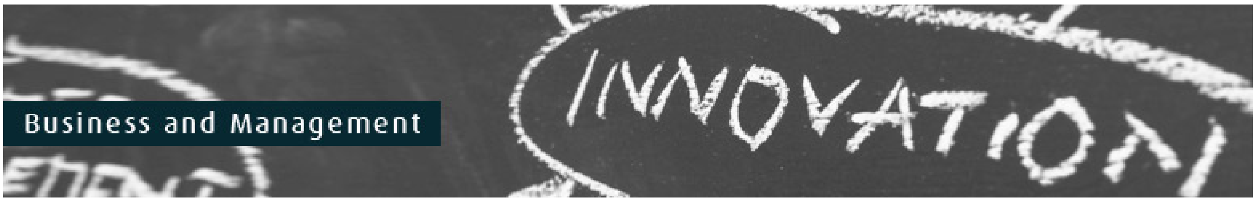


Working Paper Series

Department of Business & Management

***Macroeconomic Methodology, Theory and Economic Policy (MaMTEP)***

No. 2, 2014

# Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection

By

Lasse Bork and Stig Vinther Møller

**Lasse Bork**
Department of Business and Management
Aalborg Universitet
Fibigerstræde 2
DK-9220 Aalborg Oe
E-mail: bork@business.aau.dk

**Stig Vinther Møller**
CREATES, Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
E-mail: svm@asb.dk

February 2014

**Abstract**
We examine house price forecastability across the 50 states using Dynamic Model Averaging and Dynamic Model Selection, which allow for model change and parameter shifts. By allowing the entire forecasting model to change over time and across locations, the forecasting accuracy improves substantially. The states in which housing markets have been the most volatile are the states where model change and parameter shifts have been the most needed.

**Keywords:** Forecasting housing markets; 50 states; Kalman .ltering methods; Model change; Parameter shifts; Boom-bust cycle; Model averaging; Model selection.

**JEL classification:** C51; C52; G1; E3.

We thank seminar participants and in particular two anonymous referees for useful comments and suggestions. Møller acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

ISBN 9788791646737

# 1 Introduction

The recent collapse of the U.S. housing market has demonstrated the critical importance of being able to forecast movements in house prices. However, the choice of a proper forecasting model is not an easy task. One reason is that the U.S. housing market is characterized by many regional disparities, which implies that a given forecasting model may be useful in some regions but not others. Another reason is that the performance of specific house price predictors may be unstable over time. Thus, they may be useful at some points in time but not others.

In this paper we examine the ability to forecast house prices in each of the 50 states using Dynamic Model Averaging (DMA) and Dynamic Model Selection (DMS), which are new forecasting techniques recently developed and motivated by Raftery, Karny and Ettler (2010) and Koop and Korobilis (2012). With $m$ predictors at hand, the methods consider all the $K = 2^m$ possible model combinations in each time period $t$. The methods require computing the probability that model $k \in \{1, ..., K\}$ should be used for forecasting at time $t$, which is done using Kalman filtering methods. DMS then chooses the model with the highest probability at time $t$, while DMA uses the estimated probabilities as model weights.

The advantage of DMA and DMS is that they allow the parameters and forecasting model to change over time. Given the recent turbulence in U.S. housing markets, we argue that DMA and DMS are ideal for forecasting house prices. First, it seems likely that the marginal effects of house price predictors are time-varying due to e.g. structural breaks. Second, movements in house prices may be driven by different factors at different points in time. For example, it is possible that the best house price predictors during house price booms are different from those during house price busts. Likewise, it may also be optimal to use many house price predictors at some points in time but only a few at others. The distinguishing feature of DMA and DMS is that the methods capture not only parameter shifts but also model changes.

In the light of the strong variation in house price growth rates across different parts of the U.S., we apply DMA and DMS to forecast house price growth rates in each of the 50 states. We show that the degree of house price forecastability using DMA and DMS is strong in basically all of the 50 states. We compare with a wide range of different benchmark models and find that DMA and DMS come out as superior forecasting methods. Forecasting methods that do not allow for model change tend to perform less well in forecasting house price changes. In addition, we also find that it is important to allow the marginal effects of specific predictors to change over time, i.e. to allow for parameter shifts.

During the recent boom-bust cycle, coastal states tended to experience the largest house price increases but subsequently also the largest house price declines. We find an almost one-for-one relation between the magnitude of the boom-bust cycles and the level of forecast errors across states. The larger the boom-bust cycle, the larger the level of forecast errors. We also find a positive relation between the magnitude of the boom-bust cycles and the forecasting gains of using DMA and DMS. This result is intuitive because the benefits of model change and parameter shifts should be the highest in the most volatile housing markets.

Ghysels, Plazzi, Valkanov and Torous (2013) provide a survey of the growing empirical literature on forecasting house prices. Our contribution to this literature is to document the importance of allowing for both time-varying parameters and model change. Our paper relates most closely to Rapach and Strauss (2009) who examine differences in house price forecastilbity in the 20 largest states in the U.S. as measured by population. Rapach and Strauss (2009) do not include the recent important period with the collapse of the U.S. housing market in their analysis, so our results are not directly comparable. Nevertheless, using forecasting combination methods, they also find that it has been more difficult to obtain accurate house price forecasts in coastal states than in interior states. However, in contrast to Rapach and Strauss (2009), we tend to obtain the largest forecasting gains in coastal states.

The structure of the rest of the paper is as follows. Section 2 describes the data and provides summary statistics for the growth rates in real house prices across the 50 states. Section 3 describes how we use Dynamic Model Averaging and Dynamic Model Selection to forecast state level house prices. Section 4 provides the empirical results for each of the 50 states. Section 5 concludes.

# 2 Data and summary statistics

We use state level all-transactions house price indexes available from the Federal Housing Finance Agency (FHFA). The FHFA all-transactions indexes are constructed using repeat-sales and refinancings on the same single-family properties, and they are available on quarterly frequency back to the mid-1970s. We convert the nominal house price indexes to real units based on the consumer price index (all items) available from the Bureau of

Labor Statistics (BLS). We then calculate the annualized log real growth rate in house prices for each of the 50 states:

$$y_{i,t} = 400 \ln \left( \frac{P_{i,t}}{P_{i,t-1}} \right) \qquad i = 1, ..., 50 \tag{1}$$

where $P_{i,t}$ denotes the level of real house prices in state $i$ at time $t$.

Fig. 1 shows the mean growth rates in house prices for each of the 50 states in our sample period from 1976:2 to 2012:4 (inflation-adjusted and annualized). The highest mean growth rates for the entire period have been in states on the north-east and west coasts of the U.S., including Massachusetts, California, Vermont, Washington and New York. For these states the overall mean growth rates are above 1.5%. The lowest mean growth rates for the whole period have been in states such as West Virginia, Nevada and Mississippi where the mean growth rates have been less than $-0.5\%$. The negative growth rates are mainly driven by the recent collapse of the housing markets.

To illustrate the dramatic changes in house prices during the recent boom and bust of the housing markets, Fig. 2 plots mean growth rates over the 1995:1-2006:4 period, while Fig. 3 plots mean growth rates over the 2007:1-2012:4 period. Although most state level house prices peaked around the end of 2006, there is cross-state variation in the timing of the boom-bust cycles. To enhance comparison, we assume the same timing across all states. In the boom period, which we assume is 1995:1-2006:4, mean growth rates range from 1.07% in Indiana to 7.10% in California. In the bust period, which we assume is 2007:1-2012:4, mean growth rates range from $-14.81\%$ in Nevada to 1.49% in North Dakota; the only state with a positive growth rate during the bust period. The figures illustrate that several coastal states, e.g. California and Florida, have had very large positive growth rates during the boom but also very large negative growth rates during the bust.

In general, the states in which house prices rose the most during the boom are also the states in which house prices declined the most during the bust. Fig. 4 illustrates this empirical pattern clearly. The figure plots mean growth rates during the boom against mean growth rates during the bust. The correlation is $-0.73$. Large positive growth rates in the boom period have been followed by large negative growth rates in the bust period, while small positive growth rates in the boom period have been followed by small negative growth rates in the bust period.

The strong regional differences suggest that it is necessary with both state level and aggregate information variables to predict house prices. The most popular house price indicators in the literature include valuation ratios, labor market variables, business cycle indicators and interest rate related variables, see e.g. Hamilton and Schwab (1985), Case and Shiller (1990), Malpezzi (1999), Rapach and Strauss (2009), and Bork and Møller (2013). For each state, our set of house price predictors is based on ten variables in total; five measured at the state level and five at the national level. We transform all variables to obtain stationarity.[1] The five state level variables are the price-income ratio (in logs), the unemployment rate, real per capita income growth (in logs and annualized), labor force growth (in logs and annualized) and the lagged real house price growth rate.[2] The five national variables are the 30-years mortgage rate (in first differences), the spread between 10-year and 3-month Treasury rates, industrial production growth (in logs and annual-

[1] Across all 50 states, we have carried out Augmented Dickey-Fuller (ADF) tests for all transformed variables. The ADF tests generally reject non-stationarity.

[2] State level data on income, the labor force and the unemployment rate are available from BLS.

ized), real consumption growth (in logs and annualized) and housing starts (in logs).[3] In the out-of-sample predictions, we take into account publication lags of macroeconomic variables by lagging them an additional quarter.

# 3 DMA and DMS

Dynamic Model Averaging (DMA) and Dynamic Model Selection (DMS) are developed by Raftery, Karny, and Ettler (2010) and further motivated by Koop and Korobilis (2012). In the following we make a brief description of how we use DMA and DMS to forecast state level house prices.

DMA and DMS build on time-varying parameter (TVP) models (state indices are suppressed):

$$y_t = x'_{t-1}\beta_t + \varepsilon_t \tag{2}$$

$$\beta_t = \beta_{t-1} + \eta_t \tag{3}$$

where $y_t$ is the real house price growth rate, $x_{t-1}$ is an $m$-vector of predictors (including an intercept), $\beta_t$ is an $m$-vector of coefficients, and the innovations are distributed as $\varepsilon_t \overset{ind}{\sim} N(0, V_t)$ and $\eta_t \overset{ind}{\sim} N(0, W_t)$.

The TVP model in (2)-(3) can be estimated straightforwardly using Kalman filter methods. However, the TVP model assumes that the same set of predictors should be used in all time periods, which may not be optimal when forecasting house prices. DMA and

[3] All national variables are downloaded from the FRED database of St. Louis Fed.

DMS extend the TVP model by allowing for model change. To illustrate this, we consider the case of multiple models based on various subsets of the variables in $x_{t-1}$:

$$y_t = x_{t-1}^{(k)\prime}\beta_t^{(k)} + \varepsilon_t^{(k)} \tag{4}$$

$$\beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_t^{(k)} \tag{5}$$

where $k = 1, ..., K$ denotes a particular model. $\varepsilon_t^{(k)}$ is $N\left(0, V_t^{(k)}\right)$ and $\eta_t^{(k)}$ is $N\left(0, W_t^{(k)}\right)$. Furthermore, it is useful to introduce a model indicator $L_t$ so that $L_t = k$ means that model $k$ is selected. DMA and DMS are then implemented by computing $\Pr\left(L_t = k \mid Y^{t-1}\right)$ for $k = 1, ..., K$, where $Y^{t-1} = \{y_1, ..., y_{t-1}\}$, i.e. computing the probability that model $k$ should be used to forecast $y_t$, given information through time $t-1$. We denote these model probabilities by $\pi_{t|t-1,k} = \Pr\left(L_t = k \mid Y^{t-1}\right)$. The idea behind DMS is to choose the model with the highest probability in each time period, while DMA uses the probabilities as model weights to compute the average of the $K$ forecasts.[4] Accordingly, we calculate recursive forecasts of $y_t$ conditional on $Y^{t-1}$ using either DMA or DMS as:

$$\hat{y}_t^{DMA} = \sum_{k=1}^{K} \pi_{t|t-1,k}\, x_{t-1}^{(k)\prime}\hat{\beta}_{t-1}^{(k)} \tag{6}$$

$$\hat{y}_t^{DMS} = x_{t-1}^{(k^*)\prime}\hat{\beta}_{t-1}^{(k^*)} \tag{7}$$

where $\hat{\beta}_{t-1}^{(k)}$ is the parameter prediction and where $k^*$ in the last equation refers to the model with the maximum model probability at time $t-1$.

Aiolfi and Timmermann (2006) suggest the use of clustering techniques as a way to filter

[4] Raftery, Karny and Ettler (2010) focus on DMA in an industrial application, while Koop and Korobilis (2012) use both DMA and DMS to forecast inflation. Koop and Korobilis (2012) provide Matlab code for implementing DMA and DMS. We thankfully make use of their code in our application.

out the worst performing models. As a supplement to DMS and DMA, we implement clustering in the following way. We sort the $K$ models into $P$ equal-sized clusters $C_p$ for $p = 1, ..., P$ based on descending model probabilities. We then compute the cluster forecast with the highest model probabilities denoted by $\hat{y}_t^{DMS^{C_1}}$ following the principles in equation (6) with $k$ in the summation being replaced with $k \in C_1$ and with model probabilities appropriately rescaled to sum to one.

