Dynamic Linear Models with Adaptive Discounting

Abstract

Dynamic linear models with discounting are state-space models that are sufficiently flexible interpretable, and computationally efficient. As such they are increasingly applied in economics and finance. Successful modeling and forecasting with such models depends on an appropriate choice of the discount factor. In this work we develop an adaptive approach to sequentially estimate this parameter, which relies on the minimisation of the one-step-ahead forecast error. Simulated data and an in-depth empirical application to the problem of forecasting quarterly UK house prices shows that our approach can achieve significant improvement in forecast accuracy at a computational cost that is orders of magnitude smaller than approaches based on sequential Monte Carlo. We also conduct an extensive evaluation of diverse forecast combination methods on the task of predicting UK house prices. Our results indicate that a recent density combination method can substantially improve forecast accuracy.

Keywords: Dynamic linear model, Adaptive discount factor, Housing market

1. Introduction

Dynamic linear models (DLMs) provide a very flexible yet fairly simple state space formulation to analyse dynamic phenomena and evolving data generating processes (DGPs) (West and Harrison, 1997; Prado and West, 2010). In response to accumulating evidence that economic and financial predictors have short-lived predictive content and econometric relationships are unstable (see Rossi (2013) for a review), there has been a substantial growth of interest in the application of DLMs, and especially DLMs with discounting (D-DLMs) (Dangl and Halling, 2012; Koop and Korobilis, 2012; Koop and Tole, 2013; Byrne et al., 2018; Zhao

et al., 2016). Discounting alleviates the need to specify the covariance matrix of the state transition (system) equation (West and Harrison, 1997, Section 6.3). However, the forecast performance of D-DLMs depends critically on the choice of the discount factor; a scalar parameter that allows a continuous trade-off between estimation in a static environment, and re-initialising the estimation process by discarding all past information.

The most common approach in the literature is to treat the discount factor as a user-specified parameter, set *a priori* to a constant value. An alternative is to specify a grid of discount factor values, and through Bayesian updating sequentially estimate the posterior probability of each (West and Harrison, 1989; Zhao et al., 2016; Dangl and Halling, 2012). More recently, Irie et al. (2022) proposed a nonlinear state space model for the evolution of the discount factor. The resulting state-space model does not admit conjugate Bayesian analysis, and instead sequential Monte Carlo is used.

In this work, we develop a DLM with adaptive discounting (ADLM). The central idea underlying our approach is that the optimal value of the discount factor minimises the expectation of the one-step-ahead squared forecast error. This formulation allows us to approximate the optimal discount factor sequentially through stochastic gradient descent (SGD) (Bottou et al., 2018). This approach does not require the user to pre-select appropriate value(s) for this parameter, or to specify a model for the evolution of the discount factor (and its hyperparameters). Furthermore, the proposed approach does not increase the computational complexity of filtering with D-DLMs. On the other hand, the performance of SGD algorithms relies on the appropriate choice of their parameters, and most notably the step-size. We propose an ensemble approach to ameliorate this issue.

We assess the forecast performance of ADLMs in the context of an in-depth study of forecasting UK house prices at the national and regional level. Despite the vast interest in the dynamics of international housing markets, sparked by the latest boom-bust episode in real estate markets and its decisive role in the Great Recession, the academic literature on house price forecasting remains

relatively small (especially when compared to other financial assets), and mainly concentrates on the US market (Rapach and Strauss, 2009; Ghysels et al., 2013; Bork and Møller, 2015). In the UK, similarly to the US, housing activities account for a large fraction of GDP and of households' expenditures. Real estate property comprises the largest component of private wealth (excluding private pensions), and mortgage debt constitutes the main liability of households (Office for National Statistics, 2018). Thus, accurate forecasts of UK real estate prices are crucial for private investors and policy makers.

As we discuss in the empirical application section, there is clear evidence of time-varying patterns in the in-sample relationship between real estate valuations in the UK and conditioning macro and financial variables. This is not unexpected as similar findings have been previously reported in the literature (Aizenman and Jinjarak, 2014; Anundsen, 2015; Paul, 2020). Dynamic econometric models are therefore essential to successfully model and forecast the UK housing market. The results of the empirical evaluation show that ADLMs can achieve significant forecast accuracy improvements against competing approaches, at a fraction of the computational cost.

We complement our study of UK housing markets by evaluating a number of forecast combination methods. We consider a diverse set of methods including: density combination methods (Hall and Mitchell, 2007; Geweke and Amisano, 2011; Billio et al., 2013); the partially egalitarian LASSO (Diebold and Shin, 2019); a recent algorithm from the prediction with expert advice literature, ConfHedge (V'yugin and Trunov, 2019); as well as Bayesian model averaging, and simple averaging.

The remaining paper is organised as follows. In Section 2 we first present the D-DLM model, and then, in Section 2.1, we outline existing approaches to estimate the discount factor. The proposed adaptive approach is developed in Section 2.2. In Section 3, we employ simulated time-series to assess different approaches to tune the discount factor. Section 4 provides a comparative evaluation of different approaches to estimate the discount factor in D-DLMs with application to UK housing markets. In Section 5 we evaluate empirically

a number of forecast combination methods on the task of forecasting UK house prices. The paper ends with concluding remarks in Section 6.

2. Dynamic linear model with adaptive discounting

We consider DLMs in which the unobserved coefficient vector, $\theta_t \in \mathbb{R}^d$, follows a driftless random walk,

$$y_t = x_t^\top \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, v), \tag{1}$$

$$\theta_t = \theta_{t-1} + w_t, \qquad w_t \sim N(0, W_t). \tag{2}$$

The variance v in Eq. (1) is unknown and we impose an inverse-gamma prior for this parameter,

$$v|\mathcal{D}_0 \sim IG\left(\frac{n_0}{2}, \frac{n_0 s_0}{2}\right), \tag{3}$$

where $\mathcal{D}_t$ represents all the information available at time t, and $s_t = \mathbb{E}(v^{-1}|\mathcal{D}_t)^{-1}$ is the point estimate for v at time t. All quantities with a time-stamp zero represent prior information (belief). In the following it is convenient to express all covariance matrices as multiples of v. The prior distribution for the coefficient vector conditional on v is,

$$\theta_0|v \sim N(m_0, v\widetilde{C}_0), \tag{4}$$

while the marginal is a Student-t distribution with n_0 degrees of freedom, $\theta_0 \sim T(m_0, s_0\widetilde{C}_0, n_0)$. This specification enables a conjugate Bayesian analysis. Given the posterior (filtering) distributions at time $t-1$,

$$v|\mathcal{D}_{t-1} \sim IG(n_{t-1}/2, n_{t-1}s_{t-1}/2), \tag{5}$$

$$\theta_{t-1}|\mathcal{D}_{t-1} \sim T(m_{t-1}, s_{t-1}\widetilde{C}_{t-1}, n_{t-1}), \tag{6}$$

the prior (predictive) distribution for θ_t is,

$$\theta_t|\mathcal{D}_{t-1} \sim T(m_{t-1}, s_{t-1}(\widetilde{C}_{t-1} + \widetilde{W}_t), n_{t-1}), \tag{7}$$

where, $\widetilde{W}_t = v^{-1}W_t$. The structure and magnitude of $\widetilde{W}_t$ are crucially important for successful modelling and forecasting. This matrix controls the extent

of the stochastic variation in the evolution of the state-space model, and hence determines stability over time. Eq. (7) suggests that $s_{t-1}\widetilde{W}_t$ leads to an increase in uncertainty, or equivalently a loss of information, about θ_t between consecutive time steps. Estimating $\widetilde{W}_t$ is computationally very expensive if there is no optimal value for this matrix that is suitable for all times (West and Harrison, 1997, Sec. 6.3). Discounting is a simple and practically critical method to overcome this. The central idea underlying discounting is to set,

$$\widetilde{W}_t = \frac{1-\delta}{\delta}\widetilde{C}_{t-1}, \tag{8}$$

where $\delta \in (0, 1]$ is the *discount factor*, also called the forgetting factor. We call this model DLM with discounting (D-DLM). From the above specification of $\widetilde{W}_t$ it follows that,

$$\theta_t|\mathcal{D}_{t-1} \sim T(m_{t-1}, \frac{s_{t-1}}{\delta}\widetilde{C}_{t-1}, n_{t-1}). \tag{9}$$

According to Eq. (9), the prior variance of θ_t is that of a model with no system stochastic variation ($W_t = 0$) times a correction factor of $1/\delta$ that inflates this variance (Prado and West, 2010). The choice of δ allows a continuous range between estimation in a static environment ($\delta = 1$), and completely discarding all past data (δ approaching zero).

The one-step-ahead prior (predictive) distribution for y_t is,

$$y_t|\mathcal{D}_{t-1} \sim T\left(x_t^\top m_{t-1}, s_{t-1}\tilde{q}_t, n_{t-1}\right), \tag{10}$$

$$\tilde{q}_t = \delta^{-1}(x_t^\top \widetilde{C}_{t-1}x_t + \delta). \tag{11}$$

After observing y_t we can obtain the posterior (filtering) distributions as follows. The posterior distribution $v|\mathcal{D}_t \sim IG(n_t/2, n_t s_t/2)$ where,

$$n_t = n_{t-1} + 1, \tag{12}$$

$$s_t = \frac{n_t - 1}{n_t}s_{t-1} + \frac{\hat{\varepsilon}_t^2}{n_t\tilde{q}_t}, \tag{13}$$

where $\hat{\varepsilon}_t = y_t - x_t^\top m_{t-1}$, is the forecast error at time t. The posterior distribution

$\theta_t|\mathcal{D}_t$ is,

$$\theta_t|\mathcal{D}_t \sim T(m_t, s_t\widetilde{C}_t, n_t), \tag{14}$$

$$m_t = m_{t-1} + A_t\hat{\varepsilon}_t, \tag{15}$$

$$\widetilde{C}_t = \delta^{-1}(I - A_t x_t^\top)\widetilde{C}_{t-1}, \tag{16}$$

$$A_t = \frac{1}{\delta\tilde{q}_t}\widetilde{C}_{t-1}x_t = \frac{1}{x_t^\top\widetilde{C}_{t-1}x_t + \delta}\widetilde{C}_{t-1}x_t. \tag{17}$$

The vector A_t is known as the Kalman gain, or the adaptive coefficient vector.

The choice of δ critically affects the forecast performance of D-DLMs. In the next section we discuss existing approaches to estimate the discount factor, while in Section 2.2 we develop our adaptive approach.

2.1. Specifying the discount factor

Prado and West (2010) recommend using $\delta \in (0.9, 1)$. Other authors consider $\delta \in \{0.95, 0.99\}$ (Raftery et al., 2010; Koop and Korobilis, 2012; Koop and Tole, 2013). An alternative approach is to specify a grid of δ values, and update the posterior probability of each value sequentially,

$$p(\delta_j|\mathcal{D}_t) \propto p(y_t|\mathcal{D}_{t-1}, \delta_j)\, p(\delta_j|\mathcal{D}_{t-1}). \tag{18}$$

Posterior probabilities updated in the above manner are known to rapidly concentrate on a small set of δ values, and eventually allocate all probability mass to a single value. This can lead to significantly down-weighting δ values that could perform well in the future, and thus can degrade predictive performance (Zhao et al., 2016). Model probability power discounting (MPD) is an approach to overcome this limitation through use a *power discount factor*, $\alpha \in (0, 1]$,

$$p(\delta_j|\mathcal{D}_t) \propto p(y_t|\mathcal{D}_{t-1}, \delta_j)\, p(\delta_j|\mathcal{D}_{t-1})^\alpha, \tag{19}$$

which forces $p(\delta_j|\mathcal{D}_t)$ to depend more heavily on recent performance (West and Harrison, 1989). Zhao et al. (2016) propose to consider a grid of values for both δ and α.

