we re-estimate the model through 2007-Q4 using real-time data to produce a forecast in the first and third quarters of each year. Despite the very good in-sample fit up through 2007, the AR(1) with trend produces poor long-run forecasts after 2008. Although the pattern matches the OBR forecasts in Figure 1.1, these are much worse, probably from the trend t being explicit.

We can improve these forecasts by quasi-differencing them to obtain the robust predictor

$$\Delta\tilde{x}_{t+h|t} = 0.25 + 0.54\Delta x_{t+h-1|t} = \frac{0.25}{1-0.54} + 0.54\left(\Delta x_{t+h-1|t} - \frac{0.25}{1-0.54}\right),$$

where the differenced long-run trend becomes the long-run mean of the differenced forecast. Thus, if the trend changes, then this "robust" predictor will suffer from forecast failure.

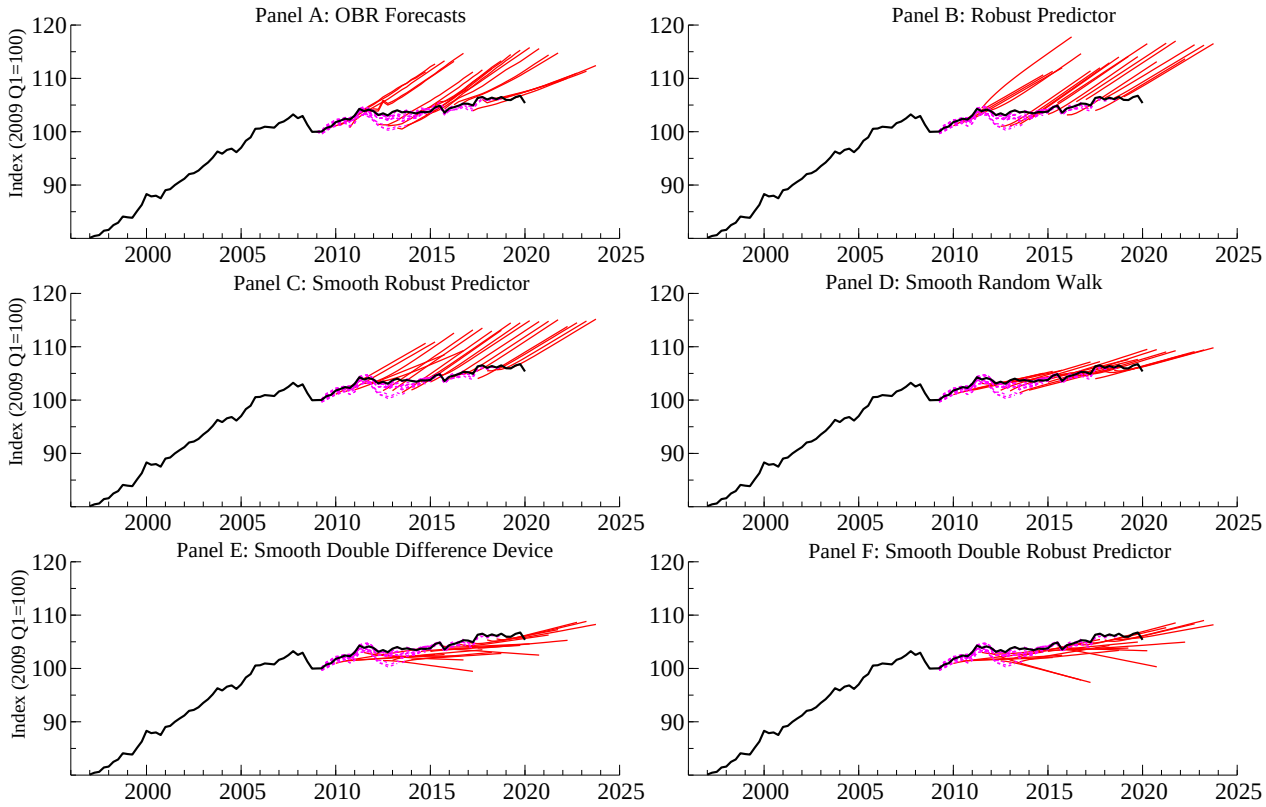
Panel B of Figure 7.2 illustrates that this is exactly what occurs as the robust predictor converges back to the historical long-run trend of 2% per annum. The OBR's forecasts as presented in Panel A of Figure 7.2 exhibit similar dynamics as they converge back to the historical long-run trend. The root mean square forecast errors (RMSFE) in Table 7.1 affirm the similarities at all horizons with the robust predictor performing slightly worse than OBR.