Three simplifying assumptions are made in Raftery, Karny and Ettler (2010) which imply that $\hat{y}_t^{DMA}$ or $\hat{y}_t^{DMS}$ can be computed in a single pass of the Kalman filter for each of the $K$ models. The first two assumptions imply that the parameter predictions $\hat{\beta}_{t-1}^{(k)}$ can be calculated independently for each model. In particular, assuming that $\beta_{t-1}^{(k)}$ is only defined when $L_{t-1} = k$ , and simplifying the estimation of the state error covariance matrix using a forgetting factor $\lambda$, simplifies the Kalman filter prediction of the parameters to:

$$\hat{\beta}_{t|t-1}^{(k)} = \hat{\beta}_{t-1|t-1}^{(k)} \tag{8}$$

$$\Sigma_{t|t-1}^{(k)} = \frac{1}{\lambda}\Sigma_{t-1|t-1}^{(k)} \tag{9}$$

where $\Sigma_{t|t-1}^{(k)}$ denotes the covariance matrix of $\beta_{t-1}^{(k)}$, while the updating equations simplify to:

$$\hat{\beta}_{t|t}^{(k)} = \hat{\beta}_{t|t-1}^{(k)} + \Sigma_{t|t-1}^{(k)} x_{t-1}^{(k)\prime} \left( V_t^{(k)} + x_{t-1}^{(k)\prime} \Sigma_{t|t-1}^{(k)} x_{t-1}^{(k)} \right)^{-1} \left( y_t - x_{t-1}^{(k)\prime} \hat{\beta}_{t-1}^{(k)} \right) \tag{10}$$

$$\Sigma_{t|t}^{(k)} = \Sigma_{t|t-1}^{(k)} - \Sigma_{t|t-1}^{(k)} x_{t-1}^{(k)\prime} \left( V_t^{(k)} + x_{t-1}^{(k)\prime} \Sigma_{t|t-1}^{(k)} x_{t-1}^{(k)} \right)^{-1} x_{t-1}^{(k)} \Sigma_{t|t-1}^{(k)} \tag{11}$$

The forgetting factor approach has been widely used in adaptive filtering and implies that estimation of the state error covariance during the Kalman filtering can be simplified to $W_t^{(k)} = \left(\lambda^{-1} - 1\right) \Sigma_{t|t-1}^{(k)}$, which in turn leads to (9).[5] The third assumption is related to the model prediction component in (6)-(7) involving another forgetting factor $\alpha$ that also leads to a very significant computational gain as we avoid specifying a large transition probability matrix. To illustrate the gain, note that we consider all possible combinations of the $m$ predictors, so that the number of models is $K = 2^m$. With $m = 10$, we therefore consider $K = 1024$ models in each time period. Since $K$ is large, it is not feasible to do Markov switching using a $K \times K$ transition matrix. The large dimension of the transition matrix would lead to imprecise inferences and excessive computation time. Consequently, the model prediction is defined by:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{\ell=1}^{K} \pi_{t-1|t-1,\ell}^{\alpha}} \tag{12}$$

where $\alpha$ is set to a value just below one, while the model updating equation is:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k \left(y_t | Y^{t-1}\right)}{\sum_{\ell=1}^{K} \pi_{t|t-1,\ell} f_\ell \left(y_t | Y^{t-1}\right)} \tag{13}$$

where $f_\ell \left(y_t | Y^{t-1}\right)$ is the predictive density of model $\ell$, i.e. the density of a $N\left(x_{t-1}^{(\ell)\prime} \hat{\beta}_{t-1}^{(\ell)}, V_t^{(\ell)} + x_{t-1}^{(\ell)\prime} \Sigma_{t|t-1}^{(\ell)} x_{t-1}^{(\ell)}\right)$ distribution evaluated at $y_t$. The equations (8)-(13) represent the complete Kalman filter prediction and updating equations.

[5] $\lambda$ is usually slightly below 1 and essentially implies that estimation is based on age-weighed data.

## 3.1 Time-varying forgetting factors

In some periods it may be attractive to allow for more rapid model switches and parameter shifts than in others. Consequently, we relax the assumption of constant forgetting factors. Similar to McCormick, Raftery, Madigan and Burd (2012), we select $\alpha_t$ as:

$$\arg\max_{\alpha_t} \sum_{k=1}^{K} f_k\left(y_t|Y^{t-1}\right) \frac{\pi_{t-1|t-1,k}^{\alpha_t}}{\sum_{\ell=1}^{K} \pi_{t-1|t-1,\ell}^{\alpha_t}} \tag{14}$$

where $\alpha_t$ can take on a finite number of values. We run the Kalman filter in parallel for each candidate value of $\alpha_t$. We then select the value of $\alpha_t$ that gives the highest predictive likelihood across the $K$ models at each time period. To limit the computational burden, $\alpha_t$ can only take on five values: $\alpha_t \in \{0.95, ..., 0.99\}$. This interval is also used by e.g. Koop and Korobilis (2012).

Typically, time-variation in $\lambda$ is modelled based on the most recent squared prediction error as in Fortescue, Kershenbaum and Ydstie (1981) and others. However, we propose to calculate model specific time-variation in $\lambda$ based on the entire history of model $k$'s squared prediction errors in order to limit excess sensitivity to a single large prediction error. Specifically, we calculate $\Psi$ quantiles of an expanding history of squared prediction errors, $\mathcal{E}_k^{t-1} = \left\{\varepsilon_{k,1}^2, ..., \varepsilon_{k,t-1}^2\right\}$, and denote these quantiles by $q_\psi\left(\mathcal{E}_k^{t-1}\right)$ for $\psi = 1, .., \Psi$. Then we select $\lambda_t^{(k)} = \lambda_{t-1}^{(k)}$ if $\varepsilon_{t,k}^2$ is within the same interval, $\left]q_{\psi-1}\left(\mathcal{E}_k^{t-1}\right), q_\psi\left(\mathcal{E}_k^{t-1}\right)\right]$, as in the previous period. On the other hand, $\lambda_t^{(k)}$ steps up if the current prediction error is relatively low, while $\lambda_t^{(k)}$ steps down if the current prediction errors is relatively high. $\lambda_t^{(k)}$ can take on the same set of values as $\alpha_t$; thus $\lambda_t^{(k)}$ can effectively move around in a bounded interval. Our empirical results are robust to allowing $\lambda_t^{(k)}$ to move around in a multi-step fashion.

### 3.2 The case of no forgetting

Raftery, Karny and Ettler (2010) show that standard Bayesian Model Averaging (BMA) (Leamer, 1978; Hoeting, Madigan, Raftery and Volinsky, 1999) is a special case of DMA. In particular, if we set $\alpha_t = \lambda_t = 1$, there is no forgetting and we recover static BMA. In a similar way, when there is no forgetting, DMS becomes equivalent to static Bayesian Model Selection (BMS). In contrast, allowing for forgetting, DMA and DMS imply that the model indicator and model parameters evolve in time, which is why these methods are called dynamic.

## 4 Empirical results

We consider the following candidate forecasting models:

1. DMA: Dynamic Model Averaging uses model probabilities as weights to compute the average forecast.

2. DMS: Dynamic Model Selection puts all the weight on the model with the highest probability.

3. DMS$^{C_1}$: Dynamic Model Selection with clustering uses the best cluster of models measured by model probabilities.[6]

[6] We use $P = 16$ clusters, implying that the number of models in each cluster is $K/P = 64$.

4. BMA: DMA with forgetting factors fixed at $\alpha_t = \lambda_t = 1$.

5. BMS: DMS with forgetting factors fixed at $\alpha_t = \lambda_t = 1$.

6. EW: Equal weighting of forecasts from $K$ OLS regression models.

7. ALL: A "kitchen sink" approach with OLS forecasts using all predictors.

8. AR1: OLS forecasts using a constant and the lagged house price growth rate.

9. MEAN: OLS forecasts using a constant only.

The OLS regressions are estimated recursively using an expanding window.

We evaluate the forecasting performance of the models using the Mean-Squared-Forecast-Error (MSFE). The out-of-sample period is from 1995:1 to 2012:4, which covers the recent boom and bust periods. Table 1 gives an overview of the main results. The table reports summary statistics of the MSFE ratios of the various models relative to the MEAN (constant only) benchmark. On average across the 50 states, DMA and DMS$^{C_1}$ perform the best. The average MSFE reduction relative to the MEAN benchmark is about 25% for these two methods. The gain in forecasting power is consistent across states: DMA and DMS$^{C_1}$ generate MSFE ratios less than one in all 50 states. Fig. 5 maps the MSFE ratios of DMA relative to the MEAN benchmark. The figure illustrates that DMA is far superior to the historical mean. The lowest ratio is obtained for Hawaii (0.545), while the highest ratio is obtained for Utah (0.961).

DMS selects the single "best" model in each time period, while DMS$^{C_1}$ selects the "best" cluster of models in each time period. In far the majority of states, the two approaches give similar results, but for a few states DMS performs worse than DMS$^{C_1}$. For instance, in Kentucky the MSFE ratio is 1.165 using DMS compared to 0.972 using DMS$^{C_1}$. Since DMS does not offer diversification gains, it may lead to inaccurate forecasts in a few cases. On the other hand, out of all models considered, DMS produces the lowest minimum MSFE ratio across the 50 states (0.501 in Nevada).

Combination of forecasts from OLS regression models have been shown in empirical studies to produce better forecasts on average than forecasts from individual OLS regression models. For example, Rapach and Strauss (2009) find that combining forecasts from OLS models does well in predicting house price changes. We confirm this result. The EW approach produces an average MSFE reduction of 17%, and it does better than the MEAN benchmark in 44 out of 50 states.

BMA and BMS also outperform the MEAN benchmark with average MSFE reductions of 14% and 9%, respectively. The MSFE ratio is less than one in 44 states using BMA and in 35 states using BMS.

The AR1 model is on average somewhat better than the historical mean benchmark. It generates an average MSFE reduction of about 4% and produces MSFE ratios less than one in 27 out of 50 states. This result reflects that the growth rates in some states are positively autocorrelated. Fig. 6 maps the MSFE ratio of DMA relative to the AR1 model. The figure shows that DMA convincingly outperforms the AR1 model: The MSFE ratio is less than one in 49 out of 50 states (Utah only exception).

The far worst method is OLS forecasts with all predictors, which most likely is due to the well-known problem that too many variables can lead to overfitting and inaccurate forecasts. Given its poor performance, we do not include this benchmark in the analysis below.

### 4.1 Statistical tests

In this section, we formally test whether the various forecasting methods generate lower MSFEs across states than the historical mean benchmark. Clark and McCracken (2001) show that the commonly used Diebold and Mariano (1995) statistic has a nonstandard distribution when testing for equal accuracy of forecasts from nested models. Clark and West (2007) propose an adjusted statistic which is approximately normally distributed when comparing forecasts from nested models. We first define:

$$f_{j,t} = (y_t - \hat{y}_{\text{MEAN},t})^2 - (y_t - \hat{y}_{j,t})^2 + (\hat{y}_{\text{MEAN},t} - \hat{y}_{j,t})^2 \tag{15}$$

where $\hat{y}_{\text{MEAN},t}$ denotes the forecast of $y_t$ using the historical mean benchmark, and $\hat{y}_{j,t}$ denotes the forecast from model $j$ = AR1, EW, BMA, BMS, DMA, DMS, DMS$^{C_1}$. To carry out the Clark-West test, we regress $f_{j,t}$ on a constant and then use the resulting $t$-statistic to test for a zero coefficient. We thereby test the null hypothesis that the MSFE ratio between model $j$ and the nested MEAN benchmark is greater than or equal to one with the alternative hypothesis that it is less than one.

Table 2 reports MSFE ratios and Clark-West $t$-statistics in parenthesis. The null hypothesis is rejected if the $t$-statistic is greater than 1.645 (for a one sided 5% test). In that case, the corresponding MSFE ratio is in bold font. From the table, we see that DMA delivers statistically significant out-of-sample gains relative to the MEAN benchmark in all 50 states. The DMS$^{C_1}$ approach significantly outperforms the MEAN benchmark in 49 states. The only exception is Kentucky where the MSFE reduction of 3% is insignificant at the 5% level. The DMS approach also performs convincingly with significant out-of-sample gains in 48 states (Kansas and Kentucky are the exceptions). For Colorado and Oklahoma, DMS generates MSFE ratios greater than one but the null that the MEAN benchmark performs equally well or better than DMS is still rejected. The reason is that when comparing a parsimonious null model to a larger model that nests the null model, the Clark-West test takes into account added estimation uncertainty from estimating parameters in the larger model that are zero under the null. In particular, under the null, we should expect a MSFE ratio greater than one due to a gain in estimation efficiency for the parsimonious null model using a finite sample.