Irie et al. (2022) propose a dynamic discount factor formulation for a Poisson-Gamma state space model for web traffic data. The nonlinear state space model

which describes the evolution of δ_t can be directly incorporated in a DLM. We therefore outline it here and consider it in the experimental results section. The discount factor is determined through the variable g_t which follows a stationary AR(1) process,

$$g_t = \mu(1-\phi) + \phi g_{t-1} + \eta_t. \tag{20}$$

where $\phi \in (0,1)$, and $\eta_t \sim N(0, \sigma_\eta^2)$, and $g_0 \sim N(\mu, \sigma_\eta^2/\sqrt{1-\phi^2})$. This choice of prior for g_0 implies that the g_t process is stationary, and the marginal distribution of g_t is identical to that of g_0 for all t. The discount factor, δ_t is obtained through,

$$\delta_t = \frac{1}{1+e^{-g_t}}, \tag{21}$$

or equivalently $g_t = \text{logit}(\delta_t)$, where logit stands for the logistic transformation. Irie et al. (2022) propose a Bayesian approach to estimate the AR(1) parameters $(\mu, \phi, \sigma_\eta^2)$. Specifically, they recommend to use the parameterisation $\phi_0 = \mu(1-\phi)$, $\phi_1 = \phi$ and $w = \sigma_\eta^{-2}$ and then assume a Normal-Gamma prior,

$$(\phi_0, \phi_1), w|\mathcal{D}_0 = NG(m_0, \frac{b_0}{a_0}C_0, a_0). \tag{22}$$

The authors use particle learning (Carvalho et al., 2010) to perform filtering and parameter estimation. Compared to a standard DLM with discounting this approach increases computational cost by a factor equal to the number of particles used in the particle learning algorithm.

2.2. Adaptive discounting DLM

The central idea underlying our approach is that the optimal value of the discount factor minimises the expectation of the one-step-ahead squared forecast error,

$$\delta_t^\star = \arg\min_\delta \frac{1}{2}\mathbb{E}_{(X_t,Y_t)}\left[\left(y_t - x_t^\top m_{t-1}\right)^2\right]. \tag{23}$$

The above expectation depends on δ because $m_{t-1} = \mathbb{E}[\theta_t|\mathcal{D}_{t-1}]$ is a function of δ (see Eqs. (15) and (17) above). Since the expectation in Eq. (23) is not

available in analytical form it is not feasible to directly minimise it. However, the observed squared forecast error at time t,

$$\hat{\varepsilon}_t^2 = \left(y_t - x_t^\top m_{t-1}\right)^2, \tag{24}$$

is an unbiased estimator of $\mathbb{E}_{(X_t,Y_t)}[\left(y_t - x_t^\top m_{t-1}\right)^2]$. Therefore we can use stochastic gradient descent (SGD) to solve the optimisation problem in (23) using only information from (24). SGD algorithms have been studied through stochastic approximation (Benveniste et al., 1990). It has been shown that their convergence can be established under relatively weak conditions: the expectation which is minimised must be reasonably well behaved, and the learning rate must decrease appropriately (Bottou et al., 2018).

To apply SGD, we need an expression for the derivative of the forecast error with respect to δ. To this end, we apply the chain rule,

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial\delta}\left\{\hat{\varepsilon}_t^2\right\} &= \frac{1}{2}\nabla_{m_{t-1}}\left\{\hat{\varepsilon}_t^2\right\}\nabla_\delta m_{t-1}, \\ &= -\hat{\varepsilon}_t x_t^\top \nabla_\delta m_{t-1}. \end{aligned} \tag{25}$$

A recursive formula for the derivative of m_t with respect to δ is obtained by differentiating Eq. (15),

$$\nabla_\delta m_t = \nabla_\delta m_{t-1} + \nabla_\delta A_t \hat{\varepsilon}_t - A_t x_t^\top \nabla_\delta m_{t-1} \tag{26}$$

$$= (I - A_t x_t^\top)\nabla_\delta m_{t-1} + \nabla_\delta A_t \hat{\varepsilon}_t. \tag{27}$$

Differentiating Eq. (17) with respect to δ yields,

$$\nabla_\delta A_t = \frac{1}{\delta\tilde{q}_t}\left[\frac{\partial\widetilde{C}_{t-1}}{\partial\delta} - \frac{1}{\delta\tilde{q}_t}\left(x_t^\top\frac{\partial\widetilde{C}_{t-1}}{\partial\delta}x_t + 1\right)\widetilde{C}_{t-1}\right]x_t. \tag{28}$$

We obtain a recursive formula for $\frac{\partial\widetilde{C}_t}{\partial\delta}$ by differentiating Eq. (16),

$$\begin{aligned}\frac{\partial\widetilde{C}_t}{\partial\delta} =& \delta^{-1}\frac{\partial\widetilde{C}_{t-1}}{\partial\delta} - \delta^{-2}\widetilde{C}_{t-1} + \frac{\tilde{q}_t + x_t^\top\frac{\partial\widetilde{C}_{t-1}}{\partial\delta}x_t + 1}{\delta^3\tilde{q}_t^2}\widetilde{C}_{t-1}x_t x_t^\top\widetilde{C}_{t-1} \\ &- \frac{1}{\delta^2\tilde{q}_t}\left(\frac{\partial\widetilde{C}_{t-1}}{\partial\delta}x_t x_t^\top\widetilde{C}_{t-1} + \widetilde{C}_{t-1}x_t x_t^\top\frac{\partial\widetilde{C}_{t-1}}{\partial\delta}\right). \end{aligned} \tag{29}$$

All the above derivatives can be computed in $\mathcal{O}(d^2)$, where d is the dimensionality of θ, since computations can be ordered so that the most expensive operation is matrix-vector multiplication. Therefore, computing $\frac{\partial}{\partial \delta}\left\{\hat{\varepsilon}_t^2\right\}$ does not increase the computational complexity of DLM filtering. Having an expression for $\frac{\partial}{\partial \delta}\left\{\hat{\varepsilon}_t^2\right\}$ allows us to apply any SGD algorithm. In all the experimental results in this paper we use the influential adaptive moment estimation (ADAM) algorithm (Kingma and Ba, 2015).

Although our analysis focuses on one-step-ahead forecasts, there is a large and growing literature that examines the predictability of numerous important economic and financial variables at different horizons. Furthermore, practitioners in policy institutions and in the private sector care about both short term and longer term predictions. The adaptive discounting formulation outlined in this section can be applied to perform direct k-step ahead forecasting as in Koop and Korobilis (2012) and Catania and Nonejad (2018).

3. Simulation Experiments

In this section we use simulated data to assess the ability of different approaches to sequentially estimate the D-DLM discount factor. We consider abruptly changing and gradually evolving DGPs. In all cases, the covariates are sampled from a Gaussian distribution $x_t \sim N(0, I)$, $t = 1, \ldots, 1000$, while the noise term in the measurement equation, Eq. (1), has a variance of 0.1. Next we specify the parameter settings for all the considered methods. The parameter values reported in Section 3.1 are used throughout the paper.

3.1. Competing methods and parameter settings

For all methods and in every experiment the priors for D-DLM are specified as, $\theta_0|\mathcal{D}_0 \sim N(m_0, s_0\widetilde{C}_0)$, with $m_0 = 0$, $s_0 = 1$, and $\widetilde{C}_0 = 100I$, while $v|\mathcal{D}_0 \sim IG(n_0/2, n_0 s_0/2)$ with $n_0 = 1$ (Catania and Nonejad, 2018). We compare ADLM with two methods for sequentially estimating the discount factor. The first employs a user-defined grid of values for δ and for the model probability power discount factor α, as in Zhao et al. (2016). Using standard Bayesian

updating we obtain the posterior probability of each δ value in the grid at every time step (West and Harrison, 1989, Ch. 12). We refer to this model as Grid-DLM. For the discount factor δ, we use a grid of values in the range $[0.95, 0.999]$ with a step-size of 0.005; while for α we use a grid in the range $[0.95, 1]$ again with stepsize 0.005 (Zhao et al., 2016). We assume a uniform prior for all δ and α values.

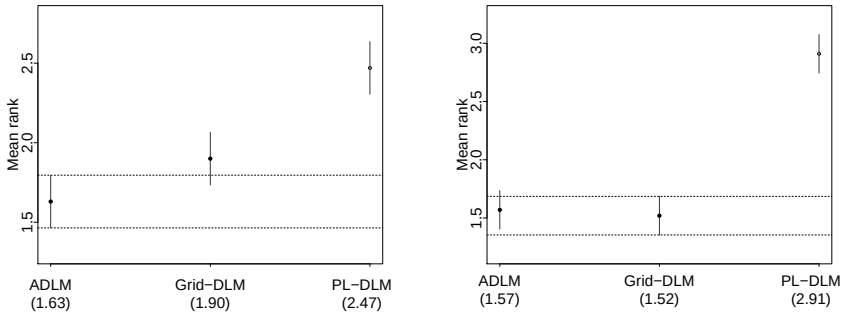
The second method we consider is based on Irie et al. (2022) and assumes that the discount factor evolves according to Eqs. (20) and (21). Because it uses particle learning we refer to this method as PL-DLM. For PL-DLM we use the following values for the prior in Eq. (22): $m_0 = [(1 - 0.9)\text{logit}(0.90), 0.9]$, $C_0 = 0.05^2 I$, $a_0 = 10$ and $b_0 = 5$. This reflects the preference for the values of the AR(1) process in Eq. (20) $(\mu, \phi, \sigma_\eta^2) = (\text{logit}(0.9), 0.9, 0.5)$. These hyperparameters for the prior are recommended by Irie et al. (2022). Finally the number of particles is set to 1000.

In ADLM we adaptively tune δ through the recently proposed and influential SGD algorithm called ADAM (Kingma and Ba, 2015). The discount factor is initialised at $\delta = 0.99$ and is constrained to lie in the interval $\delta_t \in [0.7, 0.999]$. ADAM employs three parameters (ξ, b_1, b_2), where ξ is the stepsize, while b_1, b_2 are momentum parameters. The choice of ξ and whether the step-size is constant or gradually decreasing affects the performance of ADAM. To ameliorate this problem we propose to use not one but a number of parameter specifications, and combine the resulting models. Specifically, we consider $\xi \in \{0.005, 0.1, 0.2, 0.3\}$, $b_1 \in \{0, 0.8\}$ and $b_2 = 0.9$. These settings combine parameter values that are appropriate for different types of change. A small and constant step-size like $\xi = 0.005$ allows ADAM to adapt towards the value of the optimal discount factor gradually but continuously. A larger constant step-size like $\xi = 0.1$ allows rapid adaptation in the presence of abrupt changes. However when the optimal discount factor is constant over time using a constant step-size causes the estimated δ to fluctuate around the optimal value. For this reason we also consider a decreasing step-size schedule, $\xi_t = \xi/t$, with $\xi \in \{0.1, 0.2, 0.3\}$. Finally having $b_1 = 0$ ADAM becomes equivalent to RMSprop which is another

SGD algorithm which has been shown to be effective in numerous applications. Each ADAM configuration produces a different D-DLM. The forecasts of these D-DLMs are combined through the associated model probabilities, which are updated according to the MPD approach in Eq. (19). We assume a uniform prior for all ADAM configurations and use the same grid of power discount factor values, α, as in Grid-DLM.

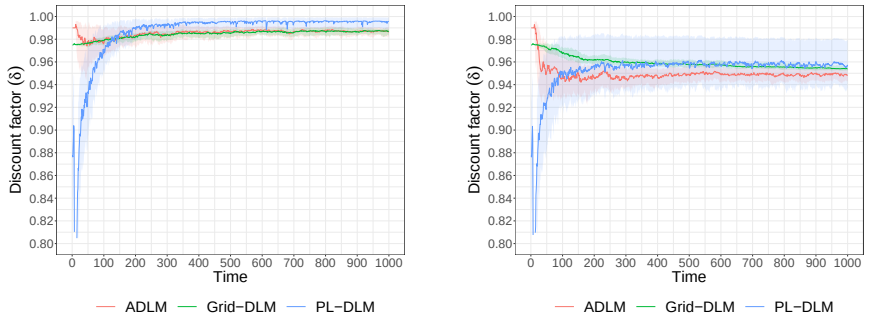
3.2. Gradual drift

In this section we consider time series where the coefficient vector changes gradually over time. We consider two such cases. In the first the DGP is a D-DLM, as defined in Eqs. (1) to (17). Our objective in this case is to compare both the forecast accuracy of ADLM, Grid-DLM and PL-DLM and to assess their ability to identify the true discount factor. We consider two values of the discount factor, $\delta \in \{0.95, 0.99\}$, and for each value we simulate 100 time series. We use Friedman's nonparametric test (Friedman, 1937, 1939) to compare the forecast performance of different methods across the 100 experiments. The null hypothesis of this test is of equal predictive performance. The p-value of this test is always extremely low which allows us to proceed with the multiple comparisons with the best (MCB) test (Koning et al., 2005).