While the random walk, the robust predictor, and their smoothed variants and combinations improve on this performance, they still tend to be biased upwards as they fail to fully account for changes in the long-run trend; see Figure 7.1. Therefore, we consider additional estimators of the

long-run that explicitly allow for breaks in the long-run mean and trend: the double difference device, see Hendry (2006), and a double robust predictor. The resulting smooth local estimator for the long-run based on the combination of these two predictors is

$$\hat{\mu}_{SH,t,h,n} = \frac{1}{n}\sum_{i=1}^{n} x_{t-i} + \left(\frac{h}{1-0.54^h}\right)\frac{1}{n}\sum_{i=1}^{n}\Delta x_{t-i} + \gamma\left(\frac{0.54}{1-0.54}\right)\left(\frac{\sum_{j=1}^{h}\left(1-0.54^j\right)}{1-0.54^h}\right)\frac{1}{n}\sum_{i=1}^{n}\Delta^2 x_{t-i}$$

where if $\gamma = 0$ then $\hat{\mu}$ simplifies to a smooth double difference device, while if $\gamma = 1$ then it becomes a smooth double robust predictor. This estimator induces additional variance by differencing the data twice. Therefore, based on the additional costs of this differencing we extend the local estimation window to $n = 16$; see Figure 7.1. Thus, the estimation window earlier in the sample includes information prior to the break in 2008 such that $n > b$.



Notes: The dashed pink lines represent the realtime data vintages starting in 2009 Q1. The solid black line is the latest actual value and the red lines are the forecasts. All forecasts are generated starting in the first quarter of 2010 and updated after every two quarters through to March 2019 to mimic the OBR forecasts. All forecasts only use parameters from the AR(1) model with a trend and a constant which is re-estimated up through 2007 Q4 for each data vintage. The smooth forecasts are estimated over the last 16 observations before the forecast starts.

Figure 7.1: Alternative Forecasts of U.K. Productivity (Output per hour worked)

Table 7.1 illustrates the performance of the forecasts and combinations thereof evaluated against the latest vintage of actual data through 2020 Q1. While the random walk outperforms the robust predictor, the smooth robust predictor and smooth random walk are able improve on that. The smooth double robust and smooth double difference also perform well at longer horizons. The best combination at longer horizons is the smooth random walk which exhibits a somewhat positive bias and the smooth double difference device which exhibits a negative bias. It is difficult to beat the smooth random walk whose RMSFE is on average more than 75 percent smaller than the OBR's forecasts extending 20-quarters-ahead and is 84 percent smaller at the longest forecast horizon. This illustrates that the smooth random walk, unlike the smooth robust predictor, is robust to both shifts in the long-run mean, as shown by the theory and simulations, and in the long-run trend.