Overall, the results show that DMA, DMS and DMS$^{C_1}$ are robust forecasting methods as they consistently outperform the MEAN benchmark. The next best forecasting methods appear to be the EW and BMA approaches. They both generate significantly lower MSFEs than the MEAN benchmark in 44 states, while BMS significantly outperforms the MEAN benchmark in 39 states. Rather than comparing with the MEAN benchmark in the Clark-West tests, we have also tried comparing with the AR1 model. This leads to the same ranking of forecasting methods. As Table 2 shows the AR1 model significantly outperforms the MEAN benchmark in 20 states only.

The final row of Table 2 provides results for the aggregate U.S. housing market. We use the same predictive variables as in the above but replace state-level variables with national variables. The DMA, DMS and DMS$^{C_1}$ methods again perform well. For these three methods the MSFE reductions are 33% or more, and the improvement in forecasting accuracy is strongly statistically significant.

## 4.2 Performance over time

To visualize the out-of-sample performance of the various forecasting methods, we follow the suggestion of Goyal and Welch (2003) and plot the difference between the cumulative squared forecast error for the MEAN benchmark and the cumulative squared forecast error for model $j$ during the out-of-sample period:

$$CDSFE_{j,t} = \sum_{t=1995:1}^{2012:4} \left(e^2_{\mathrm{MEAN},t} - e^2_{j,t}\right) \tag{16}$$

By plotting the cumulative sums up to each point in time, we can assess the stability of the out-of-sample performance of model $j$. To get the overall picture for the 50 states, Fig. 7 plots $\sum_{i=1}^{50} CDSFE_{j,i,t}$ with $j$ = AR1, BMA, EW, DMA, DMS, and where subscript $i$ refers to the $i$th state. The figure identifies the time periods where model $j$ succeeded in beating the MEAN benchmark (positive slope) and the time periods where model $j$ failed to beat the MEAN benchmark (negative slope). The figure illustrates that DMA and DMS are better at capturing the boom period from the mid-90s up to around 2006 than the other forecasting methods. DMA and DMS also do really well in capturing the initial collapse of the housing markets around the period from 2007 to 2008. However, DMA and DMS as well as the other forecasting methods all experience deterioration in the predictive ability in 2008:4 and 2009:1. In these two quarters, many states had positive growth rates after a series of negative growth rates.

## 4.3 Forecast performance and the volatility of housing markets

The results in Table 2 illustrate that DMA, DMS and DMS$^{C_1}$ work well in basically all states. The gains of using these methods are therefore not concentrated to specific geographic areas. However, looking at the level of forecast errors, interesting patterns arise across the states. In Fig. 8, we relate the MSFE level using DMA to the magnitude of the boom-bust cycles across the 50 states. The figure plots the MSFE level against the growth volatility during the out-of-sample period from 1995:1 to 2012:4. The correlation is 0.97, i.e. there is an almost one-for-one relation between growth volatility and forecasting accuracy. The higher the growth volatility (the larger the boom-bust cycle), the larger is the MSFE level. This strong pattern implies that it has been more difficult to obtain accurate forecasts in coastal states where housing markets have tended to be more volatile than in interior states. Rapach and Strauss (2009) find a similar pattern across states.

The gain in forecasting accuracy of allowing for model change and parameter shifts is highest in the most volatile housing markets. To illustrate this, we consider the top four and bottom four states in terms of housing market volatility. Arizona, California, Florida and Nevada are the states with the highest volatility, whereas Iowa, Oklahoma, Kansas and Texas are the states with the lowest volatility. Fig. 9 plots realized house price growth rates together with DMA forecasts (to keep the figure readable, it only includes DMA forecasts). Focusing on the top four volatility states, the figure shows that DMA captures a large part of the sharp decline in house price growth rates around the recent crash in housing markets. In Fig. 10, forecasted and realized growth rates are transformed into indexes with base 1994:4 = 100. Besides DMA, we include DMS, EW and the AR1 model. From the figure, we see that DMA and DMS capture a much larger part of the big housing booms in Arizona, California, Florida and Nevada than the other forecasting methods. At the same time, DMA and DMS tend to forecast the bust faster in those states relative to the other forecasting methods, which most likely reflects that DMA and DMS are able to adjust more quickly to structural changes. Turning to Iowa, Oklahoma, Kansas and Texas, we observe much more stable housing markets and therefore also that the forecasting gains of allowing for model change and parameters shifts are less obvious.

## 4.4 Dimension of models and choice of variables

DMA uses $\pi_{t|t-1,k}$ as weights for each of the $k = 1, ..., K$ models. Thus, following Koop and Korobilis (2012), we compute the expected (or average) number of predictors used by DMA at time $t$ as:

$$E\left(\text{size}_t\right) = \sum_{k=1}^{K} \pi_{t|t-1,k} \text{size}_{(k)} \tag{17}$$

where $\text{size}_{(k)}$ denotes the number of predictors in model $k$. Fig. 11 plots the medium value of $E\left(\text{size}_t\right)$ across the 50 states together with the 16th and 84th percentiles. The purpose of the figure is to give an idea about variations in the degree of parsimony over time and across states. The figure shows that the number of predictors used by DMA changes over time, but also that DMA generally tends to favor parsimonious models. Typically, the average number of predictors used by DMA is less than half out of the ten predictors in the data set. During the out-of-sample window from 1995:1 to 2012:4, the medium value ranges from around two to five predictive variables and the 84th percentile never exceeds six predictive variables. The span between the 16th and 84th percentiles indicates a certain variation across states regarding the optimal dimension of the models. To shed further light on this, Fig. 12 plots $E(\text{size}_t)$ for the top four and bottom four states in terms of housing market volatility. There is substantial shrinkage in both high and low volatility states but typically more parsimonious models are preferred in low volatility states.

It is also possible to use $\pi_{t|t-1,k}$ to illustrate which predictors are the most important across states and over time. The posterior inclusion probability of a given predictive variable is defined as the probability that DMA attaches to models that include that particular predictive variable. Fig. 13 plots posterior inclusion probabilities for each of the ten predictive variables. The figure plots the median as well as the 16th and 84th percentiles across the 50 states. The figure illustrates that the instability in housing markets around the recent collapse gives rise to significant model changes. In that period of time, the price-income ratio and housing starts become increasingly important predictive variables. In fact, in some of the states, DMA attaches a probability that is close to 1 to housing starts. Interestingly, there is a general tendency of rather large spreads in the 16th and 84 percentiles for most of the predictive variables, suggesting that it is not the same predictive variables that drive housing markets. In other words, housing markets are segmented.

Fig. 14 plots posterior inclusion probabilities of the four most important variables in the top four and bottom four states in terms of housing market volatility (the four variables that most frequently have an inclusion probability exceeding 50% during the out-of-sample period). The patterns in inclusion probabilities are complicated and there is not a single variable which is always important across states and over time. However, in the most volatile states, housing starts and the term spread tend to be important, while in the most stable states, unemployment tends to be important. Focusing on the period around the bust, housing starts tend to become increasingly important in states such as Arizona and Nevada, while the price-income ratio receives a very high weight in Florida. In general, Fig. 14 demonstrates strong variation over time and across states regarding which particular model specification works the best.

## 4.5 More benchmarks

### 4.5.1 Rolling OLS models

In the above, the OLS regressions for the EW and AR1 models are estimated recursively using an expanding window, so that all available information at the time of the forecast is used to estimate the coefficients. In case of parameter instability due to e.g. structural breaks, rolling windows may lead to more accurate forecasts than expanding windows because more rapid changes in the coefficients can be achieved using rolling OLS. It is therefore interesting to compare the forecast performance of DMA, DMS and DMS$^{C_1}$, which are designed to capture parameter shifts, with rolling OLS models. Tables 1 and 2 show that the best OLS-based method is forecast combination using equal weights (EW). Thus, here we focus on the EW model. A drawback of using a rolling window scheme is the arbitrariness of the choice of window size. We have tried using 5-year and 15-year windows. With a 5-year window, the EW model produces an average MSFE ratio relative to the MEAN benchmark across the 50 states of 0.81, and it outperforms the MEAN benchmark in 42 out of 50 states. With a 15-year window, the performance is slightly better: The average MSFE ratio is 0.79, and the EW model now outperforms the MEAN benchmark in 45 states. Still, the EW model does not match the performance of DMA, which outperforms the MEAN benchmark in all 50 states.

#### 4.5.2 Fixed model weights but time variation in coefficients

We implement BMA and BMS (DMA and DMS with $\alpha_t = \lambda_t = 1$) in a recursive way, which means that some time variation in the coefficients will arise. We have also considered DMA and DMS with $\alpha_t = 1$ and $\lambda_t \in \{0.95, ..., 0.99\}$, so that the model weights are fixed but the coefficients are allowed to change more rapidly. To be precise, we allowed for model specific time-variation in $\lambda$ following the procedure described in Section 3.1. We just briefly mention here that DMA with $\alpha_t = 1$ and $\lambda_t^{(k)} \in \{0.95, ..., 0.99\}$ gives an average MSFE ratio across the 50 states of 0.84, compared to 0.75 for DMA with $\alpha_t \in \{0.95, ..., 0.99\}$ and $\lambda_t^{(k)} \in \{0.95, ..., 0.99\}$, i.e. time-varying model weights are advantageous.

#### 4.5.3 TVP models

We have also compared with TVP models, which allow the marginal effects of the predictors to change over time but assume no model change. To analyze the performance of the TVP model approach, we have to choose a particular model, which is assumed to hold over time. That is, we have to choose one of the $K = 1024$ models and stick with that model. If we allow all predictors to enter into the model specification, the results come out almost as poorly as with the OLS kitchen sink approach. Thus, a TVP model with the full set of predictors seems to suffer from overfitting. However, simply removing some of the predictors does not necessarily help. The reason is that housing markets are not stable over time, which means that different factors drive house prices at different points in time. Thus, if too few predictors are used in a given model specification, it might lead to misspecification problems. In this respect, the useful property of DMA is that it becomes possible to include a fair amount of house price predictors without running into problems with overfitting.

# 5 Conclusion

This paper analyzes the ability to predict real house price growth rates in the 50 states using DMA and DMS, which are new forecasting methods that allow for both the model and coefficients to change over time. With DMA and DMS it is possible to gain insights into which variables that drive house prices over time. We find that there is no single variable that really stands out as being the most important. The best variables for predicting house prices vary a lot over time and across states. We also find that allowing for model change and parameter shifts leads to improvements in forecasting accuracy relative to a wide range of benchmark models. We show that the states which have had the most volatile housing markets are the states where model change and parameter shifts are the most needed.

Aiolfi, M. & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics, 135, 31-53.

Bork, L. & Møller, S.V. (2013). House price forecastability: A factor analysis. SSRN paper no 2065343 (http://ssrn.com/abstract=2065343).

Case, K.E. & Shiller, R.J. (1990). Forecasting prices and excess returns in the housing market. AREUA Journal, 18, 253-273.

Clark, T.E. & McCracken, M.W. (2001). Tests of forecast accuracy and encompassing for nested models. Journal of Econometrics, 105, 85-110.

Clark, T.E. & West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138, 291-311.

Diebold, F.X. & Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253-263.

Fortescue, T.R., Kershenbaum, L.S. & Ydstie, B.E. (1981). Implementation of self-tuning regulators with variable forgetting factors. Automatica, 17, 831-835.

Ghysels, E., Plazzi, A. Valkanov, R. & Torous, W. (2012). Forecasting Real Estate Prices. Handbook of Economic Forecasting.

Goyal, A. & Welch, I. (2003). Predicting the equity premium with dividend ratios. Management Science, 49, 639-654.

Hamilton, B.W. & Schwab, R.M. (1985). Expected appreciation in urban housing markets. Journal of Urban Economics, 18, 103-118.

Hoeting, J.A., Madigan, D. Raftery, A.E. & Volinsky, C.T. (1999). Bayesian model averaging: A tutorial. Statistical Science, 14, 382-417.

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

Leamer, E.E. (1978). Specification Searches. Wiley, New York.

Malpezzi, S. (1999). A simple error correction model of house prices. Journal of Housing Economics, 8, 27-62.

McCormick, T.M., Raftery, A.E., Madigan, D. & Burd, R.S. (2012). Dynamic logistic regression and dynamic model averaging for binary classification. Biometrics, 68, 23-30.

Raftery, A.E., Karny, M. & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to cold rolling mill. Technometrics, 52, 52-56.