(a) $\delta = 0.99$ (b) $\delta = 0.95$

Figure 1: MCB tests on time-series data sampled from the state-space model assumed by D-DLM for $\delta \in \{0.95, 0.99\}$.



(a) $\delta = 0.99$ (b) $\delta = 0.95$

Figure 2: Evolution of median and interquartile range for δ_t, for time-series data sampled from the state-space model assumed by D-DLM with $\delta \in \{0.95, 0.99\}$.

The outcomes of the MCB test are shown in Figure 1, while Figure 2 illustrates the median (solid line) and the interquartile range (shaded region) for the discount factor, δ_t, for each method. As shown in Figure 1(a) when the true value of the discount factor is 0.99 ADLM achieves the lowest median rank but its performance is not statistically significantly better than that of Grid-DLM. Figure 2(a) shows that in this case the median estimated discount factor by ADLM and Grid-DLM tends towards the true value of 0.99. The median value of δ_t obtained by PL-DLM instead converges towards unity and has higher variability.

When $\delta = 0.95$ the two best performing methods are ADLM and Grid-DLM, while PL-DLM performs significantly worse. Figure 2(b) shows that ADLM converges fastest to the true discount factor value but the estimate of δ_t exhibits higher variability compared to Grid-DLM. On the other hand the median estimate by Grid-DLM converges much more slowly. The median estimate by PL-DLM is consistently higher than the true value and the variability of is also the highest among the considered methods.

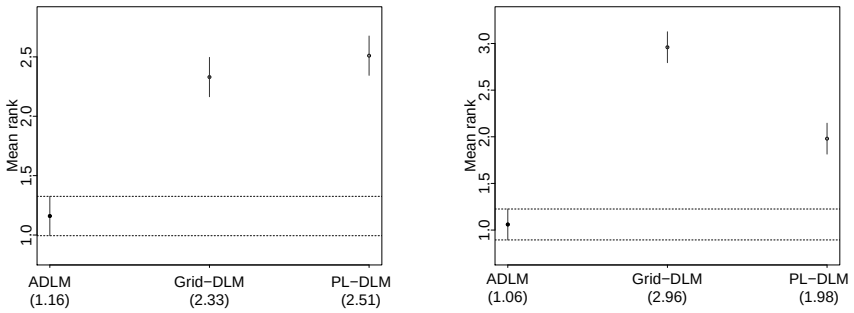
The previous example allows us to assess the ability of different methods to estimate the optimal discount factor. However, the purpose of D-DLMs is to provide a flexible framework to track time-varying DGPs when the true

dynamics are unknown. We therefore also consider a case of gradual change in which the coefficient vector follows a random walk (as in Eq. (2)) with an unknown covariance matrix W_t. We set $W_t = \sigma^2_{w,t} I$, and consider three cases: in the first two $\sigma^2_{w,t}$ is constant and takes values $\sigma^2_{w,t} \in \{5 \cdot 10^{-4}, 10^{-3}\}$. In the third case we have periods of very slow change $\sigma^2_{w,t} = 2 \cdot 10^{-4}$ that last 200 time-steps, being followed by shorter periods (100 time-steps) of more rapid change $\sigma^2_{w,t} = 10^{-3}$. As before we perform first Friedman's nonparametric test and after rejecting the null hypothesis we use the MCB test. The results of the latter are illustrated in Figure 3. As the figure shows ADLM clearly outperforms the other two methods in the first two cases. In the last case where $\sigma^2_{w,t}$ changes abruptly between two regimes ADLM still achieves the lowest median rank but its performance is statistically better than that of PL-DLM.

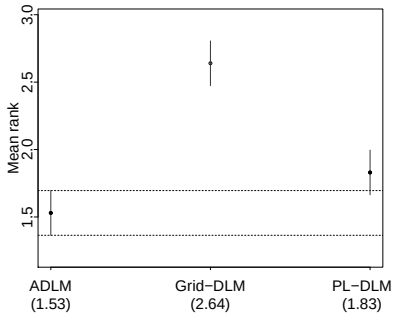
3.3. Abrupt change

For our last simulation experiment, we consider an environment in which θ_t changes at distinct points in time, and is constant in-between change points. We simulate from a single time series of θ_t and create 100 time series of y_t by different realisations of the noise term in the measurement equation, Eq. (1). The coefficient vector is determined according to $\theta_t = \gamma_t \theta_{t-1}$, for $t = 1, \ldots, 1000$, and $\theta_0 = (3, 2, 1, -1, -2)^\top$. We set $\gamma_t = 0.5, 2$ at $t = 300, 700$, respectively, and $\gamma_t = 1$ at all other time-steps.

Figure 4(b) shows that during the first 300 time-steps all three methods gradually increase δ_t towards unity, with PL-DLM being the fastest. In response to the first change point at $t = 300$ all methods reduce the discount factor. The sequential Monte Carlo (SMC) approach adopted by PL-DLM enables it to sharply reduce δ_t immediately after the change point, and subsequently rapidly increase it towards unity. This is ideal in an abruptly changing environment, which is the setting for which PL-DLM was originally designed for (Irie et al., 2022). ADLM also decreases sharply the value of δ_t after the first change point but the speed of both the decrease and the subsequent recovery are slower compared to PL-DLM. Figure 4(b) shows that following the first change point



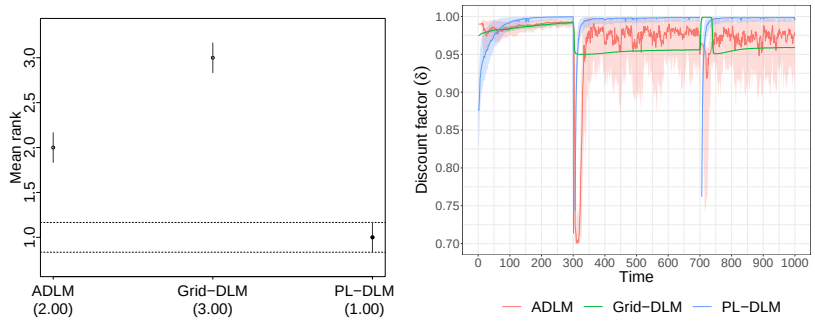
(a) $\sigma_w^2 = 5 \cdot 10^{-4}$ (b) $\sigma_w^2 = 10^{-3}$



(c) Changing σ_w^2

Figure 3: MCB tests on time-series data sampled from the state-space model in which the coefficient vector follows a random walk with an unknown covariance matrix $W_t = \sigma_{w,t}^2 I$.

there is an increase in the variability of δ_t by ADLM. This is because after the change point ADLM assigns a large weight to the ADAM variant with a large constant step-size. In response to the second change point PL-DLM and ADLM behave as they did after the first. Figure 4(b) shows that Grid-DLM behaves differently. In response to the first change point the median discount factor becomes very close to the smallest value in the grid, 0.95. The subsequent recovery is very slow, and by the time of the second change point δ_t is still very close to 0.95. In response to the second change point the discount factor first increases and then decreases again towards the minimum value in the grid. The above described behaviour of the discount factor explains the outcome of the



(a) MCB test

(b) Evolution of δ_t

Figure 4: Top: MCB test on time-series with abrupt change in the coefficient vector. Bottom: Evolution of median and interquartile range for δ_t.

MCB test illustrated in Figure 4(a). PL-DLM is always the best performing method in this setting, while ADLM always outperforms Grid-DLM.

Overall, our results on simulated time-series suggest that there is not one method for estimating the discount factor that is always best. PL-DLM is designed to handle static environments which experience sudden large changes, and performs best in these conditions. The main advantages of Grid-DLM are its simplicity, and the very low performance variability. However, this method can perform poorly if the grid of discount factor values does not contain an appropriate δ for a given time-series, or when there is no fixed value of δ that is optimal for the entire time-series (as in the case of abrupt change). The reported results illustrate that ADLM adapts the discount factor towards appropriate values under different types of variation in the DGP. This allows ADLM to be outperform Grid-DLM in all cases and PL-DLM in the case of gradual change. Finally we note that the computational cost of PL-DLM is in a different order of magnitude compared to that of Grid-DLM and ADLM. Specifically to process the 100 time-series of length 1000 for each simulation PL-DLM with 1000 particles required on average 4.7 hours while the average time for Grid-DLM and ADLM was 34 seconds (a relative speed-up of more than 513 times).

4. Forecasting UK house price inflation through DLMs

Our empirical application focuses on the prediction of quarterly seasonally adjusted house price indices. The data on house price indices has been provided by Nationwide (the largest building society in the world and one of the largest mortgage providers in the UK) and covers the period from 1982:Q1 to 2017:Q4. We consider the UK national market, as well as each of the thirteen regional UK markets as defined by Nationwide's classification. The regional markets are: East Anglia (EA), East Midlands (EM), Greater London (GL), Northern Ireland (NI), North (NT), North West (NW), Outer Metropolitan (OM), Outer South East (OSE), Scotland (SC), South West (SW), Wales (WW), West Midlands (WM), and Yorkshire and Humberside (YH).

4.1. Data

To transform nominal into real prices, we divide by the consumer price index (all items), obtained from the OECD Database of Main Economic Indicators, and then compute the annualised log transformation of real property price inflation as,

$$y_{r,t} = 400 \times \ln\left(\frac{P_{r,t}}{P_{r,t-1}}\right), \quad r = 1, \ldots, 14, \tag{30}$$

where $P_{r,t}$ stands for the level of the real house price index of market r at time t. Time series plots of the annualised house price inflation index for all markets are provided in Appendix A.

We consider eight economic variables as potential predictors of future house price movements: three regional-level and five national-level predictors. The variables measured at the regional level include the price-to-income ratio (which proxies for affordability), income growth, and the unemployment rate. National-level predictors consist of the real mortgage rate, the number of housing starts, growth in real consumption, a new measure of house price uncertainty (HPU), which we construct using the news based methodology of Baker et al. (2016), and a credit conditions index (CCI) (Fernandez-Corugedo and Muellbauer, 2006). For GL, we also consider an additional financial variable, the spread between

yields on long-term and short-term government securities. Appendix B provides a detailed description of the data sources and all the variables.

From the above predictors, the first six have been used by Bork and Møller (2015) to forecast house price movements in US metropolitan states. CCI and HPU have not been previously employed in a forecasting context but as we argue next there are good reasons to be considered. Credit supply conditions in the UK economy, especially in the mortgage market, have changed dramatically since the 1970s. As argued by several authors, such changes were at the heart of the housing boom that preceded the Great Recession. It therefore seems natural to investigate whether an index of credit conditions may contain valuable information for forecasting.[1] With regard to HPU, there is a growing literature on the effect of uncertainty on macroeconomic outcomes. In this context, the impact of house price uncertainty on housing investment and real estate construction decisions has been recognised in a number of studies (Cunningham, 2006; Banks et al., 2015; Oh and Yoon, 2020).

In addition to the relationship of house prices with macroeconomic and financial variables, there is also a substantial empirical literature that documents the existence of strong spatial linkages between UK regional markets in sample (see, e.g., Drake, 1995; Meen, 1999; Cook and Thomas, 2003; Holly et al., 2010; Antonakakis et al., 2018, *inter alia*). To accommodate this, we incorporate in the set of potential predictors lagged property price growth in contiguous regions. The number of neighbouring regions for each of the thirteen real estate markets under consideration lies in the range of one to five.

We employ Hausman-type and Chow-type tests proposed by Chen and Hong (2012) to assess whether there is sufficient evidence of structural instability

[1] A deficiency of simple proxies for credit conditions, such as interest-rate spreads and unsecured credit to income ratios, is that they fail to control for the economic environment, and are thus subject to an endogeneity problem. The methodology of Fernandez-Corugedo and Muellbauer (2006) mitigates this problem by making use of a large number of economic and demographic controls.

between house price inflation and individual predictors. The results of both tests, reported in Appendix C, provide strong evidence of structural instability, and thus justify the use of dynamic econometric models.

