Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2020)
183, *Part* 3, *pp.* 801–827

A similarity-based approach for macroeconomic forecasting

Y. Dendramis,

Athens University of Economics and Business, Greece, and University of Cyprus, Nicosia, Cyprus

G. Kapetanios

King's College London, UK

and M. Marcellino

Bocconi University, Milan, Italy

[Received March 2019. Final revision March 2020]

Summary. In the aftermath of the recent financial crisis there has been considerable focus on methods for predicting macroeconomic variables when their behaviour is subject to abrupt changes, associated for example with crisis periods. We propose similarity-based approaches as a way to handle parameter instability and apply them to macroeconomic forecasting. The rationale is that clusters of past data that match the current economic conditions can be more informative for forecasting than the entire past behaviour of the variable of interest. We apply our methods to predict both simulated data in a set of Monte Carlo experiments, and a broad set of key US macroeconomic indicators. The forecast evaluation exercises indicate that similarity-based approaches perform well, in general, in comparison with other common time-varying forecasting methods, and particularly well during crisis episodes.

Keywords: Empirical similarity; Forecast comparison; Kernel estimation; Macroeconomic forecasting; Parameter time variation

1. Introduction

The Great Recession in 2007–2009 was the major recent economic event that stressed the deficiency of macroeconomists to provide reliable forecasts in periods of turmoil. The poor performance of the economic models was mainly driven by their weakness to predict abrupt changes in economic series that are usually observed around crisis periods (see, for example, Ferrara *et al.* (2015) and Potter (2011) *inter alia*).

In the macroeconomic and financial literature, it is well known that forecasting is affected significantly by parameter instability (see, for example, Clements and Hendry (1998), Hendry (2000), Stock and Watson (1996) and Pesaran *et al.* (2011) *inter alia*). For instance, assuming wrongly a fixed model structure will result in inconsistent parameter estimates and, probably, major forecast failures. Various methods have been developed to incorporate parameter changes in econometric models. A common approach is to assume that parameters change continuously over time. In this case, the breaks are generally of smaller size and occur at ev-

Address for correspondence: Y. Dendramis, Department of Economics, Athens University of Economics and Business, 76 Patission Street, Athens 10434, Greece.
E-mail: ydendramis@aueb.gr

0964–1998/20/183801

ery period, resulting in a slowly changing parameter vector. In particular, random-coefficient models assume parameters that evolve stochastically over time, typically as persistent stochastic processes. These specifications are commonly adopted in the macroeconometric modelling and forecasting literature (see, for example, Nicholls and Pagan (1985), Cogley and Sargent (2005) and Primicery (2005)).

An alternative approach assumes that changes occur rarely and are abrupt. A key reference in this context is Hamilton (1989), whose model with Markov switching (MS) coefficients has later been extended in various directions. Kapetanios and Tzavalis (2010) and Dendramis *et al.* (2015) have provided a related avenue for modelling structural breaks in the level or volatility of economic series.

Both random-coefficient and MS models are characterized by unobservable parameter changes. In threshold and smooth transition models, the parameter evolution is, instead, driven by observable characteristics; see for example Tong (1990) and Terasvirta (1998). Ghysels and Marcellino (2018) have provided an excellent review on forecasting approaches in the presence of breaks.

In this paper we propose to handle structural change and parameter instability in a different manner. Drawing from notions of similarity and learning, we allow the time-varying parameter (TVP) vector to be driven by one or more observable variables, which are collected in vector z_t. This is in line with the threshold and smooth transition approaches, which assume a parametric specification for linking the z_t and the TVP β_t. In our framework, the z_t is the 'trigger' variable that drives the coefficient vector β_t through a kernel weighting scheme, allowing substantial flexibility in the patterns of parameter time variation. This permits rather simple yet flexible local estimation of models with auto-regressive (AR) and moving average (MA) components, which is generally known to perform remarkably in out-of-sample forecasting. Our approach works as follows: through the observables z_t, we identify past periods that are similar to the present in terms of economic conditions and exploit these specific periods to learn about the current model parameters β_t. Hence, parameters are estimated by data that are more similar (or relevant) to the current economic regime.

Our approach can be directly linked to nearest neighbourhood (NN) techniques and kernel-based non-parametric regression. It can nest NNs, for specific choices of weighting scheme, trigger variable z_t and set of regressors x_t. It is also related to locally weighted regression (LWR), which is a generalization of the NN technique that was proposed by Cleveland (1979), and refined by Cleveland and Devlin (1988). Diebold and Nason (1990) studied extensively the forecasting performance of this approach for 10 major dollar spot rates in the post-1973 float, finding no out-of-sample forecasting benefits from LWR. LWR relates the weighting scheme to the realizations of an observable variable included in the set of regressors, whereas our trigger variable z_t is not restricted to this set. Moreover, the computational burden that is required by LWR obliged the authors to consider a constant tuning parameter that accounts for the neighbourhood on which the forecasting equation is estimated, whereas we allow for time-varying tuning.