Rapach, D.E. & Strauss, J.K. (2009). Differences in housing price forecastability across U.S. states. International Journal of Forecasting, 25, 351-372.

Table 1. Forecast errors across states

| | Model | Avg. | Std. dev. | Min. | Max. | # < 1 |
|---|---|---|---|---|---|---|
| 1. | DMA | 0.751 | 0.12 | 0.546 | 0.961 | 50 |
| 2. | $DMS^{C_1}$ | 0.752 | 0.12 | 0.556 | 0.972 | 50 |
| 3. | DMS | 0.796 | 0.15 | 0.501 | 1.165 | 46 |
| 4 | EW | 0.831 | 0.24 | 0.548 | 1.657 | 44 |
| 5. | BMA | 0.858 | 0.14 | 0.548 | 1.177 | 44 |
| 6. | BMS | 0.903 | 0.15 | 0.591 | 1.188 | 35 |
| 7. | AR1 | 0.957 | 0.12 | 0.599 | 1.228 | 27 |
| 8. | ALL | 1.208 | 0.95 | 0.606 | 5.565 | 31 |

The table shows summary statistics of MSFE ratios of various forecasting models relative to the MEAN benchmark. Avg., Std. dev., Min. and Max. refer to the average, standard deviation, minimum and maximum of the MSFE ratios across the 50 states. The "# < 1" column reports the number of MSFE ratios less than one. The out-of-sample window is 1995:1 to 2012:4.

Table 2. Test for equal predictive ability

| | AR1 | BMS | BMA | EW | DMS | DMS$^{C_1}$ | DMA |
|---|---|---|---|---|---|---|---|
| Alabama | 1.02 | **0.89** | **0.91** | **0.92** | **0.86** | **0.80** | **0.81** |
| | (−0.70) | (3.78) | (3.06) | (1.73) | (2.49) | (2.68) | (2.69) |
| Alaska | 1.01 | **0.81** | **0.82** | 1.64 | **0.88** | **0.86** | **0.84** |
| | (−2.72) | (3.92) | (4.01) | (−1.87) | (3.10) | (3.06) | (3.12) |
| Arizona | 0.97 | **0.70** | **0.69** | **0.74** | **0.64** | **0.60** | **0.62** |
| | (1.17) | (3.77) | (3.61) | (2.66) | (3.16) | (3.35) | (3.44) |
| Arkansas | 1.00 | **0.89** | **0.88** | **0.87** | **0.96** | **0.88** | **0.89** |
| | (0.16) | (2.86) | (3.13) | (3.13) | (3.02) | (2.80) | (2.83) |
| California | **0.60** | **0.59** | **0.59** | **0.60** | **0.56** | **0.56** | **0.56** |
| | (3.06) | (3.29) | (3.35) | (3.52) | (3.27) | (3.40) | (3.40) |
| Colorado | 0.97 | 1.12 | 1.09 | **0.77** | **1.15** | **0.92** | **0.90** |
| | (1.20) | (1.37) | (1.59) | (3.44) | (1.93) | (2.64) | (2.77) |
| Connecticut | **0.91** | **0.79** | **0.82** | **0.63** | **0.65** | **0.92** | **0.65** |
| | (3.23) | (4.04) | (3.82) | (4.95) | (4.06) | (3.72) | (4.19) |
| Delaware | **0.96** | **1.00** | **0.77** | **0.64** | **0.66** | **0.60** | **0.62** |
| | (2.74) | (2.58) | (3.38) | (4.24) | (4.21) | (4.18) | (4.20) |
| Florida | **0.89** | **0.67** | **0.73** | **0.76** | **0.60** | **0.61** | **0.61** |
| | (2.20) | (3.59) | (3.72) | (3.47) | (3.46) | (3.68) | (3.64) |
| Georgia | 1.01 | 0.91 | **0.85** | **0.73** | **0.57** | **0.69** | **0.55** |
| | (0.22) | (1.49) | (2.25) | (2.91) | (4.74) | (4.70) | (4.76) |
| Hawaii | 1.00 | **0.82** | **0.74** | **0.73** | **0.69** | **0.57** | **0.55** |
| | (−1.51) | (4.04) | (4.18) | (3.13) | (4.70) | (4.74) | (4.76) |
| Idaho | **0.92** | 1.18 | **0.87** | **0.78** | **0.80** | **0.72** | **0.73** |
| | (1.76) | (0.42) | (2.14) | (2.45) | (2.62) | (2.68) | (2.65) |
| Illinois | **1.00** | **0.72** | **0.74** | **0.75** | **0.70** | **0.67** | **0.68** |
| | (1.81) | (3.36) | (3.27) | (2.67) | (2.85) | (3.02) | (3.07) |
| Indiana | 1.23 | 1.04 | 0.99 | 0.99 | **0.99** | **0.95** | **0.95** |
| | (−0.94) | (−0.07) | (1.62) | (1.34) | (1.72) | (1.78) | (1.76) |
| Iowa | 1.05 | **0.97** | **0.96** | **0.87** | **0.91** | **0.95** | **0.88** |
| | (0.35) | (3.76) | (3.50) | (2.62) | (3.42) | (2.61) | (3.22) |
| Kansas | 1.16 | 1.19 | 1.18 | **0.96** | 1.12 | **0.94** | **0.93** |
| | (0.54) | (0.52) | (0.53) | (1.67) | (1.47) | (2.25) | (2.44) |
| Kentucky | 1.01 | 1.04 | 0.98 | **0.86** | 1.16 | 0.97 | **0.90** |
| | (−0.45) | (0.26) | (1.39) | (2.99) | (0.38) | (1.38) | (2.32) |
| Louisiana | 0.95 | **1.06** | **0.90** | **0.81** | **0.80** | **0.78** | **0.80** |
| | (1.40) | (3.35) | (4.00) | (4.06) | (3.83) | (4.22) | (4.16) |
| Maine | **0.98** | **0.90** | **0.83** | **0.63** | **0.75** | **0.68** | **0.67** |
| | (2.25) | (2.91) | (3.09) | (4.41) | (4.00) | (4.06) | (4.06) |
| Maryland | **0.73** | **0.65** | **0.61** | **0.62** | **0.59** | **0.58** | **0.58** |
| | (3.83) | (4.62) | (4.72) | (4.53) | (4.71) | (4.74) | (4.70) |

Table continues on next page.

| | AR1 | BMS | BMA | EW | DMS | DMS$^{C_1}$ | DMA |
|---|---|---|---|---|---|---|---|
| Massachusetts | **0.79** | **1.02** | **0.95** | **0.62** | **0.77** | **0.72** | **0.70** |
| | (4.01) | (4.39) | (4.55) | (4.95) | (4.65) | (4.50) | (4.50) |
| Michigan | 1.05 | **0.80** | **0.82** | **0.75** | **0.71** | **0.72** | **0.75** |
| | (0.62) | (3.35) | (3.11) | (2.78) | (3.39) | (3.53) | (3.57) |
| Minnesota | 1.00 | **0.65** | **0.68** | **0.74** | **0.66** | **0.65** | **0.65** |
| | (0.60) | (3.12) | (3.07) | (2.80) | (3.20) | (3.20) | (3.14) |
| Mississippi | 1.01 | **0.99** | **0.99** | 1.34 | **0.88** | **0.83** | **0.84** |
| | (−0.95) | (1.66) | (1.67) | (−0.39) | (3.00) | (3.10) | (3.10) |
| Missouri | 1.00 | **0.86** | **0.83** | **0.83** | **0.75** | **0.80** | **0.81** |
| | (−0.14) | (3.21) | (3.07) | (2.27) | (2.96) | (2.94) | (3.00) |
| Montana | 0.98 | 1.15 | 1.13 | **0.84** | **0.78** | **0.78** | **0.80** |
| | (1.63) | (0.15) | (0.08) | (1.95) | (2.96) | (3.19) | (3.19) |
| Nebraska | 1.00 | **0.90** | **0.90** | **0.91** | **0.92** | **0.93** | **0.92** |
| | (−0.32) | (2.51) | (2.43) | (2.72) | (2.30) | (1.93) | (2.13) |
| Nevada | **0.80** | **0.68** | **0.62** | **0.70** | **0.50** | **0.57** | **0.58** |
| | (2.99) | (2.46) | (3.13) | (3.20) | (2.94) | (3.26) | (3.30) |
| New Hampshire | **0.85** | **1.00** | **0.97** | **0.67** | **0.74** | **0.75** | **0.72** |
| | (3.86) | (3.61) | (3.61) | (3.61) | (4.29) | (4.16) | (4.10) |
| New Jersey | **0.74** | **0.91** | **0.76** | **0.55** | **0.63** | **0.60** | **0.60** |
| | (5.12) | (4.60) | (4.75) | (5.04) | (4.80) | (4.82) | (4.82) |
| New Mexico | **0.92** | **0.93** | **0.93** | **0.81** | **0.75** | **0.74** | **0.77** |
| | (2.00) | (1.94) | (1.92) | (2.95) | (3.20) | (3.21) | (3.17) |
| New York | **0.95** | **0.98** | **0.93** | **0.65** | **0.71** | **0.72** | **0.72** |
| | (3.01) | (3.71) | (3.74) | (3.62) | (4.69) | (4.15) | (4.12) |
| North Carolina | 1.00 | 1.06 | **0.88** | **0.80** | **0.66** | **0.65** | **0.66** |
| | (0.51) | (0.28) | (1.92) | (2.08) | (2.57) | (2.60) | (2.59) |
| North Dakota | 1.16 | **0.73** | **0.73** | **1.22** | **0.78** | **0.80** | **0.78** |
| | (−3.41) | (5.31) | (5.31) | (1.68) | (4.61) | (4.36) | (4.69) |
| Ohio | 1.12 | 1.12 | **1.00** | **0.91** | **0.96** | **0.93** | **0.95** |
| | (−0.50) | (0.96) | (2.09) | (2.13) | (1.81) | (1.73) | (1.75) |
| Oklahoma | 1.03 | **0.79** | **0.82** | **0.83** | **1.02** | **0.90** | **0.85** |
| | (0.69) | (5.01) | (4.73) | (3.75) | (2.66) | (3.56) | (3.96) |
| Oregon | **0.84** | **0.81** | **0.78** | **0.80** | **0.71** | **0.71** | **0.73** |
| | (3.54) | (3.85) | (3.71) | (2.20) | (3.54) | (3.64) | (3.75) |
| Pennsylvania | 0.98 | **0.94** | **0.87** | **0.68** | **0.80** | **0.73** | **0.74** |
| | (1.14) | (2.92) | (3.23) | (3.82) | (3.81) | (3.90) | (3.87) |
| Rhode Island | **0.84** | **0.96** | **0.83** | **0.66** | **0.63** | **0.64** | **0.62** |
| | (3.91) | (4.89) | (5.01) | (4.29) | (4.89) | (5.00) | (4.96) |
| South Carolina | 1.00 | 0.94 | **0.89** | **0.79** | **0.75** | **0.69** | **0.71** |
| | (0.24) | (1.60) | (2.73) | (2.63) | (2.43) | (2.78) | (2.81) |

| | AR1 | BMS | BMA | EW | DMS | DMS$^{C_1}$ | DMA |
|---|---|---|---|---|---|---|---|
| South Dakota | 1.01 | **1.02** | **1.01** | 1.66 | **0.84** | **0.90** | **0.90** |
| | (−0.61) | (2.48) | (2.64) | (0.54) | (3.79) | (3.35) | (3.38) |
| Tennessee | 1.01 | 1.14 | 1.11 | 0.93 | **0.99** | **0.86** | **0.88** |
| | (0.25) | (−0.32) | (−0.12) | (1.54) | (1.98) | (2.24) | (2.23) |
| Texas | 0.99 | **0.74** | **0.87** | **0.78** | **0.76** | **0.76** | **0.76** |
| | (0.54) | (4.23) | (3.27) | (3.37) | (3.95) | (3.96) | (3.95) |
| Utah | **0.90** | **1.03** | **1.08** | **0.81** | **0.99** | **0.95** | **0.96** |
| | (2.62) | (2.40) | (2.71) | (3.00) | (3.14) | (3.01) | (3.05) |
| Vermont | 1.06 | **0.75** | **0.76** | 1.26 | **0.81** | **0.73** | **0.73** |
| | (−3.03) | (4.24) | (4.18) | (1.18) | (4.18) | (4.45) | (4.49) |
| Virginia | **0.80** | **0.82** | **0.76** | **0.62** | **0.79** | **0.66** | **0.63** |
| | (3.05) | (4.23) | (4.31) | (4.78) | (4.69) | (4.65) | (4.59) |
| Washington | **0.81** | **0.91** | **0.80** | **0.63** | **0.82** | **0.68** | **0.69** |
| | (3.50) | (3.09) | (3.25) | (2.70) | (3.29) | (3.27) | (3.28) |
| West Virginia | 1.02 | **0.87** | **0.87** | **1.02** | **0.82** | **0.76** | **0.77** |
| | (−1.03) | (3.05) | (3.02) | (3.14) | (4.10) | (4.95) | (4.66) |
| Wisconsin | 1.00 | **0.93** | **0.90** | **0.75** | **0.91** | **0.79** | **0.79** |
| | (−0.25) | (2.38) | (2.70) | (3.79) | (2.50) | (3.23) | (3.47) |
| Wyoming | **0.82** | **0.63** | **0.62** | **0.69** | **0.68** | **0.61** | **0.62** |
| | (4.05) | (3.92) | (4.28) | (4.07) | (4.30) | (4.77) | (4.75) |
| U.S. | **0.93** | **0.71** | **0.71** | **0.68** | **0.58** | **0.65** | **0.67** |
| | (1.76) | (3.40) | (3.42) | (3.15) | (3.37) | (3.36) | (3.33) |

The table reports MSFE ratios of various forecasting methods relative to the MEAN benchmark. Below the MSFE ratios are $t$-statistics from the Clark and West (2007) test in parenthesis. We use Newey-West heteroskedasticity and autocorrelation consistent standard errors. The out-of-sample window is 1995:1 to 2012:4.