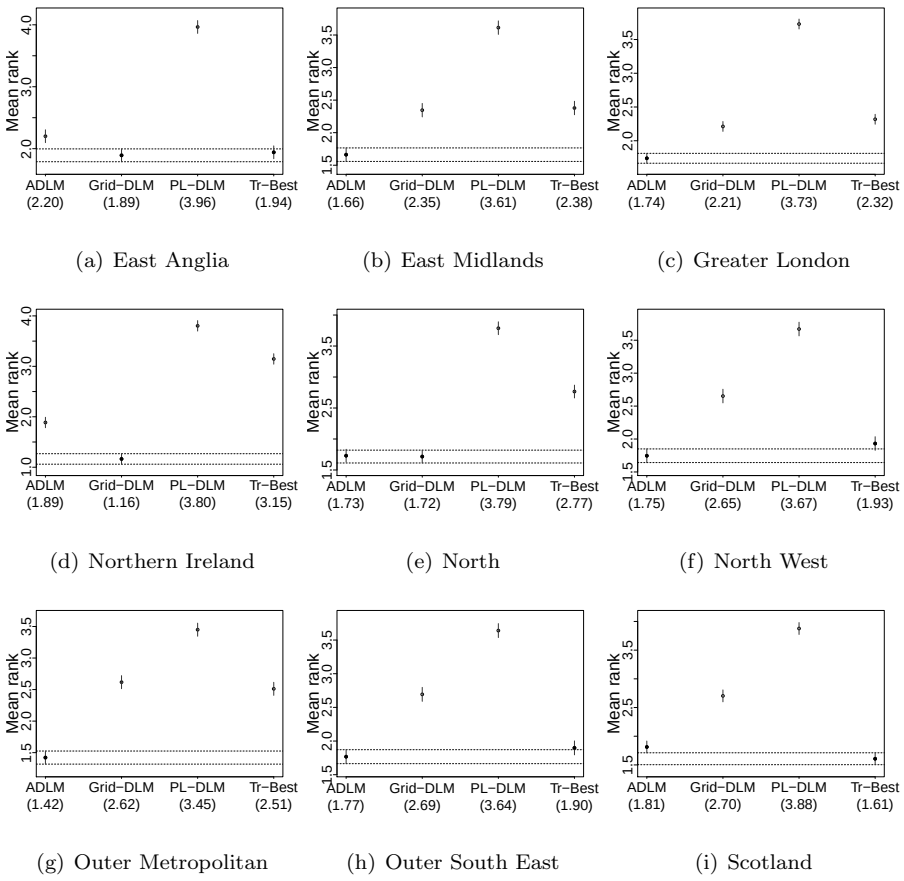
4.2. Empirical evaluation

In this section we evaluate the performance of ADLM, Grid-DLM, and PL-DLM on the task of predicting UK house price inflation. For each market we construct a model set (pool) which consists of a DLM for every possible combination of the predictors listed previously. The total number of models, K, is equal to $2^9 = 512$ for each regional market, except GL where due to the inclusion of the spread variable we have $2^{10} = 1024$ models. For the UK national market the number of models is $2^8 = 256$, since there are no contiguous regions at the national level.

In addition to ADLM, Grid-DLM, and PL-DLM we also consider a simple and intuitive approach to selecting a fixed discount factor. Specifically, for each model specification we estimate a separate D-DLM for each value of the discount factor, $\delta \in \{0.9, 0.91, \ldots, 0.99, 0.999\}$. At the end of the in-sample period we identify the value of δ that produced the lowest Mean Squared Forecast Error (MSFE) and use this to predict in the out-of-sample period. We abbreviate this method as Tr-Best.

The first evaluation criterion we consider is MSFE. To compare the performance of different methods across the different model specifications in each market we first use Friedman's nonparametric test (Friedman, 1937, 1939). In all markets the p-value for the null hypothesis of equal predictive performance across ADLM, Grid-DLM and PL-DLM, is extremely small (smaller than 10^{-9}). We are therefore justified to proceed with the MCB test (Koning et al., 2005). Figures 5 and 6 visualises the outcomes of this test for every market.

ADLM performs statistically significantly better than all other methods in the UK national market, and in six regional markets (namely EM, GL, OM, SW, WM, and WW). In NW and OSE ADLM achieves the lowest median rank but its performance is not statistically significantly better than that of Tr-Best. In

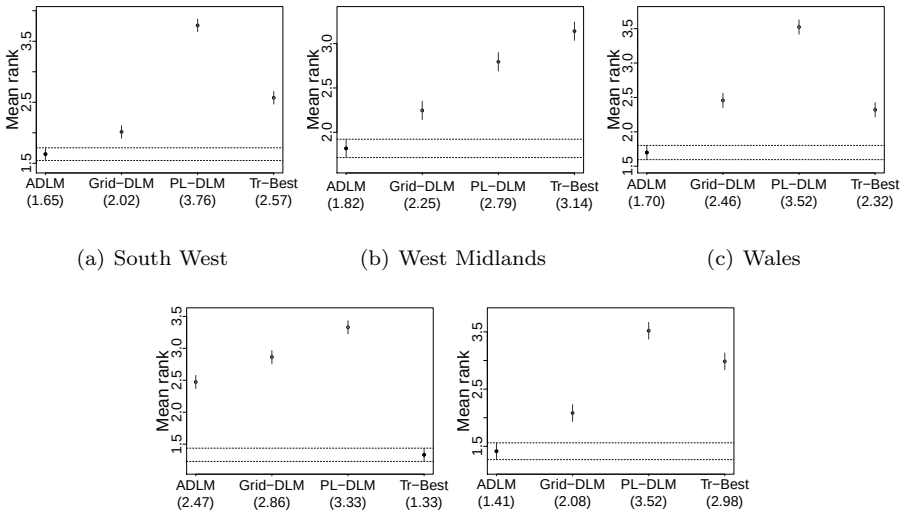


(a) East Anglia (b) East Midlands (c) Greater London

(d) Northern Ireland (e) North (f) North West

(g) Outer Metropolitan (h) Outer South East (i) Scotland

Figure 5: MCB tests to compare different methods to select the DLM discount factor. Each subfigure illustrates the outcome of the MCB test across all model specifications for a given housing market.



(a) South West (b) West Midlands (c) Wales

(d) Yorkshire & Humberside (e) UK national market

Figure 6: MCB tests to compare different methods to select the DLM discount factor. Each subfigure illustrates the outcome of the MCB test across all model specifications for a given housing market.

NT and SC the performance of ADLM is not statistically worse than that of the best performing method. Grid-DLM performs statistically significantly better than all other methods in NI, while Tr-Best outperforms all other methods in YH. In EA the best performing methods are Grid-DLM and Tr-Best.

In conclusion across the fourteen UK housing markets ADLM achieves a performance that is statistically significantly better than Tr-Best in nine markets, Grid-DLM in eleven markets, and PL-DLM in all fourteen markets. Note that Tr-Best is statistically significantly better than ADLM in one market (YH), while Grid-DLM is statistically significantly better than ADLM in two markets (EA and NI). This provides strong evidence that the proposed ADLM is effective at adapting the discount factor to improve forecast performance.

5. Forecast combination

In this section we investigate whether and to what extent forecast combination can improve the forecastability of UK house prices. Based on the results of the performance comparison in the previous section, here we only consider combining forecasts produced by ADLMs. In Sections 5.1 to 5.5 we briefly introduce the forecast combination methods we consider, and in Section 5.6 we discuss the findings of the empirical evaluation.

5.1. Averaging

Although averaging is the simplest approach to model combination, an extensive literature attests to the difficulty of outperforming it. The "equal weights puzzle" (Clemen, 1989; Diebold, 1989) has been extensively studied. In a contribution that is particularly relevant to our case Smith and Wallis (2009) show that the finite-sample error in the estimation of the combination weights can cause sophisticated forecast combination methods to perform poorly. Clearly the possibility and severity of "overfittting" the combination weights increases as the number of models, K, increases relative to the time series length, T. In our case this is an important consideration since the time series length is $T = 143$, while the number of models is $K \in \{256, 512, 1024\}$.

5.2. Bayesian model averaging and power discounting

In Bayesian model averaging (BMA) and model probability power discounting (MPD) the probability assigned to each model M_j is sequentially updated according to,

$$p(M_j|\mathcal{D}_t) \propto p(y_t|\mathcal{D}_{t-1}, M_j)\, p(M_j|\mathcal{D}_{t-1})^{\alpha},$$

where $\alpha \in (0, 1]$ is the power discount factor (West and Harrison, 1989). Setting $\alpha = 1$ yields BMA, while in MPD $\alpha < 1$ to allow $p(M_j|\mathcal{D}_t)$ to weigh more heavily recent predictive performance. As mentioned in Section 2.1 the motivation for MPD is to prevent situations in which models that predict well at later time steps are assigned a weight that is very small (or effectively zero) in early stages. In the implementation of MPD in this section we use a grid of values, $\alpha \in [0.95, 1]$ with a step size of 0.005 (Zhao et al., 2016). Note that the more recent dynamic model averaging (Raftery et al., 2010; Koop and Korobilis, 2012) combines a pool of DLMs with constant discount factors, and power discounting with a single value for α (typically $\alpha \in \{0.99, 0.95\}$).

5.3. Prediction with expert advice

The problem of incrementally updating the weights assigned to individual models so as to minimise a measure of forecast error has been extensively studied in the field of machine learning known as prediction with expert advice (PEA) (Cesa-Bianchi and Lugosi, 2006). PEA studies the following online learning problem: At time-step t each of the K forecasters (experts) provides a forecast, $\hat{y}_{k,t}$, and the forecast of the aggregation algorithm is a convex combination of these, $\hat{y}_t = \sum_{k=1}^{K} w_{k,t}\, \hat{y}_{k,t}$. Note that no assumptions are made concerning how the individual $\hat{y}_{k,t}$ are generated. After observing y_t each expert incurs a loss $\ell(y_t, \hat{y}_{k,t})$, which depends on the forecast error, and the weight of each expert, $w_{k,t}$, is updated based on these losses. The analysis of PEA algorithms in dynamic environments considers a partition of the time series into an arbitrary number of M segments, $t_{(0)} = 1 < t_{(1)} < t_{(2)} < \cdots < t_{(M-1)} < T = t_{(M)}$. Given this partition and the forecasts of the experts the minimum cumulative

loss over the length of the time series is,

$$L_T^{\star}\left(t_{(1)}, t_{(2)}, \ldots, t_{(M-1)}\right) = \sum_{m=1}^{M} \min_{k} \left\{ \sum_{t=t_{(m-1)}}^{t_{(m)}} \ell(y_t, \hat{y}_{k,t}) \right\}.$$

The optimal partition of the time series into M segments minimises the above loss with respect to the location of the change points,

$$L_T^{\star}(M) = \min_{1<t_{(1)}<\cdots<t_{(M-1)}<T} L_T^{\star}\left(t_{(1)}, t_{(2)}, \ldots, t_{(M-1)}\right).$$

The literature on PEA provides aggregating algorithms whose cumulative loss,

$$L_T = \sum_{t=1}^{T} \ell(y_t, \hat{y}_t),$$

is within a known upper bound with respect to $L_T^{\star}(M)$, for any M and T. It is important to note that these upper bounds are functions of T, K, and M. The dependence on the latter is inevitable because as M increases the problem becomes more difficult (and in the limiting case $M = T$ no meaningful bound is possible). Starting with Herbster and Warmuth (1998) a number of PEA algorithms have been proposed that achieve optimal upper bounds for this problem, but most work assumes that the loss is bounded (typically the range of the loss function is assumed to be $[0, 1]$). In most forecasting applications this assumption is not tenable. The first (and to the best of our knowledge only) PEA algorithm that can handle unbounded losses and dynamic environments is ConfHedge (CH) (V'yugin and Trunov, 2019).

BMA, MPD and CH are designed to identify the "true" (or best) model contained in the model set. If the model set is incomplete, that is if it does not include the "true" model, then BMA and MDP will assign all the probability mass to the model with the smallest Kullback-Leibler divergence to the true DGP. Similarly, CH will assign a weight of one to the model with the smallest loss. This will be the case even if there exists a convex combination of forecasts which achieves better forecast accuracy. The remaining three approaches we consider overcome this limitation in different ways.