Machine learning methods such as neural networks and random forests (see Breiman (2001), *inter alia*) can be seen, in some sense, as extensions of the NN approach. For instance, the random-forests method is an ensemble of fully grown regression trees estimated on different bootstrap subsamples of the original data. A regression tree forecasts a dependent variable by splitting the space that is spanned by the covariates into a significant number of regions. In each region, the forecast of a dependent variable y_t is defined as its local average. Then, the out-of-sample prediction of the dependent (target) variable depends on the prevailing regime, as summarized by the observed regressors. Similarly, neural networks, with their ability to approximate arbitrary unknown functions, are important alternatives to non-parametric regression, especially when extended to 'deep' multilayer architectures. Further, recent successful architectures such as long short-term memory networks can 'remember' relevant events from the distant past, in an analogous fashion to our approach. The success of such machine learning methods in forecasting inflation has been documented recently by Medeiros *et al.* (2019).

There is a limited related literature on similarity-based forecasting. Guerron-Quintana and Zhong (2017) used clustering techniques to identify similar economic periods, which are then fed to auto-regressive integrated moving average models. In an additional step, they proposed to adjust the forecasts by adding an error term that is constructed from matched blocks of data. Guerron-Quintana and Zhong (2017) also considered a combination of NN models that have previously performed well instead of selecting a single parameterization. Overall, the algorithms proposed work sufficiently well in recessions, compared with standard auto-regressive integrated moving average models, but the theoretical rationale is not fully specified. Gilboa *et al.* (2011) proposed a related approach that combines the notion of similarity with non-parametric regression. Yet, they focused on the theoretical axiomatization of their proposals, without presenting a comprehensive econometric methodology or an empirical application. Pesaran *et al.* (2013) also associated the evolution of the parameter vector β_t with that of observables. They derived theoretically optimal weighting schemes of past observations under specific assumptions (such as known size and timing) on the break process of the parameter vector, for one-step-ahead forecasting. In a related paper, Eklund *et al.* (2010) considered two groups of forecasting strategies. In the first, the forecaster monitors the happening of a change and adjusts the forecasting method once a change has been detected. In the second strategy, the forecaster does not attempt to identify breaks and uses instead break robust forecasting strategies that essentially downweight data from past periods. Although moving in an interesting direction, Eklund *et al.* (2010) did not elaborate on the extent and shape of the downweighting of past data. Clearly, both issues affect the forecasting performance of the model. Monotonic discounting has been extensively studied by Giraitis *et al.* (2013), whereas our proposed similarity-based forecasting approach can account for non-monotonicity of past data discounting. The rationale and importance of this is straightforward: if economic regimes come and go, then data from periods that are similar to the current period are more suitable for efficient forecasting, rather than the more recent data only. In our approach, the trigger variable z_t governs the evolution of the TVP estimates, indicating and exploiting periods with similar economic characteristics.

Our proposed similarity-based, kernel-driven, TVP estimator can be viewed as a form of local linear regression. For this, standard theoretical results on consistency, rates and asymptotic normality, such as those provided in Pagan and Ullah (1999) and Robinson (1983), easily apply. Yet, our approach depends crucially on the choice of the bandwidth parameter. We provide empirical and simulation evidence that supports the appropriateness of cross-validation as a tool for calibrating the bandwidth parameter. Moreover, although we focus on univariate time-varying auto-regressive moving average (ARMA) models, analogous methods can be applied to account for time variation in general univariate regression models or multivariate vector AR-type models, or factor models.

To assess empirically the forecasting performance of our similarity-based methods relative to either stable models or a variety of common time-varying models, we focus on a set of key monthly US macroeconomic and financial variables. These include payments, unemployment, earnings, real personal income, industrial production, capacity utilization, housing starts, federal funds rate, 3-month interest rate, money stock, consumer credit, the consumer price index CPI and producer price index PPI. In terms of the trigger variable for our similarity approaches, we consider two main possibilities. First, following proposals in threshold and smooth transition models, we use a smooth transformation of the target variable that needs to be forecasted.

Second, we explore alternative macroeconomic indicators whose behaviour could affect the dynamics of the target variable of interest: oil prices as a measure of external shocks, the federal funds rate as a measure of the monetary policy stance and the housing starts index, as a leading indicator of economic conditions. In the on-line appendix, we extend the set of trigger variables to consider also summary indicators of financial and real conditions based on large information sets.

Overall, we find that the forecasting performance of stable AR(1) models can be improved by either adding more lags or an MA component, which both capture additional persistence that can be either real or due to unaccounted parameter changes. Evidence in favour of the latter option is provided by the overperformance of the TVPs in many cases. Within this class of models, our newly proposed similarity-based methods behave satisfactorily in a substantial number of cases, indicating the potential of this type of econometric modelling.