Table 7.1: Accuracy of U.K. Productivity Forecasts (2010 Q1 - 2020 Q1)

	RMSFE	**Relative RMSFE**								
h	**OBR**	**RP**	**RW**	**SRP**	**SRW**	**SA**$_{RW,RP}$	**SDD**	**SDR**	**SA**$_{DD,DR}$	**SA**$_{RW,DD}$
1	1.34	0.95	0.90	**0.84**	0.86	**0.84**	0.97	0.98	0.98	0.92
2	1.42	1.00	0.93	**0.75**	0.85	0.79	1.03	1.02	1.02	0.93
3	1.28	1.06	0.88	**0.55**	0.80	0.65	1.14	1.16	1.15	0.96
4	1.24	1.36	0.96	**0.60**	0.84	0.66	1.28	1.31	1.29	1.04
5	1.25	1.46	0.87	0.58	0.67	**0.47**	1.24	1.29	1.26	0.92
6	1.43	1.47	0.79	0.76	0.67	**0.56**	1.20	1.28	1.23	0.89
7	1.62	1.55	0.71	0.86	**0.45**	0.49	1.06	1.16	1.10	0.71
8	2.02	1.41	0.51	0.88	**0.33**	0.50	0.86	0.95	0.90	0.53
9	2.42	1.37	0.50	0.93	**0.25**	0.52	0.76	0.86	0.80	0.44
10	2.93	1.26	0.33	0.92	**0.15**	0.52	0.64	0.75	0.69	0.32
11	3.37	1.23	0.35	0.93	**0.17**	0.54	0.57	0.67	0.61	0.27
12	3.97	1.19	0.32	0.91	**0.18**	0.54	0.54	0.65	0.58	0.26
13	4.54	1.15	0.33	0.90	**0.19**	0.55	0.48	0.59	0.53	0.23
14	4.91	1.11	0.26	0.91	**0.19**	0.55	0.45	0.57	0.50	**0.19**
15	5.51	1.09	0.29	0.91	0.22	0.56	0.44	0.56	0.49	**0.21**
16	6.00	1.05	0.25	0.89	0.19	0.55	0.43	0.53	0.47	**0.18**
17	6.58	1.03	0.24	0.89	0.21	0.56	0.40	0.50	0.44	**0.17**
18	6.95	1.07	0.22	0.90	**0.18**	0.55	0.39	0.49	0.43	**0.18**
19	7.50	1.05	0.23	0.90	0.19	0.55	0.37	0.45	0.40	**0.16**
20	7.78	1.04	0.17	0.89	**0.16**	0.54	0.38	0.46	0.41	**0.16**
Ave.	3.63	1.13	0.30	0.87	**0.24**	0.55	0.58	0.66	0.61	0.33

Notes: RP: Robust Predictor with Drift, RW: Random Walk with Drift, SRP: Smooth Robust Predictor with Short-run Drift, SRW: Smooth Random Walk with Drift, SA: Smooth Average, SDD: Smooth Double Difference Device, SDR: Smooth Double Robust Predictor. Averages are equaly weighted. Calculated based on at most 20 observations out through 2020 Q1 which decrease every other horizon so that at 20 horizons there are 11 observations. Ave. is the Average RMSFE across all horizons caluculated using 11 observations for each horizon. Bolded values indicate the lowest RMSFE at each horizon.

7.2 U.S. 10-year Treasury yields

It is common to use survey forecasts to augment model forecasts. Survey forecasts are useful because they incorporate external information and pool different sources of information across multiple models; e.g. see Wright (2013). At the same time, there is evidence that survey forecasts struggle to beat naive forecasts of financial variables; see Kladivko et al. (2019). In this section we evaluate how quarterly forecasts of the U.S. 10-year Treasury yield from the Survey of Professional Forecasters (SPF) perform relative to naive and smooth robust methods.

We start by estimating an AR(1) model of the quarterly average of 10-year Treasury yields during a relatively stable period in the Great Moderation from 1999-Q1 through 2007-Q3. We account for any in-sample shifts using Step Indicator Saturation (see Castle et al. 2015b) with a target gauge of 1 percent. This implies that under the null hypothesis of no shifts, we expect to retain less than 0.4 irrelevant indicators on average given the sample size. Our estimated model is:

$$\begin{aligned}
\hat{x}_t =& \underset{(0.38)}{1.86} + \underset{(0.08)}{0.59}\, x_{t-1} + \text{Steps} \\
\text{Steps} =& \underset{(0.12)}{0.50} \times \mathrm{I}_{\leq 2000Q2} + \underset{(0.16)}{0.65} \times \mathrm{I}_{\leq 2002Q2} - \underset{(0.12)}{0.47} \times \mathrm{I}_{\leq 2003Q2} \\
& \hat{\sigma}_\epsilon = 0.22 \;\; R^2 = 0.91,
\end{aligned}$$

where the standard errors are shown in parentheses.