5.4. Partially egalitarian LASSO

The partially egalitarian LASSO (peLASSO) (Diebold and Shin, 2019) follows a two-step procedure to combine forecasts. First, individual model forecasts are used as predictors in a LASSO penalised linear regression (Tibshirani, 1996). The resulting regression however is not used for forecasting, but only to identify a subset of relevant models. The final forecast is obtained by averaging the predictions of the models selected by LASSO. Through this two step procedure peLASSO attempts to combine the benefits of reducing the size of the model set, and avoid overfitting in the estimation of the combination weights. The performance of peLASSO depends on the choice of the LASSO penalty parameter, λ.[2] We select λ through 5-fold cross validation. To accommodate time-variation in the optimal model combination peLASSO estimates uses a window of most recent observations. In all the reported results we use a window length of 40 which is equal to the length of the in-sample period.

5.5. Density combination

Density combination methods aim to optimally combine the information in the probability density functions of individual models rather than just the point forecasts. The optimal prediction pool (OPP), proposed by Hall and Mitchell (2007) and elaborated by Geweke and Amisano (2011), minimises the Kullback-Leibler divergence between the combined density and the unknown, true probability density function. This is equivalent to maximising the log predictive score function, with respect to the combination weights, which results in a convex optimisation problem.

We also consider a Bayesian combination approach called Density Combination (DeCo) (Billio et al., 2013). DeCo relies on a state-space formulation to model the time variation in the true, but unobserved combination weights. In

[2] Diebold and Shin (2019) suggest an approach that avoids LASSO (and hence the determination of λ) through an exhaustive search over all possible subsets of a certain size. In our case this is infeasible since it would involve estimating billions of averages at every time step.

particular the true combination weights are obtained through a logistic transformation of latent variables which evolve according to a linear and Gaussian process. The resulting state-space model is therefore nonlinear and hence filtering is performed through sequential Monte Carlo (particle filtering). In DeCo the predictive density of each individual model is represented as a finite sample from the corresponding distribution. Therefore the two critical parameters for this algorithm are the number of particles in the particle filter, and the number of points with which the predictive distribution of each model is represented at each time-step. We set both to 1000 which is the default setting in the `DeCo R` package (Casarin et al., 2015).

5.6. Performance evaluation

Figures 7 and 8 present histograms of the MSFE of individual models in the out-of-sample period. In the same figures the MSFE of each forecast combination method is illustrated with a vertical line and a dot below the horizontal axis. These figures enable us to compare the performance of each combination method relative to the distribution of the performance by the individual ADLMs which comprise the pool of models. For completeness we also indicate the out-of-sample performance of the ADLM which achieved the lowest MSFE in the in-sample period, which we abbreviate as Tr-Best. We also illustrate the out-of-sample MSFE of three standard time-series forecasting methodologies: AR(1), the established benchmark in house price forecasting (Bork and Møller, 2015), ARIMA and exponential smoothing (ETS). For ARIMA and ETS we used automatic model selection through the `forecast R` package (Hyndman and Khandakar, 2008). Table 2 reports the precise MSFE values across housing markets for all the considered forecasting methods. In this table the best performing method is indicated with bold, and the second best performing method with *italic*.

DeCo achieves the best performance in all markets, and is the only combination method that consistently outperforms the best individual model in the out-of-sample period. In eleven out of the fourteen housing markets the simple

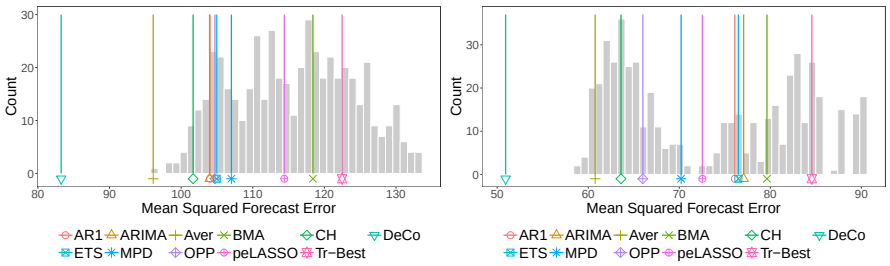
average is the second best performing method. This is not surprising since the number of models is very large compared to the time-series length. In six markets (EA, NI, NW, OSE, SW, and WM) the simple average achieves an MSFE that is lower than that of the best individual ADLM. Out of the three methods that sequentially update the combination weights CH is the best performing, while BMA is frequently the worst combination method. The performance of peLASSO exhibits high variability. In most markets OPP is ranked in the top four best performing methods. Tr-Best performs very poorly in all housing markets except in WW. In fact in all markets except WW the MSFE of Tr-Best is on the right tail of the empirical distribution of MSFE. With a few exceptions ARIMA, ETS and AR(1) are among the worst performing methods.

The above findings are confirmed by an MCB test that compares the out-of-sample performance of the different methods across the fourteen housing markets. Figure 9 presents the outcome of this test. The performance of DeCo is statistically significantly better than all other methods except the average and ConfHedge. The average achieves the second lowest median rank, and its performance is statistically significantly better than peLASSO, BMA, AR(1), ARIMA, and ETS.

Next we evaluate density forecasting performance and consider the log predictive density ratio (LPDR) on the out-of-sample period. Based on the previous analysis the natural benchmark for this measure is DeCo, but DeCo does not provide a density score. Therefore we use as benchmark the simple average. Table 2 reports the LPDR computed as,

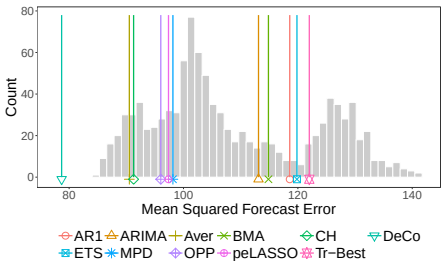
$$\text{LPDR} = \sum_{t=T_{\text{train}}}^{T-1} \log\left(p_f(y_{t+1}|\mathcal{D}_t)/p_{\text{aver}}(y_{t+1}|\mathcal{D}_t)\right),$$

where $p_f(y_{t+1}|\mathcal{D}_t)$ is the predictive density of method f, and $p_{\text{aver}}(y_{t+1}|\mathcal{D}_t)$ is the predictive density of the simple average. A negative LPDR indicates that method f performs worse than the average in terms of density forecasting over the out-of-sample period. The opposite is true when LPDR is positive. Table 2 reveals that in terms of LPDR the average is the best performing method in four markets: EA, NI, SC and the national UK market. In all other markets there

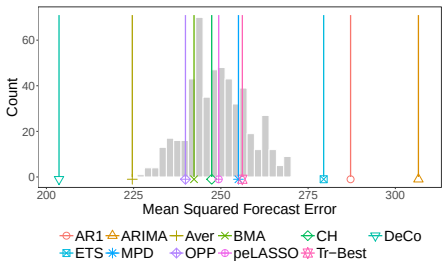


(a) East Anglia

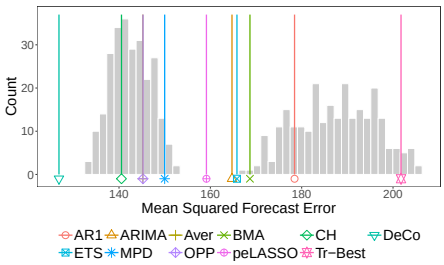
(b) East Midlands



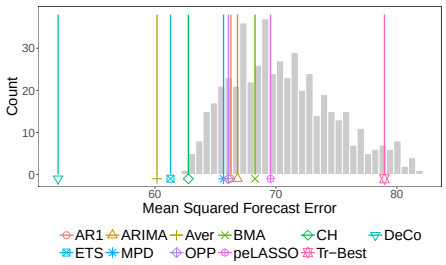
(c) Greater London



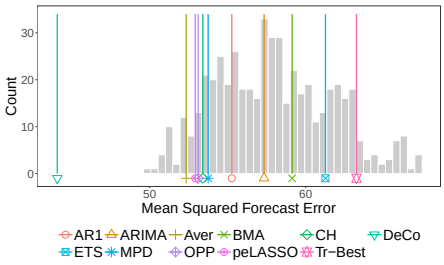
(d) Northern Ireland



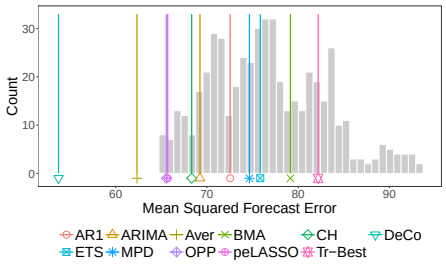
(e) North



(f) North West

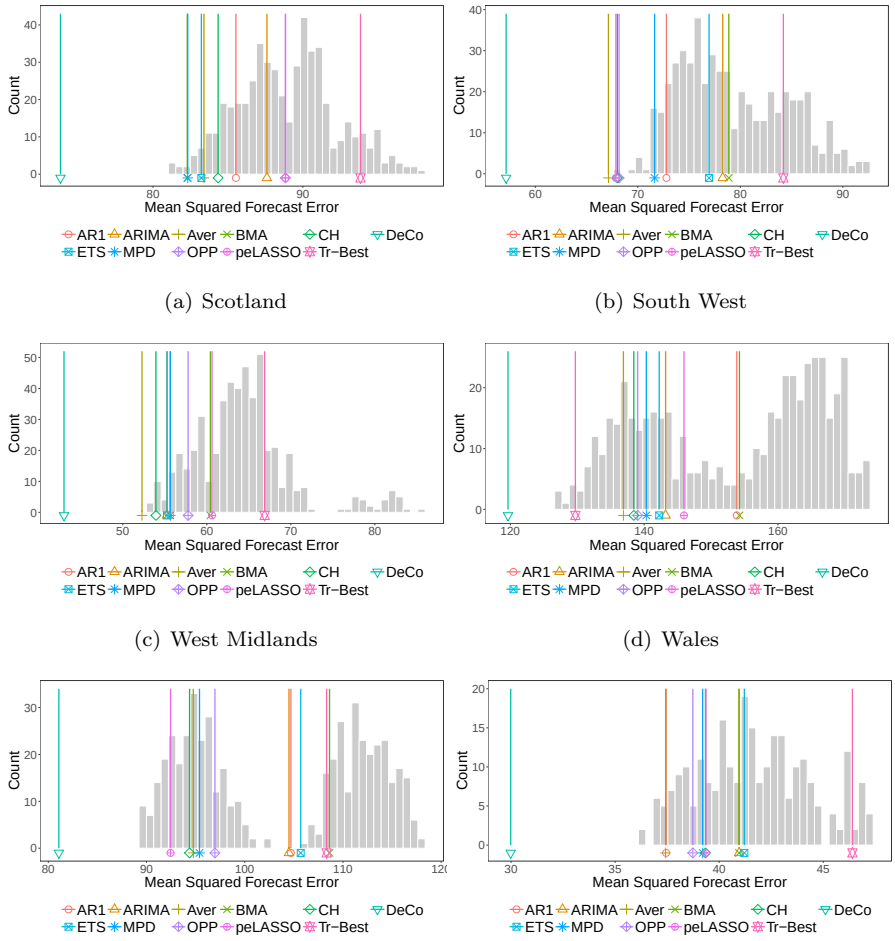


(g) Outer Metropolitan



(h) Outer South East

Figure 7: Histograms of out-of-sample mean squared forecast error (MSFE) of individual ADLMs in different housing markets. The MSFE of each forecast combination method is represented with a dot and a vertical line.



(a) Scotland

(b) South West

(c) West Midlands

(d) Wales

(e) Yorkshire & Humberside

(f) UK national market

Figure 8: Histograms of out-of-sample mean squared forecast error (MSFE) of individual ADLMs in different housing markets. The MSFE of each forecast combination method is represented with a dot and a vertical line.

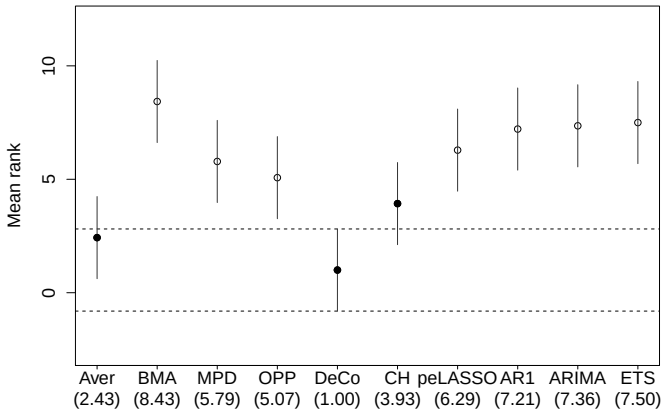


Figure 9: Multiple comparison with the best (MCB) test that compares different model combination methods across the fourteen housing markets.