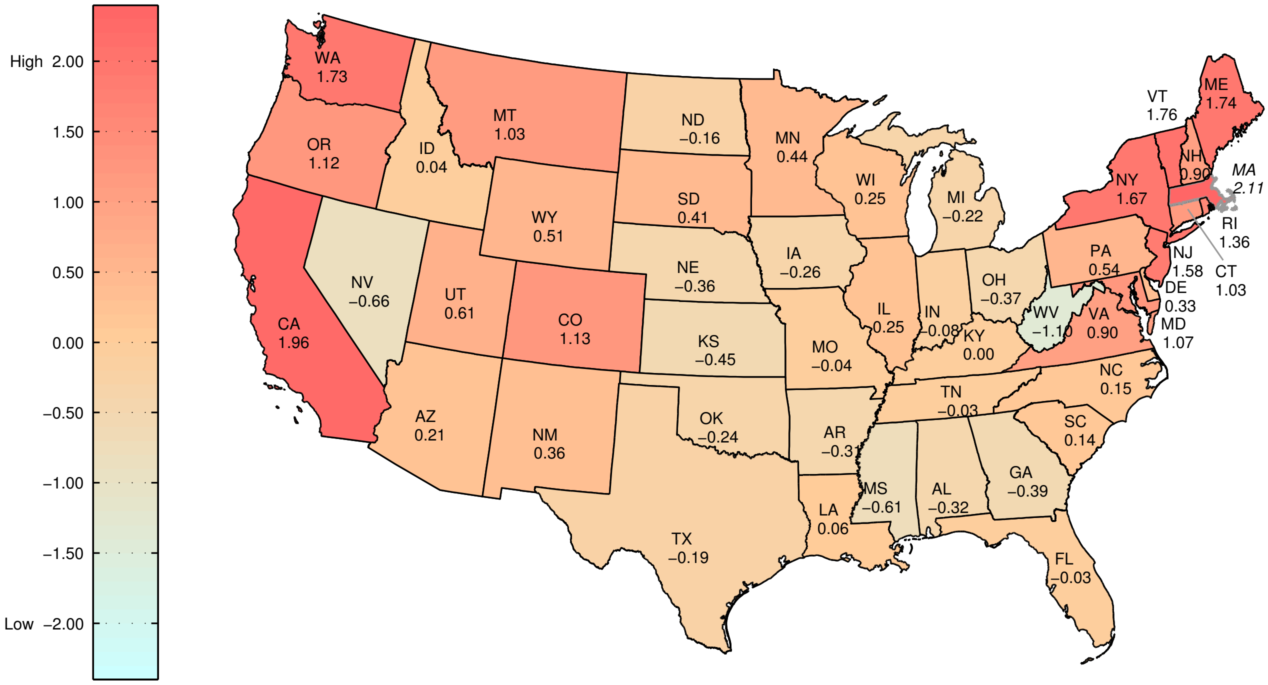


Fig. 1. Mean real growth rates, 1976:2-2012:4. The growth rates are annualized and inflation-adjusted.

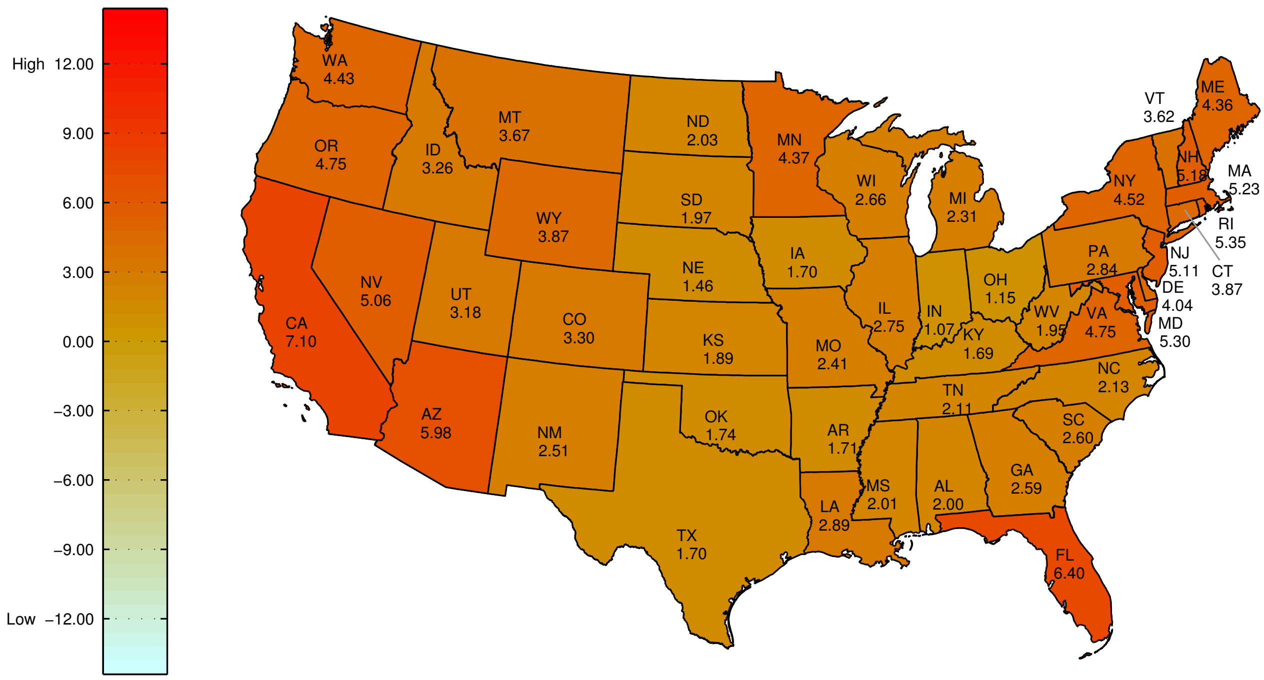


Fig. 2. Mean real growth rates during the boom period, 1995:1-2006:4. The growth rates are annualized and inflation-adjusted.

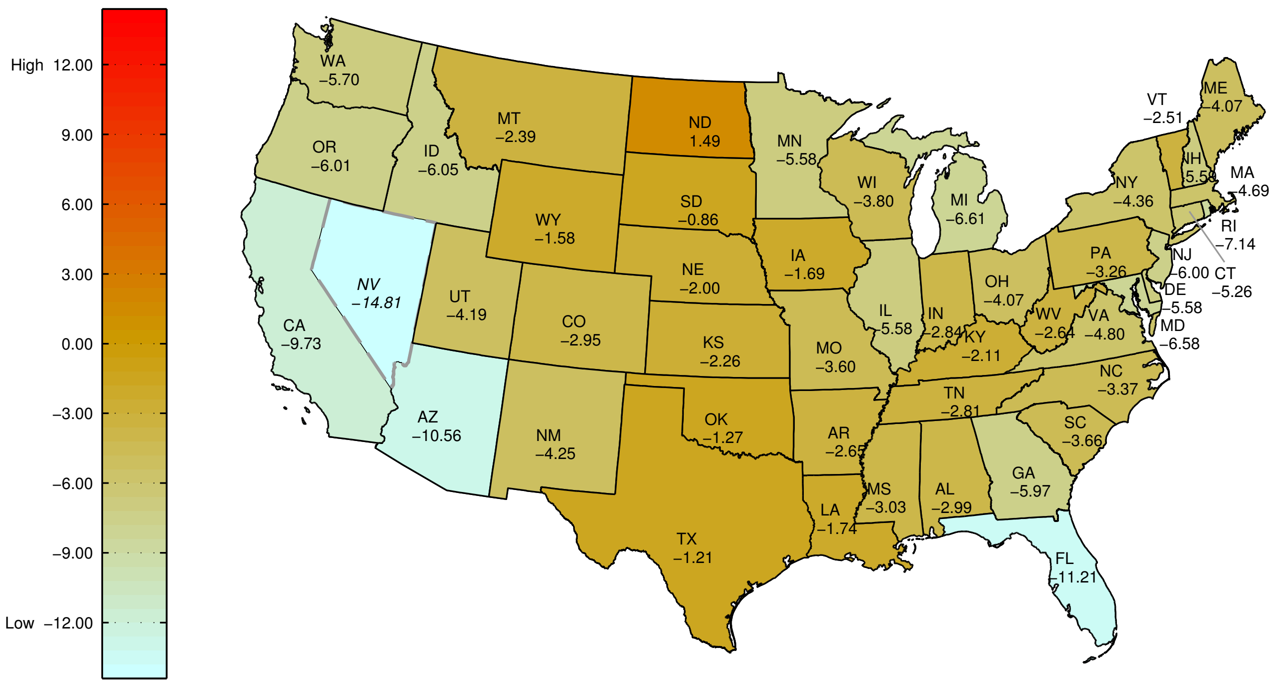


Fig. 3. Mean real growth rates during the bust period, 2007:1-2012:4. The growth rates are annualized and inflation-adjusted. Numbers beyond the left-hand side scale are indicated by italic font face.

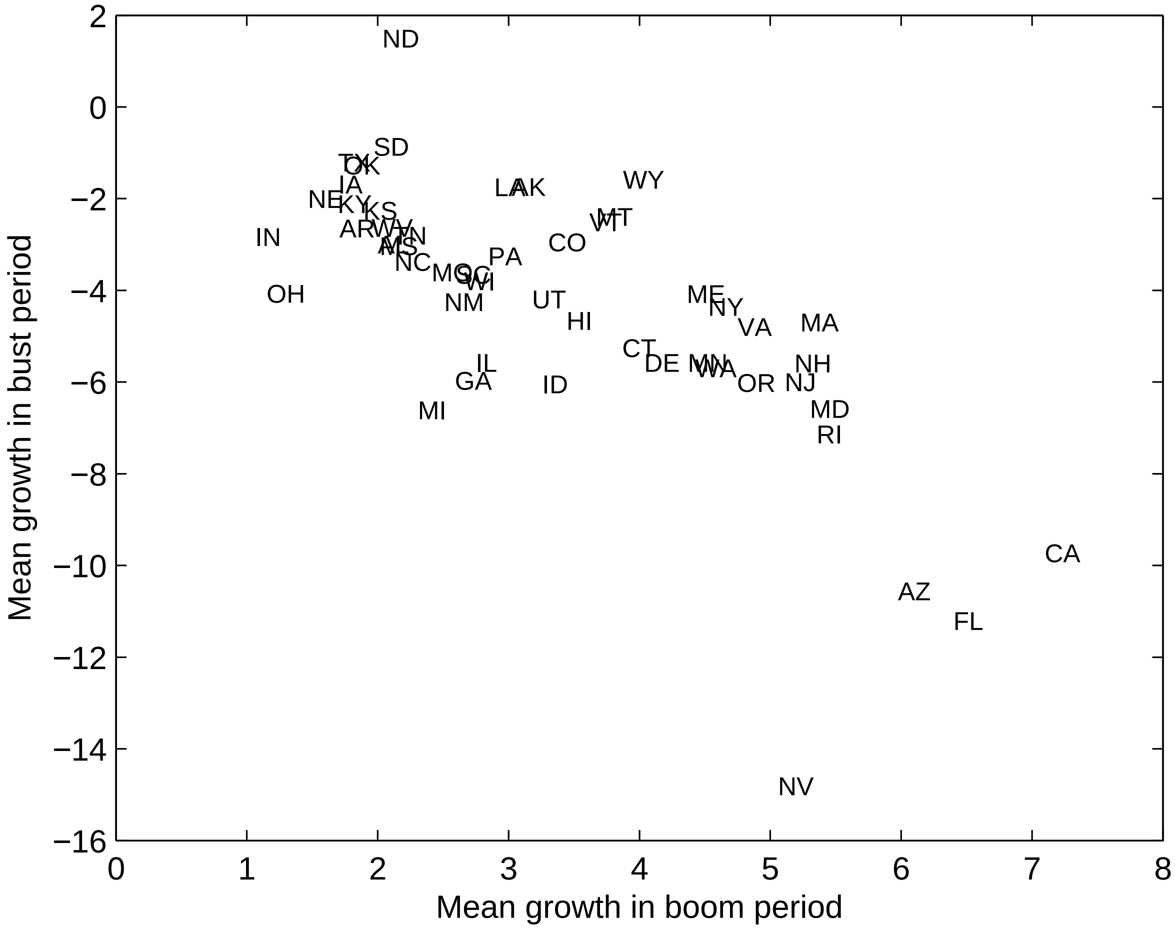


Fig. 4. Mean growth in bust period against mean growth in boom period. The boom period is assumed to be 1995:1-2006:4, while the bust period is assumed to be 2007:1-2012.4. The growth rates are annualized and inflation-adjusted.