We find three step shifts between 2000 and 2003, which is consistent with other findings that there was a structural break in bond yields around the 2001 recession; see Bauer and Rudebusch (2013) and Campbell et al. (2020). The shifts are included to obtain a more accurate estimate of the parameters of interest but do not play a direct role in the forecasts themselves. Thus, the model is in line with the theoretical discussion and simulations above and the resulting estimates imply that the long-run mean of the 10-year Treasury bond after 2003 was 4.5%.

We modify the AR(1) model slightly by treating the one-step-ahead SPF forecast as known in order to account for private forecasters access to additional high frequency information

$$\hat{x}_{t+h|t} = \mu_t + 0.59^{h-1} \left(SPF_{t+1|t} - \mu_t \right),$$

where the long-run mean is $\mu_{t,AR(1)} = 4.5$. We seek to improve upon the longer-horizon SPF forecasts by modifying this model further using smooth robust predictors. In this setting, the general local estimator of the long-run mean of the predictors is expressed as

$$\mu_{t,SH(\gamma,n)} = \frac{1}{n}\left[SPF_{t+1|t} + \sum_{j=0}^{n-2} x_{t-j} + \gamma\frac{0.81}{1-0.81}\left(SPF_{t+1|t} - x_{t-n+1}\right)\right],$$

where if $\gamma = 0$ we obtain the smooth random walk and if $\gamma = 1$ we obtain the smooth robust predictor. If $n = 1$ we obtain the non-smoothed variants. For the smooth variants we choose an estimation window of $n = 4$ as suggested by Hendry (2006).

We evaluate the relative performance of these forecasts using RMSFEs calculated from 2009 Q1 - 2020 Q1. Table 7.2 illustrates that accounting for the SPF information advantage at $h = 1$ implies that all forecasts do equally well at that horizon. However, the robust predictors quickly outperform the SPF thereafter with every predictor doing better by $h = 2$. All of the smoothed robust predictors outperform their un-smoothed counterparts. The smooth random walk does best for all horizons followed closely by the smooth average. The results are consistent with our theory and simulations when using a fixed estimation window which suggests that the smooth random walk dominates because the smooth robust predictor suffers from a higher forecast error variance. However, unlike in the simulations, in the application there is not enough differentiation in the biases for the smooth average to improve on the smooth random walk.

Table 7.2: Accuracy of 10-year Treasury Yield Forecasts (2009 Q1 - 2020 Q1)

	RMSFE	**Relative RMSFE**					
h	**SPF**	**RW**	**RP**	**RA**	**SRW**	**SRP**	**SA**$_{RW,RP}$
1	**0.18**	1.00	1.00	1.00	1.00	1.00	1.00
2	0.50	**0.90**	0.95	0.91	**0.90**	**0.90**	**0.90**
3	0.70	0.85	0.94	0.88	**0.84**	0.85	**0.84**
4	0.86	0.79	0.88	0.82	**0.77**	0.78	**0.77**
5	1.05	0.72	0.81	0.76	**0.70**	0.72	0.71

Notes: RP: Robust Predictor, RW: Random Walk , SRP: Smooth Robust Predictor, RA: Robust Average, SRW: Smooth Random Walk , SA: Smooth Average. Average is equally weighted. Smoothed forecasts are based on a fixed window of $n = 4$. Bolded values indicate the lowest RMSFE at each horizon.