Region	DeCo	Aver	BMA	MPD	OPP	CH	peLASSO	AR(1)	ARIMA	ETS
EA	**83.24**	*96.1*	118.39	107.01	104.72	101.67	114.4	104.02	103.97	104.97
EM	**50.93**	*60.76*	79.63	70.2	66	63.6	72.54	76.1	77.07	76.49
GL	**78.69**	*90.53*	114.85	98.15	96.03	91.28	97.38	118.58	113.12	119.84
NI	**203.62**	*224.61*	242.31	255.04	239.85	247.37	249.38	287.17	306.62	279.46
NT	**126.85**	145.22	168.67	150	145.25	*140.54*	159.14	178.44	164.74	165.82
NW	**51.99**	*60.16*	68.27	65.67	66.07	62.76	69.55	66.27	66.82	61.28
OM	**44.07**	*52.33*	59.13	53.74	53.1	53.4	52.92	55.26	57.33	61.26
OSE	**53.72**	*62.32*	79.15	74.65	65.54	68.32	65.69	72.54	69.23	75.84
SC	**73.82**	83.4	*82.29*	82.32	88.85	84.35	88.84	85.53	87.6	83.23
SW	**57.15**	*67.13*	78.87	71.63	68.17	68	67.86	72.79	78.28	76.95
WM	**43**	*52.28*	60.42	55.66	57.77	53.94	60.62	55.64	55.25	55.26
WW	**119.65**	*136.9*	154.24	140.32	139.04	138.45	145.95	153.85	143.21	142.25
YH	**81.07**	94.76	108.63	95.39	96.96	94.38	*92.44*	104.7	104.5	105.7
UK	**29.98**	*37.43*	40.95	39.21	38.73	39.34	39.37	37.45	40.96	41.21

Table 1: Mean squared forecast error (MSFE) on the out-of-sample period for different forecast combination methods as well as three benchmark time-series forecasting methods.

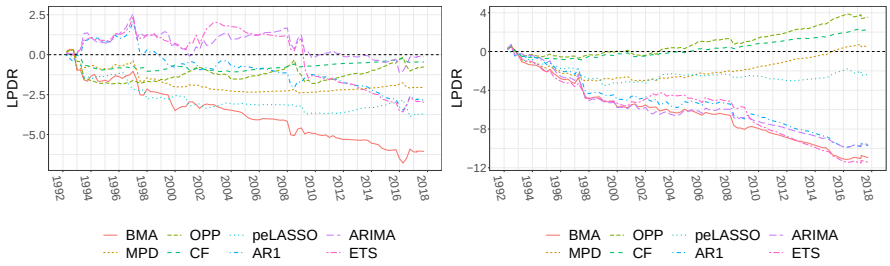
exists at least one forecast combination method which achieves a positive LPDR. The method whose performance clearly stands out is OPP, which achieves a positive LPDR in ten markets, and outperforms all other methods (including the simple average) in eight markets. The other two methods that perform notably well in terms of LPDR are CH and peLASSO. It is worth noting that all simple time-series forecasting methodologies achieve a negative LPDR in all markets. Therefore in UK housing markets averaging ADLM models always produces better density forecasts compared to AR(1), ETS, and ARIMA.

However the LPDR at the end of the out-of-sample period does not capture the entire picture. Figures 10 and 11 illustrate the evolution of LPDR for each housing market. In most markets (namely EA, NI, OM, OSE, SC, SW, WM, WW, YH, and UK) we observe a sudden decrease in the LPDR of AR(1), ETS, and ARIMA which assume stationarity. The most notable example of this is the NI market, where the magnitude of the decrease in LPDR for AR(1), ETS and ARIMA is much larger than that observed in any other market. In all markets except YH, the abrupt decline occurs in 2008 suggesting a connection to the onset of the financial crisis. It can also be observed that in most markets (EA, NT, OM, OSE, SC, SW, WM, YH, and the UK national market), at least one of the standard forecasting methods has a positive LPDR early in the out-of-sample period. In contrast, in most markets OPP improves its relative performance later in the out-of-sample period.

In conclusion, the empirical evaluation of the seven forecast combination methods considered suggests that in terms of MSFE DeCo is the only method that is capable of consistently outperforming the best individual model in the pool, and the simple average in the out-of-sample period. This comes at a very high computational cost however. DeCo required 7.84, 9.68 and 34.12 hours to combine 256, 512, and 1024 models, respectively. DeCo does not provide an analytic predictive distribution and hence we use the average as a baseline to evaluate LPDR. Our results suggest that OPP is the best performing method in terms of density forecasting. Out of BMA, MPD and Confhedge which sequentially update model probabilities in less than a second our results

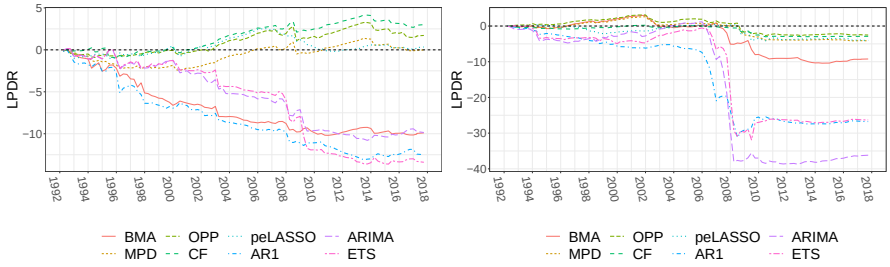
	BMA	MPD	OPP	CH	peLASSO	AR1	ARIMA	ETS
EA	-6.05	-2.05	-0.77	*-0.46*	-3.71	-2.84	**-0.13**	-2.97
EM	-10.95	0.59	**3.54**	*2.23*	-2.38	-9.72	-9.59	-11.43
GL	-9.9	-0.02	*1.72*	**3.01**	0.34	-12.5	-9.85	-13.4
NI	-9.23	-4.14	**-2.52**	*-2.88*	-4.03	-26.77	-36.18	-26.31
NT	0.88	3.89	**3.99**	*3.93*	-3.33	-10.08	-4.93	-6.09
NW	-5.19	-1.13	**0.77**	*0.05*	-3.83	-4.5	-2.5	-1.92
OM	-4.6	-0.27	**0.83**	0.02	*0.58*	-2.53	-3.72	-7.59
OSE	-8.95	-4.32	**0.92**	-1.62	*-0.22*	-6.07	-2.71	-8.3
SC	-1.05	**-0.08**	-2.68	*-0.53*	-2.36	-4.07	-4.09	-2.69
SW	-5.65	-0.37	*1.83*	-0.07	**1.86**	-3.72	-5.55	-6.7
WM	-4.12	*-0.2*	**0.21**	-0.99	-2.97	-1.25	-0.54	-2.98
WW	-5.45	0.13	**0.72**	0.54	-3.44	-8.57	-3.69	-4.88
YH	-5.82	*2.14*	**3.72**	1.33	1.66	-5.16	-4.18	-7.14
UK	-3.12	-1.1	**-0.34**	-1.28	*-0.92*	-1.04	-4.33	-5.66

Table 2: Log predictive density ratio (LPDR) on the out-of-sample period for different forecast combination methods as well as three benchmark time-series forecasting methods.



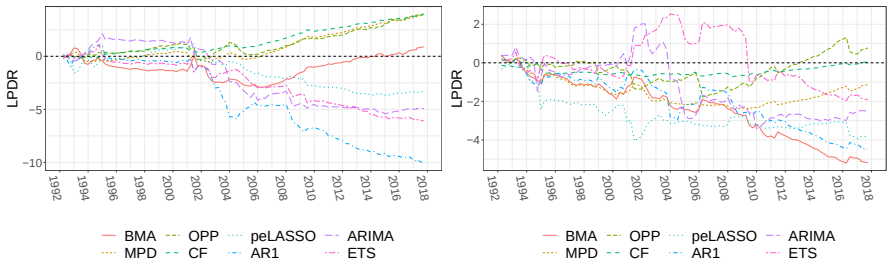
(a) East Anglia

(b) East Midlands



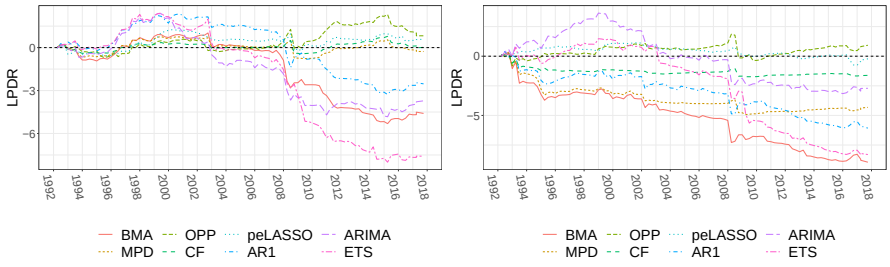
(c) Greater London

(d) Northern Ireland



(e) North

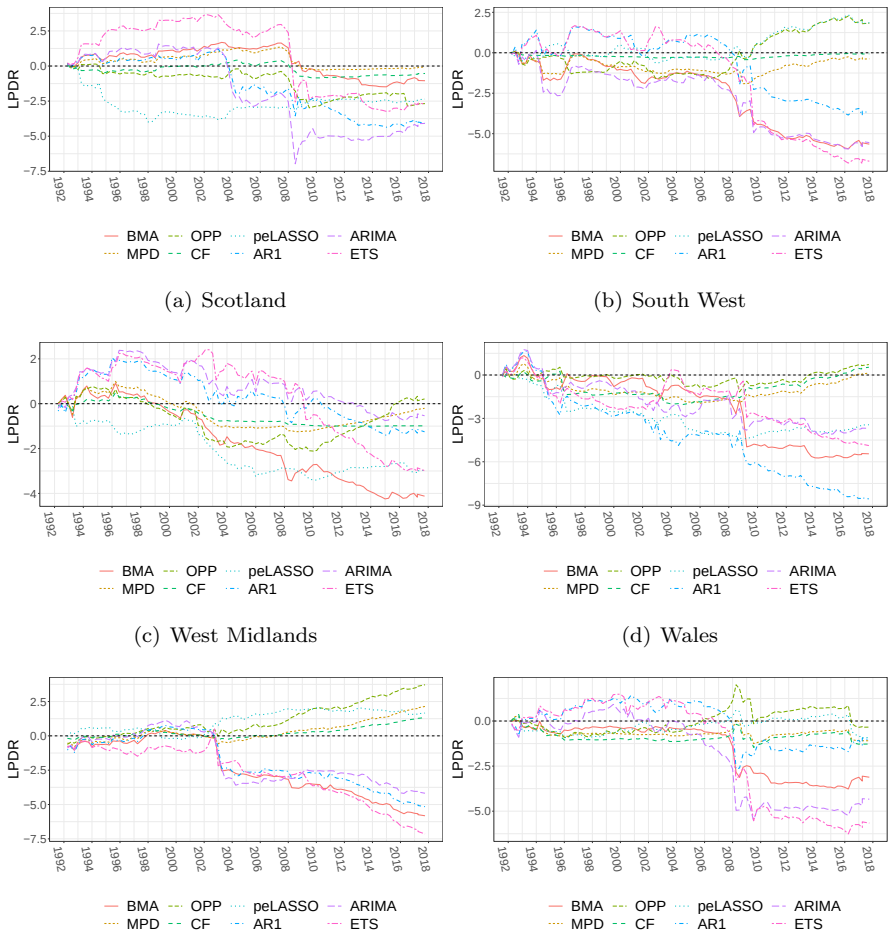
(f) North West



(g) Outer Metropolitan

(h) Outer South East

Figure 10: Evolution of likelihood predictive density ration (LPDR) in the out-of-sample period for different model combination methods and benchmark time-series forecasting methodologies.