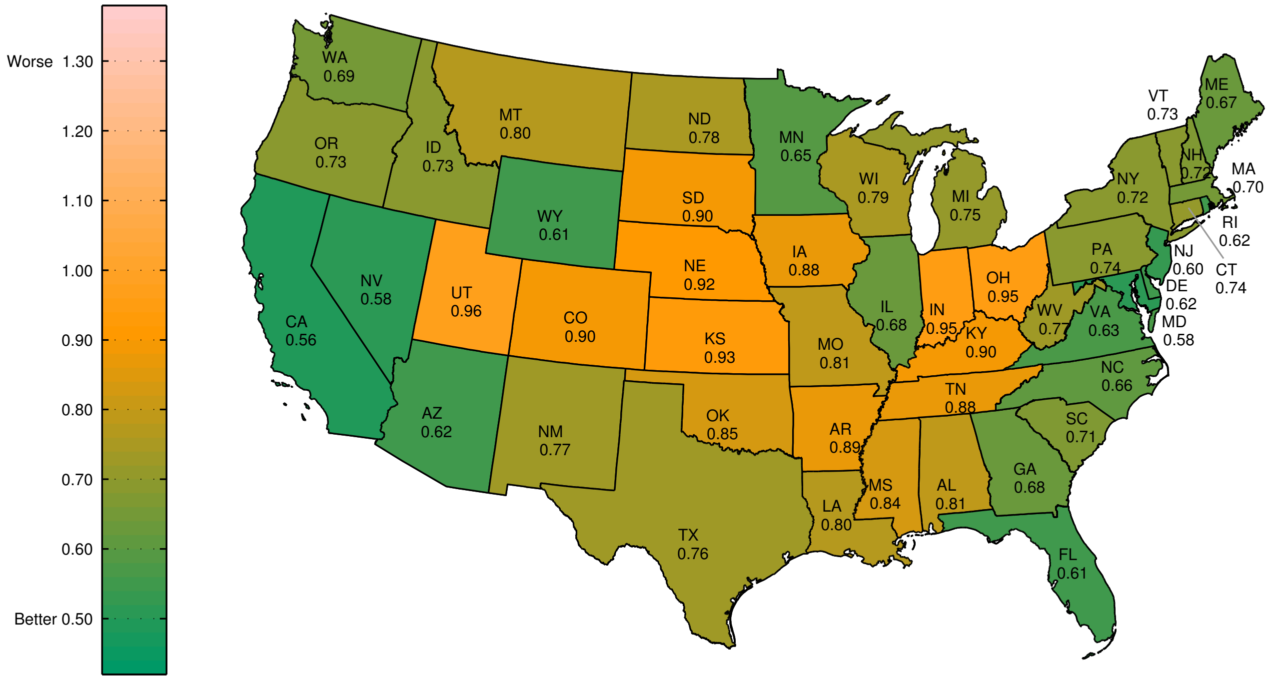


Fig. 5. DMA vs. historical mean. The figure maps the MSFE ratio of DMA relative to the historical mean. The evaluation period is 1995:1-2012:4.

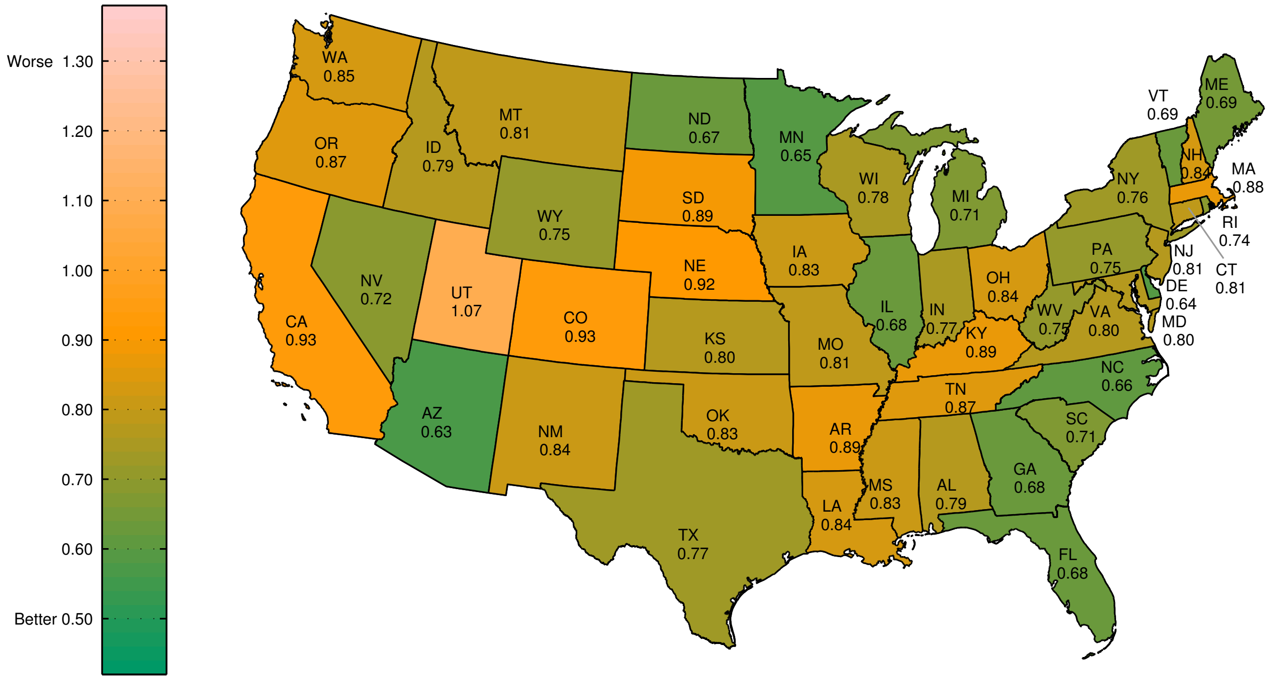


Fig. 6. DMA vs. AR1 model. The figure maps the MSFE ratio of DMA relative to the AR1 model. The evaluation period is 1995:1-2012:4.

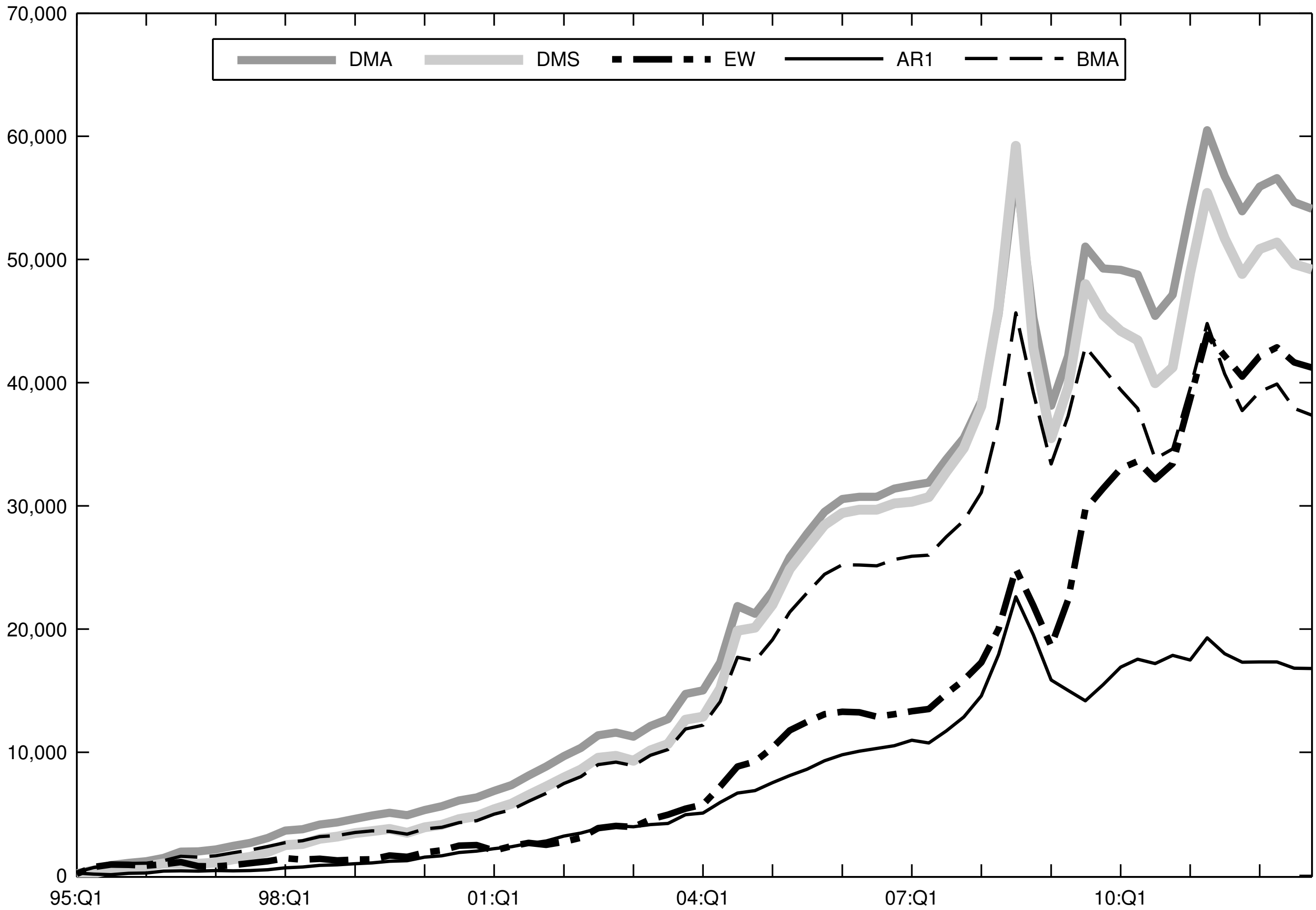


Fig. 7. Cumulative squared forecast error difference. The evaluation period is 1995:1-2012:4.