8 Conclusions

As almost all macroeconomic forecasting models can be expressed as equilibrium correction specifications with long-run means, and shifts in that mean can lead to systematic forecast failure, we explore approaches for transforming economic forecasts after, or in the midst of, ongoing structural breaks using local estimates of the long-run mean. We show that as the forecast moves away from the initial break date then it is possible to mitigate persistent forecast failure at longer forecast horizons even when the forecast model is incorrectly specified, an incorrect transformation of the mean is used, and the true parameters are unknown. This is done by recognizing that naive forecasting predictors can be thought of as using local estimators of the long-run mean, with the benefit of adapting to any new long-run mean very quickly. We illustrate the advantages of these predictors and show how they can be embedded within economic and policy models. We then show that their costs can be mitigated by extending the window over which estimation of the long-run mean occurs to account for information since the break. Simulations show the potential benefits of smoothing and two applications forecasting U.K. productivity and U.S. 10-year bond yields show that it is possible to use smooth robust methods to dramatically improve multi-horizon forecast performance over forecasts from economic policy models and professional forecasters.

Our results have important implications for macroeconomic forecasts following COVID-19. The pandemic has had a massive impact on many important macroeconomic time-series and their future trajectories remains deeply uncertain depending on the widespread distribution of a vaccine and supporting macroeconomic policies. In this period of uncertainty about the shape of the eventual recovery, smooth robust methods may be able to provide guidance by using the relatively limited amount of data since the start of the pandemic to forecast the recovery. In particular, combining the smooth robust predictor and the smooth random walk using an expanding estimation window of the long-run mean may perform well to capture and understand the nascent recovery.

References

Atkeson, A. and Ohanian, L. E. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

Ball, L. and Mazumder, S. (2011). Inflation dynamics and the great recession. *Brookings Papers on Economic Activity*, 2011(1):337–381.

Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bauer, M. D. and Rudebusch, G. D. (2013). What caused the decline in long-term yields? *FRBSF Economic Letter*, 19:208.

Campbell, J. Y., Pflueger, C., and Viceira, L. M. (2020). Macroeconomic drivers of bond and equity risks. *Journal of Political Economy*, 128(8):3148–3185.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2015a). Robust Approaches to Forecasting. *International Journal of Forecasting*, 31(1):99–112.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2018). Selecting a Model for Forecasting. Department of Economics Discussion Paper Series No. 861, Univeristy of Oxford, Oxford U.K.

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

Clements, M. P. and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Zeuthen Lecture Book Series. Cambridge, MA: MIT Press.

Cooper, J. P. and Nelson, C. R. (1975). The Ex-Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and some results on Composite Predictors. *Journal of Money, Credit and Banking*, 7(1):1–32.

Giraitis, L., Kapetanios, G., and Price, S. (2013). Adaptive Forecasting in the Presence of Recent and Ongoing Structural Change. *Journal of Econometrics*, 177(2):153–170.

Hendry, D. F. (2006). Robustifying Forecasts from Equilibrium-Correction Systems. *Journal of Econometrics*, 135(1-2):399–426.

Hendry, D. F. (2018). Deciding between alternative approaches in macroeconomics. *International Journal of Forecasting*, 34(1):119–135.

Hendry, D. F. and Clements, M. P. (1994). On a Theory of Intercept Corrections in Macro-economic Forecasting. In Holly, S., editor, *Money, Inflation and Employment: Essays in Honour of James Ball*, pages 160–182. Aldershot: Edward Elgar.

Hendry, D. F. and Clements, M. P. (2004). Pooling of Forecasts. *The Econometrics Journal*, 7(1):1–31.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Kladivko, K., Österholm, P., et al. (2019). Market Participants' Forecasts of Financial variables–Can Survey Data Outperform the Random Walk? Working Paper 10/2019, Örebro University School of Business, Örebro, Sweden.

Nelson, C. R. (1972). The Prediction Performance of the FRB-MIT-PENN Model of the US Economy. *American Economic Review*, 62(5):902–917.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Smith, B. B. (1929). Judging the Forecast for 1929. *Journal of the American Statistical Association*, 24(165A):94–98.

Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, 1:135–196.

Wright, J. H. (2013). Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior. *Journal of Applied Econometrics*, 28(5):762–776.

A Appendix

A.1 Additional Simulation Results (Finite Window: n=4)

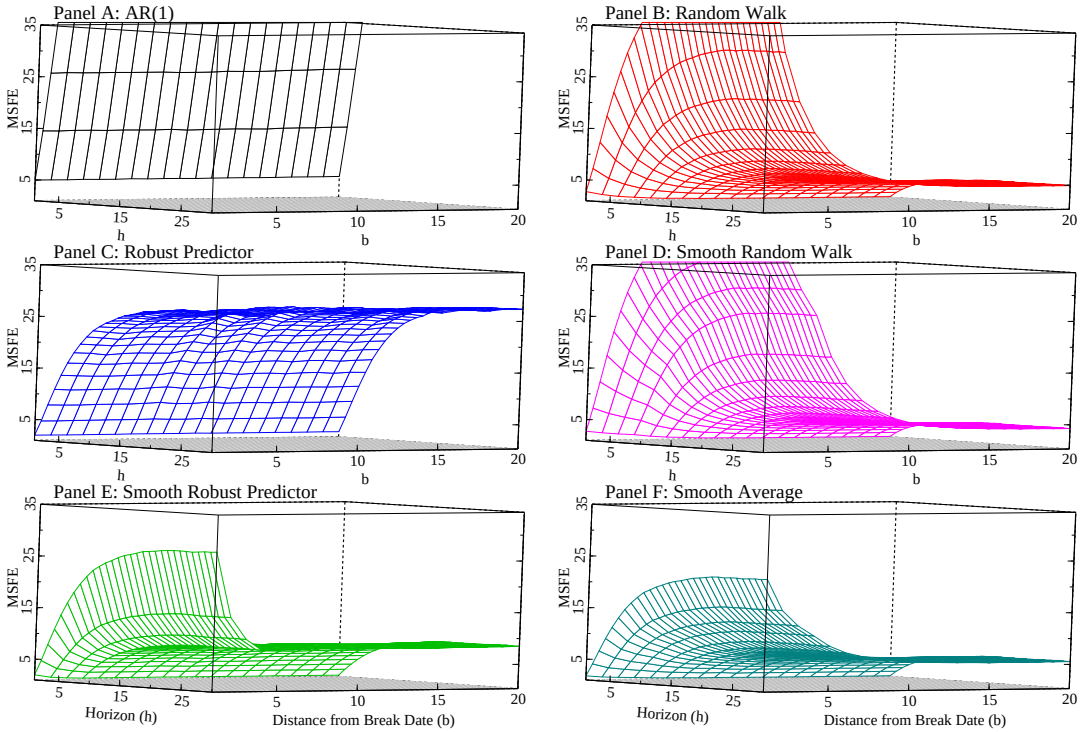


Figure A.1: Simulations: Break in Long-Run Mean without Slope Shift

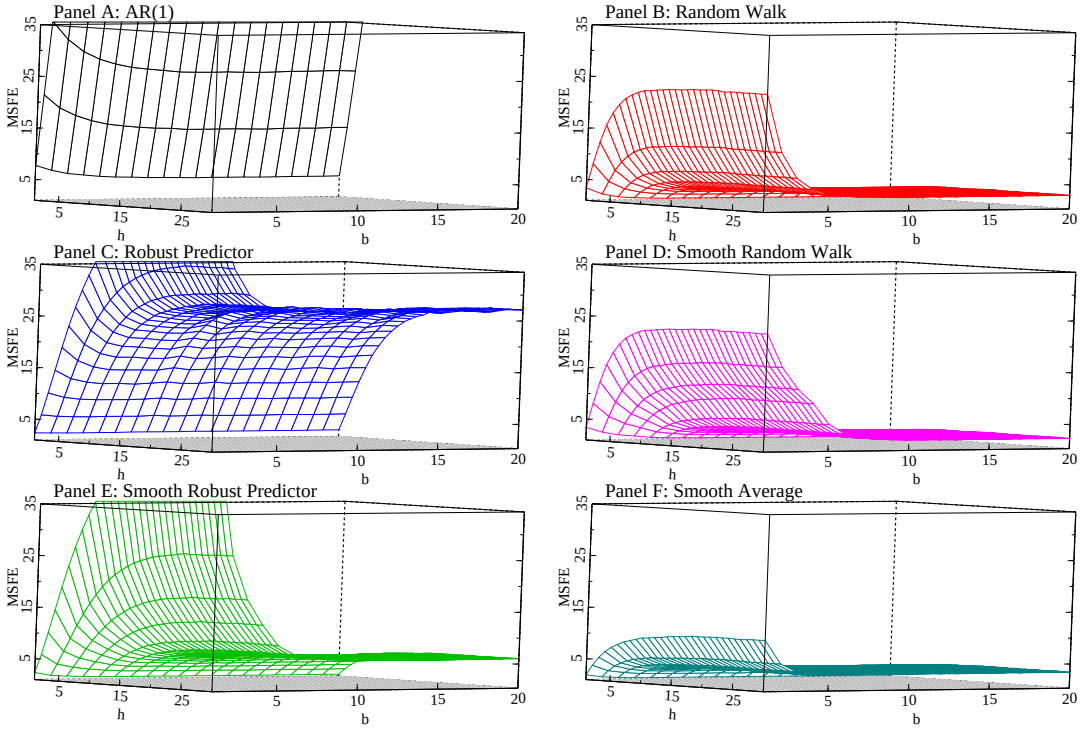
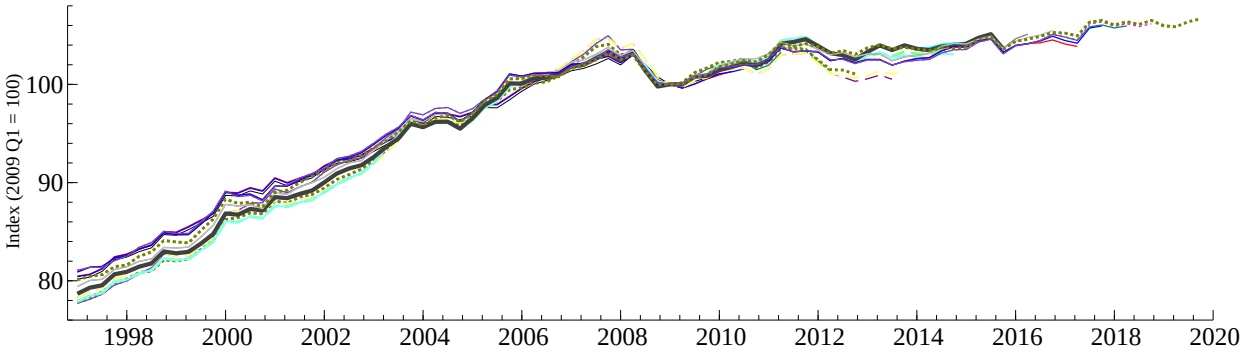


Figure A.2: Simulations: Break in Long-Run Mean due to Slope Shift

A.2 Historical Data Vintages of U.K. Productivity



Notes: The data was obtained from historical releases from the Office for National Statistics and the National Archives. Output is measured as GVA excluding Oil and Gas (ONS code UIZY or KLS2) while hours worked is measured as total actual weekly hours worked (ONS code YBUS). Both variables are seasonally adjusted. The source of most historical variation is due to changing base years for which the chained values of GVA are calculated. All vintages are re-normalized so that 2009 Q1=100.

Figure A.3: Historical Data Vintages of U.K. Productivity (Output per Hour Worked)