(a) Scotland (b) South West

(c) West Midlands (d) Wales

(e) Yorkshire & Humberside (f) UK national market

Figure 11: Evolution of likelihood predictive density ration (LPDR) in the out-of-sample period for different model combination methods and benchmark time-series forecasting methodologies.

suggest that CH performs best, while BMA performs worse. Finally, combining the forecasts of ADLMs consistently outperforms standard time series methods like ETS and ARIMA, and the AR(1) model which is the established benchmark in forecasting housing markets (Bork and Møller, 2015; Rapach and Strauss, 2009).

6. Conclusions

In this paper we proposed an adaptive approach to estimate the discount factor in dynamic linear models. In our approach the optimal discount factor minimises the expected one-step-ahead squared forecast error. This formulation enables us to apply stochastic gradient descent to sequentially estimate the optimal value of this parameter. The proposed approach allows for closed-form updating and does not increase the computational complexity of filtering. We conducted the empirical evaluation of the proposed approach on the task of forecasting UK house prices, at the national and regional level. The results of this study suggest that the adaptive dynamic linear model can achieve significant forecast accuracy improvements over competing methods.

We also considered and assessed a range of forecast combination methods to predict UK house prices. In agreement with a large empirical literature we find that it is difficult to outperform the simple average in terms of mean squared forecast error. Our findings however show that a recently proposed density combination method which allows for time-varying combination weights consistently outperforms the simple average. In terms of density forecasting the optimal prediction pool consistently outperforms the simple average.

Acknowledgements

We want to thank the associate editor and two anonymous reviewers for providing us with constructive comments which enabled us to substantially improve this work.

References

Aizenman, J., Jinjarak, Y., 2014. Real estate valuation, current account and credit growth patterns, before and after the 2008–9 crisis. Journal of International Money and Finance 48 (Part B), 249–270.

Antonakakis, N., Chatziantoniou, I., Floros, C., Gabauer, D., 2018. The dynamic connectedness of UK regional property returns. Urban Studies 55 (14), 3110–3134.

Anundsen, A. K., 2015. Econometric regime shifts and the US subprime bubble. Journal of Applied Econometrics 30 (1), 145–169.

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. The Quarterly Journal of Economics 131 (4), 1593–1636.

Banks, J., Blundell, R., Oldfield, Z., Smith, J. P., 2015. House price volatility and the housing ladder. Working Paper 21255, National Bureau of Economic Research.

Benveniste, A., Priouret, P., Métivier, M., 1990. Adaptive algorithms and stochastic approximations. Springer-Verlag New York, Inc. New York, NY, USA.

Billio, M., Casarin, R., Ravazzolo, F., van Dijk, H. K., 2013. Time-varying combinations of predictive densities using nonlinear filtering. Journal of Econometrics 177 (2), 213–232.

Bork, L., Møller, S. V., 2015. Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. International Journal of Forecasting 31 (1), 63–78.

Bottou, L., Curtis, F. E., Nocedal, J., 2018. Optimization methods for large-scale machine learning. SIAM Review 60 (2), 223–311.

Byrne, J. P., Korobilis, D., Ribeiro, P. J., 2018. On the source of uncertainty in exchange rate predictability. International Economic Review 59 (1), 329–357.

Carvalho, C. M., Johannes, M. S., Lopes, H. F., Polson, N. G., 2010. Particle Learning and Smoothing. Statistical Science 25 (1), 88–106.

Casarin, R., Grassi, S., Ravazzolo, F., van Dijk, H. K., 2015. Parallel sequential monte carlo for efficient density combination: The DeCo MATLAB toolbox. Journal of Statistical Software 68 (13), 1–30.

Catania, L., Nonejad, N., 2018. Dynamic model averaging for practitioners in economics and finance: The eDMA package. Journal of Statistical Software 84 (11), 1–39.

Cesa-Bianchi, N., Lugosi, G., 2006. Prediction Learning and Games. Cambridge University Press.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80 (3), 1157–1183.

Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting 5 (4), 559–583.

Cook, S., Thomas, C., 2003. An alternative approach to examining the ripple effect in UK house prices. Applied Economics Letters 10 (13), 849–851.

Cunningham, C. R., 2006. House price uncertainty, timing of development, and vacant land prices: Evidence for real options in seattle. Journal of Urban Economics 59 (1), 1–31.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106 (1), 157–181.

Diebold, F. X., 1989. Forecast combination and encompassing: Reconciling two divergent literatures. International Journal of Forecasting 5 (4), 589–592.

Diebold, F. X., Shin, M., 2019. Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. International Journal of Forecasting 35 (4), 1679–1691.

Drake, L., 1995. Testing for convergence between UK regional house prices. Regional Studies 29 (4), 357–366.

Fernandez-Corugedo, E., Muellbauer, J., 2006. Consumer credit conditions in the United Kingdom. Bank of England working papers 314, Bank of England.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32 (200), 675–701.

Friedman, M., 1939. A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 34 (205), 109–109.

Geweke, J., Amisano, G., 2011. Optimal prediction pools. Journal of Econometrics 164 (1), 130–141.

Ghysels, E., Plazzi, A., Valkanov, R., Torous, W., 2013. Forecasting real estate prices. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2. Elsevier, pp. 509 – 580.

Hall, S. G., Mitchell, J., 2007. Combining density forecasts. International Journal of Forecasting 23, 1–13.

Herbster, M., Warmuth, M. K., 1998. Tracking the best expert. Machine Learning 32 (2), 151–178.

Holly, S., Pesaran, M. H., Yamagata, T., 2010. Spatial and temporal diffusion of house prices in the UK. IZA Discussion Papers 4694, Institute of Labor Economics (IZA).

Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 (3), 1–22.

Irie, K., Glynn, C., Aktekin, T., 2022. Sequential modeling, monitoring and forecasting of streaming web traffic data. Annals of Applied Statistics 16 (1), 300–325.

Kingma, D. P., Ba, J. L., 2015. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR).

Koning, A. J., Franses, P. H., Hibon, M., Stekler, H. O., 2005. The M3 competition: Statistical tests of the results. International Journal of Forecasting 21 (3), 397–409.

Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review 53 (3), 867–886.

Koop, G., Tole, L., 2013. Forecasting the European carbon market. Journal of the Royal Statistical Society: Series A 176 (3), 723–741.

Meen, G., 1999. Regional house prices and the ripple effect: A new interpretation. Housing Studies 14 (6), 733–753.

Muellbauer, J., Cameron, G., 2006. Was there a British house price bubble? Evidence from a regional panel. Economics Series Working Papers 276, University of Oxford, Department of Economics.

Office for National Statistics, 2018. UK national accounts, the blue book: 2018. Tech. rep., Office for National Statistics.

Oh, H., Yoon, C., 2020. Time to build and the real-options channel of residential investment. Journal of Financial Economics 135 (1), 255–269.

Paul, P., 2020. The time-varying effect of monetary policy on asset prices. Review of Economics and Statistics 102 (4), 690–704.

Prado, R., West, M., 2010. Time Series: Modeling, Computation, and Inference. CRC Texts in Statistical Science. Chapman & Hall.

Raftery, A. E., Kárný, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52 (1), 52–66.

Rapach, D. E., Strauss, J. K., 2009. Differences in housing price forecastability across US states. International Journal of Forecasting 25 (2), 351–372.

Rossi, B., 2013. Advances in forecasting under instability. In: Handbook of economic forecasting. Vol. 2. Elsevier, pp. 1203–1324.

Smith, J., Wallis, K. F., 2009. A simple explanation of the forecast combination puzzle*. Oxford Bulletin of Economics and Statistics 71 (3), 331–355.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58 (1), 267–288.

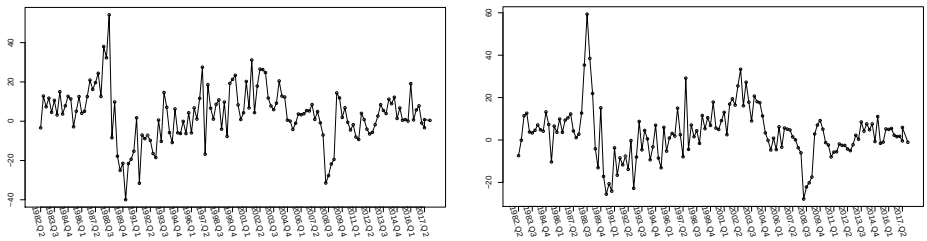
V'yugin, V., Trunov, V., 2019. Online aggregation of unbounded losses using shifting experts with confidence. Machine Learning 108 (3), 425–444.

West, M., Harrison, P. J., 1989. Bayesian Forecasting and Dynamic Models, 1st Edition. Springer: New York.

West, M., Harrison, P. J., 1997. Bayesian Forecasting and Dynamic Models, 2nd Edition. Springer: New York.

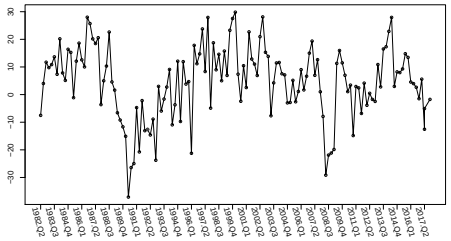
Zhao, Z. Y., Xie, M., West, M., 2016. Dynamic dependence networks: Financial time series forecasting and portfolio decisions. Applied Stochastic Models in Business and Industry 32 (3), 311–332.