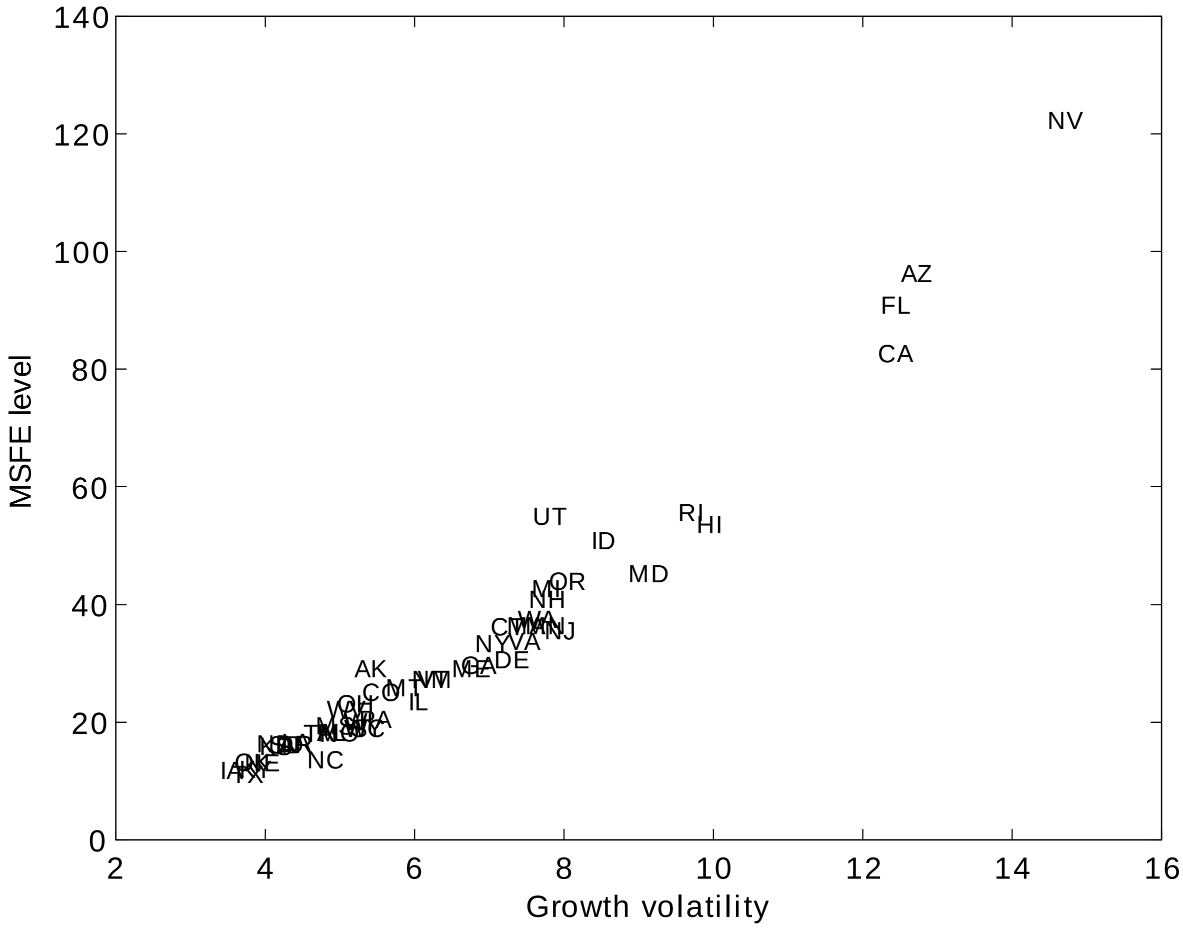


Fig. 8. MSFE level vs. volatility. The figure plots the MSFE level against the volatility in real growth rates, 1995:1-2012:4.

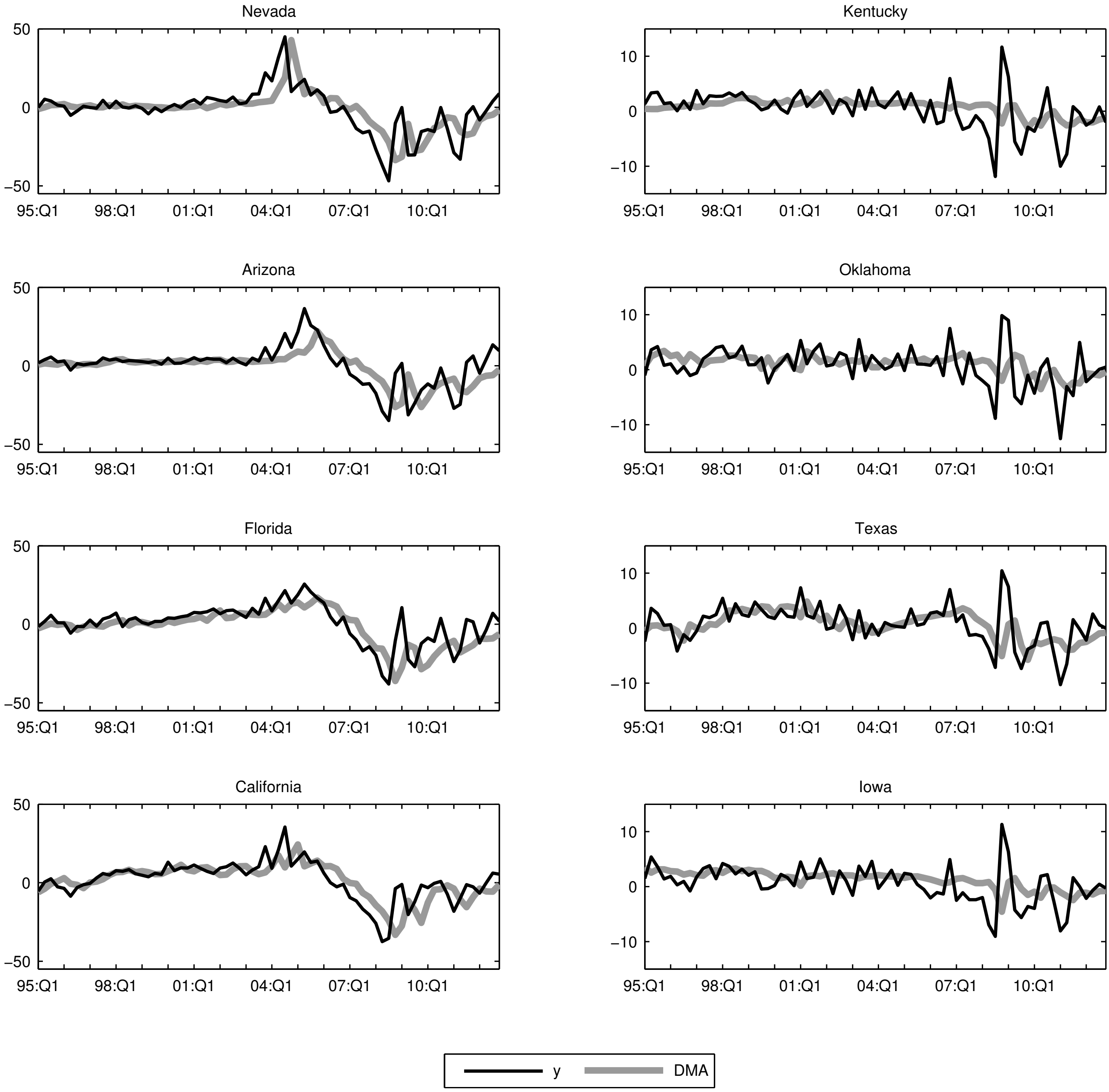


Fig. 9. Forecasted vs. realized house price changes: Top four volatility states (left) vs. bottom four (right; different scale). The evaluation period is 1995:1-2012:4.

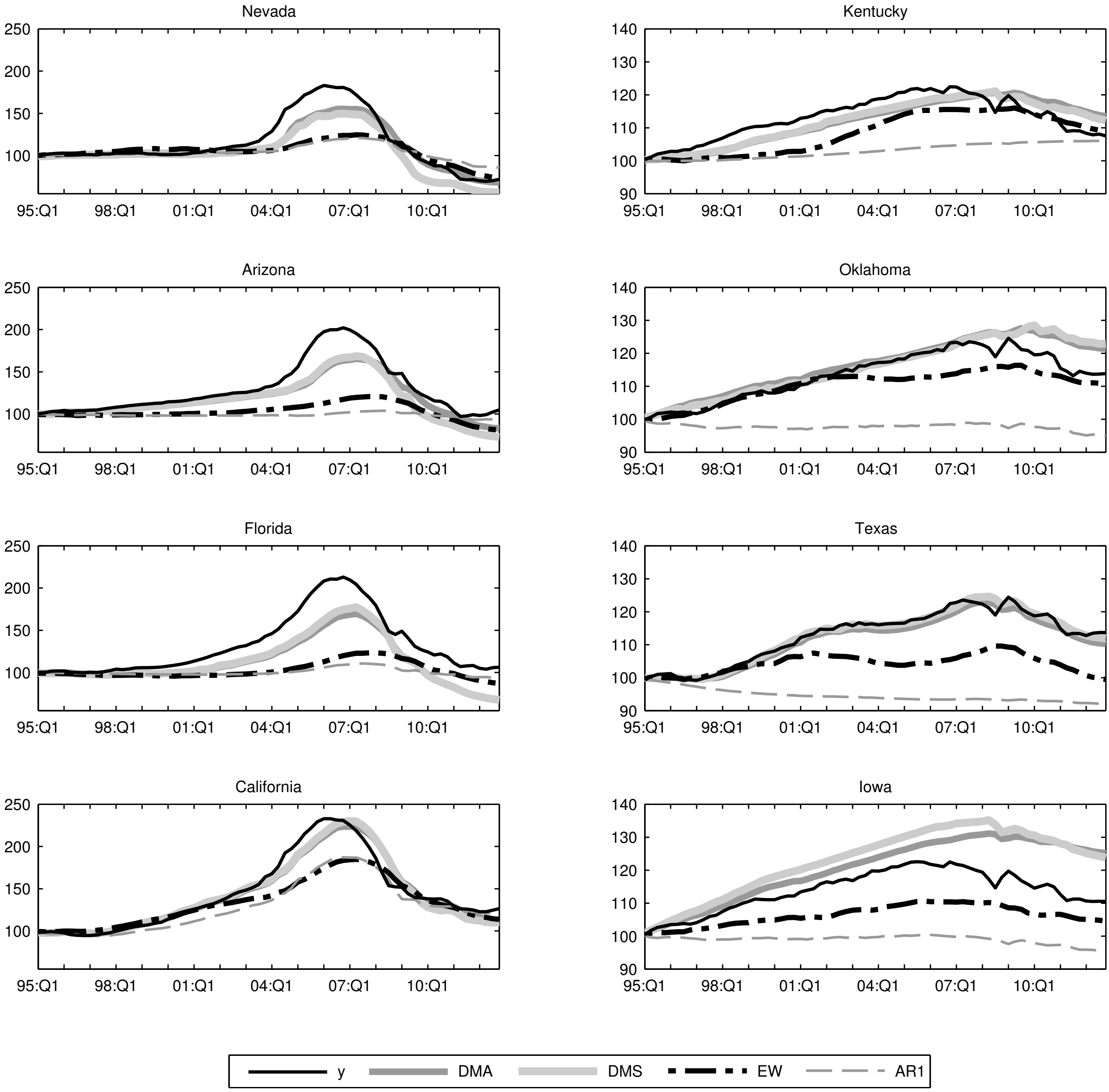


Fig. 10. Forecasted vs. realized levels of house prices: Top four volatility states (left) vs. bottom four (right; different scale). The evaluation period is 1995:1-2012:4.

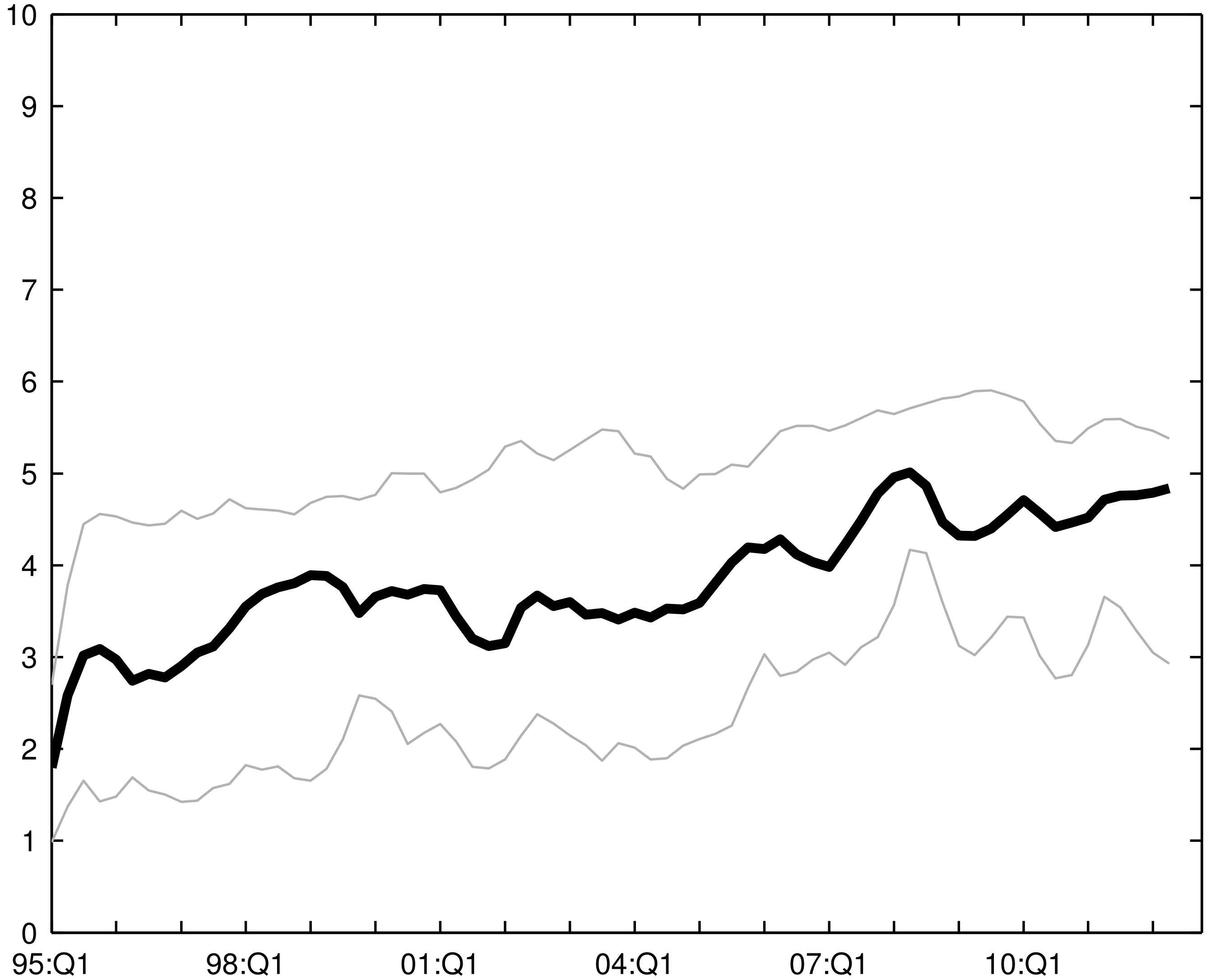


Fig. 11. Median of expected model size.

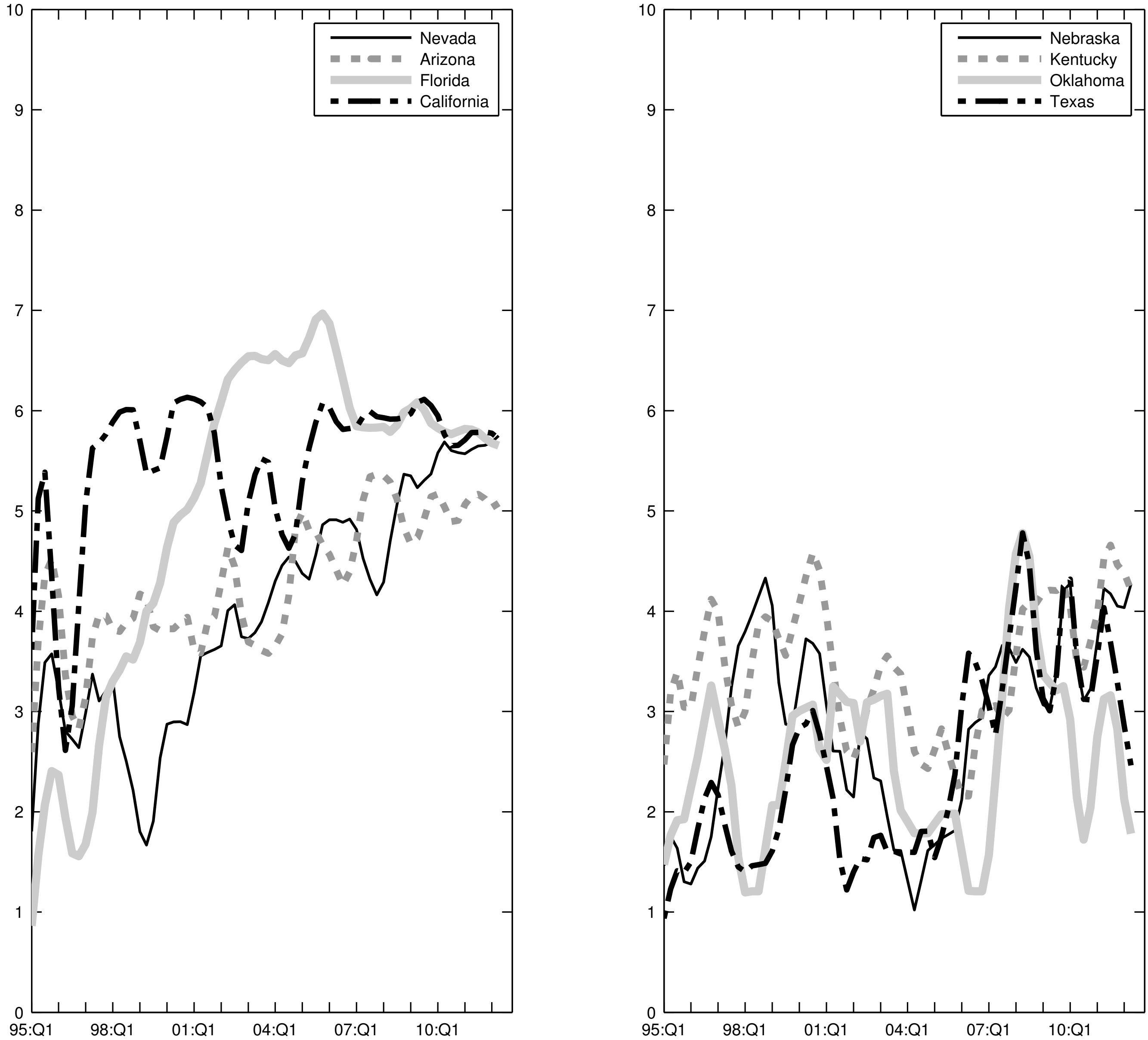


Fig. 12. Expected model size: Top four volatility states (left) versus bottom four (right)

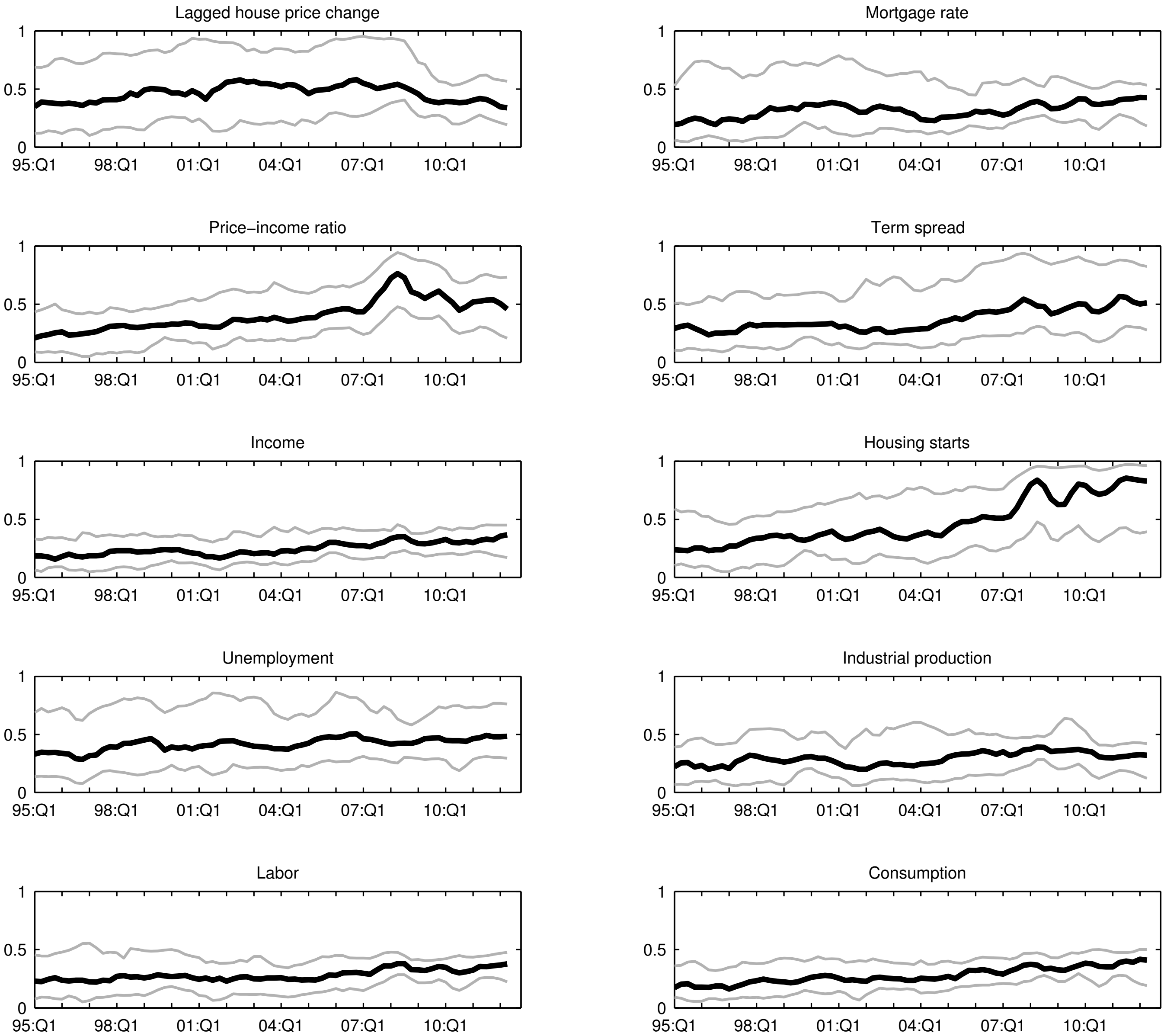


Fig. 13. Median of posterior inclusion probabilities across states

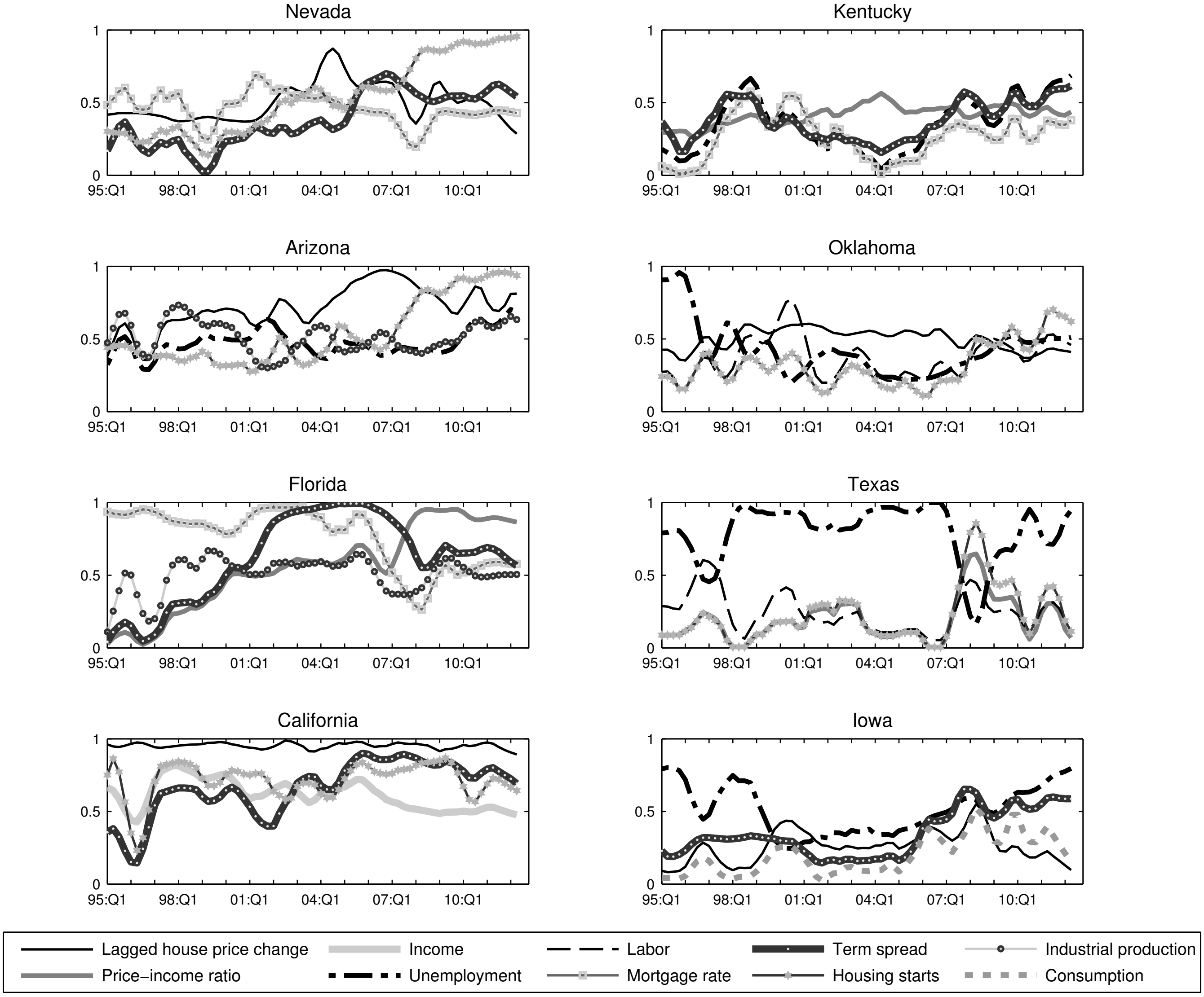


Fig. 14. Posterior inclusion probabilities of the four most important variables: Top four volatility states (left) versus bottom four (right)