Appendix A Time series of annualised real house price inflation

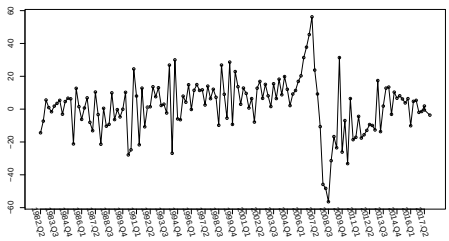


(a) East Anglia

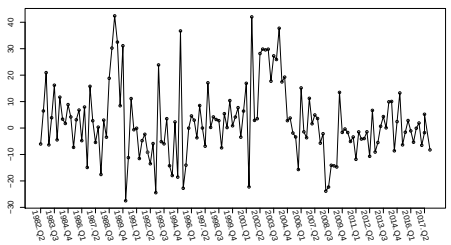
(b) East Midlands



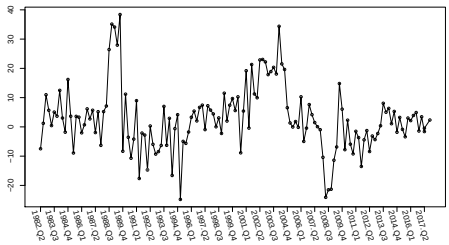
(c) Greater London



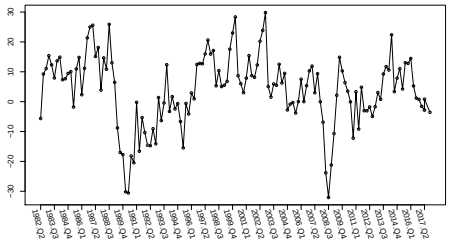
(d) Northern Ireland



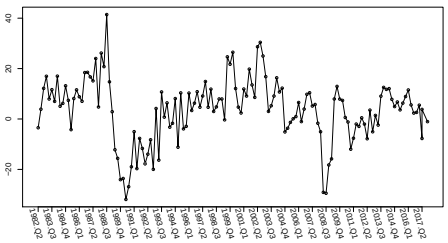
(e) North



(f) North West

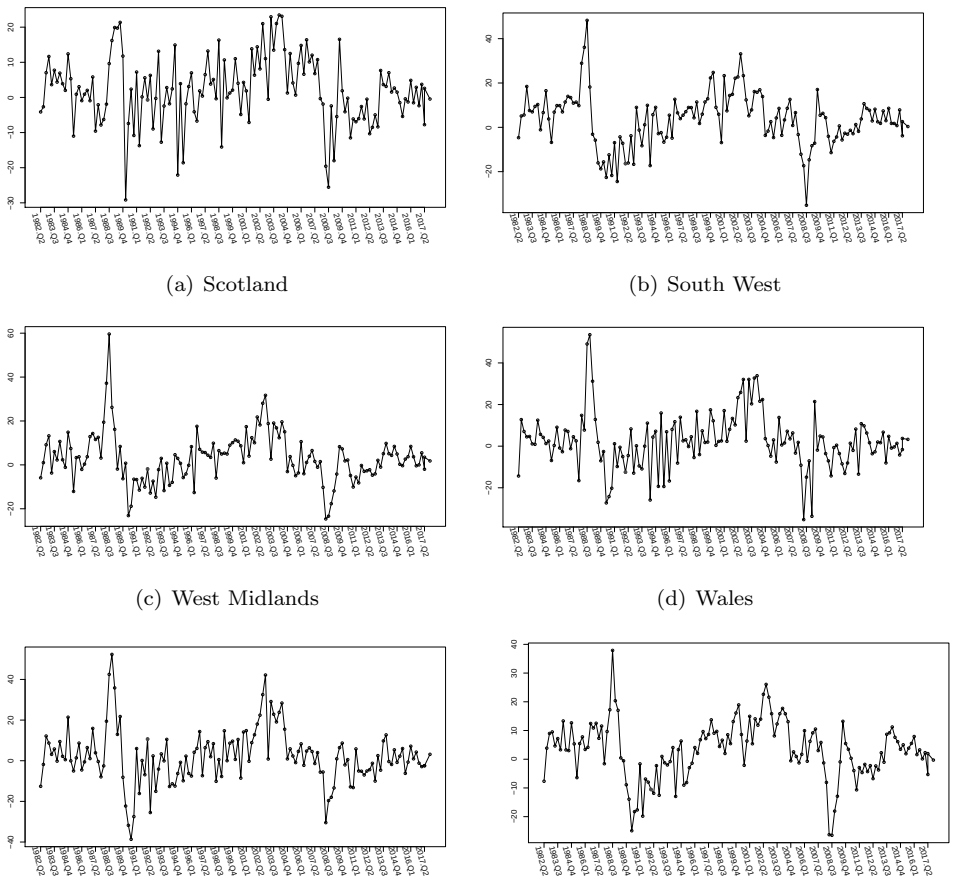


(g) Outer Metropolitan



(h) Outer South East

Figure 12: Time series plots of annualised real house price inflation in eight regional UK housing markets.



(a) Scotland

(b) South West

(c) West Midlands

(d) Wales

(e) Yorkshire and Humberside

(f) UK national market

Figure 13: Time series plots of annualised real house price inflation in five regional, as well as the national UK housing market.

Appendix B Variable Definitions and Data Sources

House Prices Real regional house price index (all houses, seasonally adjusted). Source: Nationwide.

Income Real average total household's weekly expenditure. Source: Family Expenditure Survey (FES).

Price-to-Income Ratio Quarterly changes in the log of the ratio of house prices to income.

Income Growth Annualised quarterly changes in the log of real income.

Unemployment Rate Quarterly changes in the ratio of unemployed people to the labour force times 100. Source: LFS.

Real Mortgage Rate Quarterly changes in the real mortgage rate of building societies, adjusted for the cost of mortgage tax relief as in Muellbauer and Cameron (2006). Sources: OECD Main Economic Indicators and HM Revenue & Customs.

Spread Difference between the 10-year government bond yield and the rate of discount on 3-month treasury bills. Sources: Saint Louis FRED Economic Data and the Bank of England.

Real Consumption Growth Annualised quarterly changes in the log of real final consumption expenditure of households and non-profit institutions serving households (seasonally adjusted, millions of UK sterling pounds). Source: ONS.

Housing Starts Log of the number of all permanent dwellings started in the UK. Source: Department for Communities and Local Government.

Index of Credit Conditions Designed as a linear spline function, this index is estimated using a two-equation system of secured and unsecured lending. For details about the methodology and sources of the data used in the estimation please refer to the supplementary Appendix to Yusupova et al., (2019).

House Price Uncertainty Index Constructed using the methodology outlined in Baker et al. (2016) to proxy for economic policy uncertainty. The HPU is an index of search results from five large newspapers in the UK: The Guardian, The Independent, The Times, Financial Times and Daily Mail. We

use LexisNexis digital archives of these newspapers to obtain a quarterly count of articles that contain the following three terms: 'uncertainty' or 'uncertain'; 'housing' or 'house prices' or 'real estate'; and one of the following: 'policy', 'regulation', 'Bank of England', 'mortgage', 'interest rate', 'stamp-duty', 'tax', 'bubble' or 'buy-to-let' (including variants like 'uncertainties', 'housing market' or 'regulatory'). To meet the search criteria an article must contain terms in all three categories. The resulting series of search counts is then scaled by the total number of articles in the given newspaper and in the given quarter. Finally, to obtain the HPU index, we average across the five newspapers by quarter and normalise the index to a mean of 100.

Appendix C Structural Instability in UK House Prices

For completeness, in this section we examine whether there is evidence of structural instability in the relationship between real house price inflation and individual house price predictors in-sample. To do so, we employ two tests proposed by Chen and Hong (2012). The first is a Hausman-type (H) test that compares time-varying parameter estimates obtained by local linear regression to constant estimates obtain by ordinary least squares. The second is a Chow-type (C) test which compares the sum of squared residuals between the constant parameter and local linear regression models. The null hypothesis in both tests is that of time-invariant regression coefficients.

These tests correspond well to the ADLM of Section 2 because they impose minimal restrictions on the functional form of the time-varying parameters and, thus, are consistent with both smooth and abrupt structural change. Furthermore, they require no prior information regarding the timing and the number of breaks; they are asymptotically pivotal, and they do not involve trimming of the boundary region near the end points of the sample period.

Tables 3 and 4 report wild-bootstrap p-values of the H and C tests for each of the 11 house price predictors considered and for each of the 13 regions. We observe that, out of the 286 p-values, none exceeds 10 percent, four exceed five percent, and the vast majority lie below the one percent threshold. This strong evidence of structural instability motivates the use of dynamic econometric models for forecasting house price inflation.

Table 3: Stability Test Results

	EA		EM		GL		NI		NT		NW		OM	
	H	*C*	*H*	*C*	*H*	*C*	*H*	*C*	*H*	*C*	*H*	*C*	*H*	*C*
RATIO	$1.1{\cdot}10^{-3}$	$1.5{\cdot}10^{-3}$	$1.8{\cdot}10^{-3}$	$1.9{\cdot}10^{-3}$	$3{\cdot}10^{-4}$	·	·	·	0.0169	$1.3{\cdot}10^{-3}$	0.0808	0.0322	$1{\cdot}10^{-4}$	$2{\cdot}10^{-4}$
GROWTH	$6{\cdot}10^{-4}$	$8{\cdot}10^{-4}$	$8{\cdot}10^{-4}$	·	·	·	·	·	$1.7{\cdot}10^{-3}$	$1.1{\cdot}10^{-3}$	$2{\cdot}10^{-4}$	·	·	·
UR	·	·	$3{\cdot}10^{-4}$	·	·	·	$1.8{\cdot}10^{-3}$	$2.4{\cdot}10^{-3}$	·	·	·	·	·	·
LF	·	·	·	·	·	·	0.0202	$6.5{\cdot}10^{-3}$	$3.9{\cdot}10^{-3}$	$2.0{\cdot}10^{-3}$	$1{\cdot}10^{-4}$	·	$1{\cdot}10^{-4}$	·
HS	·	·	$1{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	·	·	$1.4{\cdot}10^{-3}$	0.0127	$5.2{\cdot}10^{-3}$	$4{\cdot}10^{-4}$	$2{\cdot}10^{-4}$	·	·	·
CONS	$1{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	·	·	·	$2{\cdot}10^{-4}$	$9{\cdot}10^{-4}$	$3.4{\cdot}10^{-3}$	$6.1{\cdot}10^{-3}$	$1{\cdot}10^{-4}$	$3{\cdot}10^{-4}$	·	·
INDUS	$3{\cdot}10^{-4}$	·	·	·	·	·	$1.8{\cdot}10^{-3}$	$2.3{\cdot}10^{-3}$	$2{\cdot}10^{-4}$	$3{\cdot}10^{-4}$	·	·	·	·
RABMR	$4{\cdot}10^{-4}$	·	$6{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	·	·	0.0222	$3.3{\cdot}10^{-3}$	$1.5{\cdot}10^{-3}$	·	·	·	·	·
SPREAD	·	·	·	·	·	·	0.0752	0.0859	·	$1{\cdot}10^{-4}$	·	·	·	·
CCI	·	·	$2.7{\cdot}10^{-3}$	·	·	·	·	$8{\cdot}10^{-4}$	0.0197	$3.4{\cdot}10^{-3}$	$2.8{\cdot}10^{-3}$	$4{\cdot}10^{-4}$	·	·
HPU	$1{\cdot}10^{-4}$	0.0199	·	·	·	·	·	·	$1.2{\cdot}10^{-3}$	$3.7{\cdot}10^{-3}$	·	·	·	$3{\cdot}10^{-4}$

Notes: The table reports wild-bootstrap p-values, based on $B = 9999$ iterations, of the Hausman- (H) and Chow-type (C) structural stability tests of Chen and Hong (2012) for univariate predictor regressions. The symbol · corresponds to a p-value less than $5 \cdot 10^{-5}$.

Table 4: Stability Test Results (Cont.)

	OSE		SC		SW		WM		WW		YH	
	H	*C*	*H*	*C*	*H*	*C*	*H*	*C*	*H*	*C*	*H*	*C*
RATIO	·	·	$3.7{\cdot}10^{-3}$	$5.8{\cdot}10^{-3}$	0.0180	$1.8{\cdot}10^{-3}$	0.0181	0.0112	·	·	$4.3{\cdot}10^{-3}$	$1.7{\cdot}10^{-3}$
GROWTH	·	·	$2{\cdot}10^{-4}$	·	$2{\cdot}10^{-4}$	$2{\cdot}10^{-4}$	$3{\cdot}10^{-4}$	$1{\cdot}10^{-4}$	$2.3{\cdot}10^{-3}$	$1.3{\cdot}10^{-3}$	$3{\cdot}10^{-3}$	$7.8{\cdot}10^{-3}$
UR	·	·	$1.5{\cdot}10^{-3}$	·	·	·	·	·	$1{\cdot}10^{-4}$	·	·	·
LF	·	·	$1.2{\cdot}10^{-3}$	$3{\cdot}10^{-4}$	·	·	$6{\cdot}10^{-4}$	$2{\cdot}10^{-4}$	·	·	$5.9{\cdot}10^{-3}$	$1.9{\cdot}10^{-3}$
HS	·	·	0.0192	0.0357	·	·	·	·	$4.1{\cdot}10^{-3}$	$1.3{\cdot}10^{-3}$	$8{\cdot}10^{-4}$	$9{\cdot}10^{-4}$
CONS	·	·	$1.8{\cdot}10^{-3}$	$4.8{\cdot}10^{-3}$	·	·	$3{\cdot}10^{-4}$	·	$1.0{\cdot}10^{-3}$	$8{\cdot}10^{-4}$	$5{\cdot}10^{-4}$	$1.6{\cdot}10^{-3}$
INDUS	·	·	·	$1{\cdot}10^{-4}$	·	·	·	·	0.0242	$5.7{\cdot}10^{-3}$	$2{\cdot}10^{-4}$	$3{\cdot}10^{-4}$
RABMR	·	·	$6{\cdot}10^{-4}$	·	·	·	$8{\cdot}10^{-4}$	·	$2{\cdot}10^{-4}$	·	0.0134	$4.9{\cdot}10^{-3}$
SPREAD	·	·	$1.2{\cdot}10^{-3}$	$4{\cdot}10^{-4}$	·	·	·	·	·	·	·	·
CCI	·	·	$1{\cdot}10^{-4}$	$3{\cdot}10^{-4}$	·	·	·	·	0.0446	$8{\cdot}10^{-4}$	0.0408	$6.6{\cdot}10^{-3}$
HPU	·	$1.3{\cdot}10^{-3}$	·	$1.0{\cdot}10^{-3}$	$1{\cdot}10^{-4}$	$8{\cdot}10^{-4}$	·	$1.4{\cdot}10^{-3}$	$3{\cdot}10^{-4}$	$2.6{\cdot}10^{-3}$	·	$3{\cdot}10^{-4}$

Notes: The table reports wild-bootstrap p-values, based on $B = 9999$ iterations, of the Hausman- (H) and Chow-type (C) structural stability tests of Chen and Hong (2012) for univariate predictor regressions. The symbol · corresponds to a p-value less than $5 \cdot 10^{-5}$.