The paper is structured as follows. In Section 2, we introduce our similarity-based forecasting approach and discuss its theoretical properties and cross-validation schemes for choosing the tuning parameters, which are important for the empirical implementation of the approach. In Section 3, we briefly review alternative existing time-varying forecasting models and forecast comparison criteria. In Section 4, we conduct Monte Carlo experiments to assess the relative performance of our method in a controlled environment. In Section 5, we present the extensive empirical application related to forecasting US macroeconomic variables. In Section 6, we summarize the main results and conclude. Additional results are gathered in an on-line appendix.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`https://rss.onlinelibrary.wiley.com/hub/journal/1467985x/series-a-datasets.`

2. Similarity-based forecasting

In this section we present three model specifications that are associated with the notion of similarity and learning. In our first proposal, we extend the kernel-based, non-parametric regression model. The second draws on the threshold regression model and in our final proposal we modify appropriately the local averaging model. In these models, similar economic regimes are identified endogenously by the values of a trigger variable.

2.1. Trigger time-varying parameter model

We consider the following linear regression model for the dependent variable y_t:

$$y_t = x_{t-1}\beta_{t-1} + u_t, \qquad t = 1, \ldots, n, \quad u_t \sim \mathrm{IID}(0, \sigma_u^2), \tag{1}$$

where x_t is a $1 \times k$ vector of relevant covariates that may include an intercept, p lags of the dependent variable, q lags of the errors u_t, and/or other exogenous regressors. The forecast of y_{n+1} made in period n, denoted by $\hat{y}_{n+1}$, depends crucially on the estimate of the $k \times 1$ vector β_n, denoted by $\hat{\beta}_n$. It is

$$\hat{y}_{n+1} = x_n \hat{\beta}_n. \tag{2}$$

To estimate β_n, we adopt a non-parametric approach combined with the notion of similarity. Specifically, periods that match the current evolution of x_t affect significantly the parameter estimate, and vice versa for periods that are very different. To identify these periods, we relate the value of β_n to that of a trigger variable z_n, and we define the kernel estimator as

$$\hat{\beta}_n = \left(\sum_{l=1}^{n} k_{n_l,H} x'_{l-1} x_{l-1} \right)^{-1} \sum_{l=1}^{n} k_{n_l,H} x'_{l-1} y_l, \tag{3}$$

with the weights $k_{n_l,H} = K\{(z_n - z_l)/H\}$, where $K(x)$, $x \in \mathbb{R}$, is a continuous bounded function and H is the bandwidth parameter. In practice, $K(\cdot)$ is generally specified as a probability density function (e.g. a normal kernel). Other popular choices for $K(\cdot)$ include a rolling window kernel with $K(u) = I(0 \leqslant u \leqslant 1)$, and the exponential weighted moving average with $K(u) = \exp(-u)$, for $u \in [0, \infty)$. The parameter H, that governs the relative magnitude of the weight, is set equal to $H = \{\max(\{z_l\}_{l=1}^n) - \min(\{z_l\}_{l=1}^n)\}h$. The tuning parameter h controls the magnitude of the effect that the trigger variable z_t has on the parameter estimate. A sufficiently small H implies that, for periods l in which the trigger z_l is far from z_n (in the squared error sense in the case of a symmetric kernel), the kernel weight that is placed on the observation pair (x_{l-1}, y_l) is relatively small compared with periods where z_l is closer to z_n. For a sufficiently large bandwidth H (or equivalently h), the estimator $\hat{\beta}_n$ in equation (3) is very similar to the full sample ordinary least squares (OLS) estimator.

In case of an AR(p) model, the vector x_{t-1} includes lags of the dependent variable. To add an MA(q) component, the estimation procedure must be slightly modified. In the first step, the estimated errors $\hat{u}_t$ are derived from a long AR model AR(m) with m large. In the second step, $\hat{u}_{t-1}, \ldots, \hat{u}_{t-q-1}$ are included as covariates in x_{t-1}, and the resulting model is estimated by the estimator in equation (3). This two-step procedure is a direct extension of that often adopted for constant parameter ARMA models; see for example Dufour and Pelletier (2008).

It is important to note that the estimator in equation (3) nests other popular time-varying estimators proposed in the literature. In particular, when the trigger variable z_t equals the time index t, i.e. $z_t = t$, then the estimator in equation (3) becomes the time-varying kernel estimator that was developed by Giraitis *et al.* (2018). In this case, $\hat{\beta}_n$ is associated with a monotone weighting scheme of past data, whereas in our approach the weights depend on the behaviour of the trigger variable, which can clearly imply non-monotonic weighting schemes. As we shall see in the next section, model (1), accompanied by the estimator in equation (3), also nests other popular approaches in the literature, like exponential smoothing and threshold regressions.

In practice, the trigger z_t can be any variable that is informative about the likely evolution of the β_t-parameters. Past values of the dependent variable are a candidate. For example, the dynamics of inflation or growth can depend on whether these variables are high or low. Economic theory can also provide some clues about the choice of z_t. For instance, the stance and extent of fiscal or monetary policy can influence the magnitude of the effects of changes of foreign variables (or shocks) on the domestic indicators.

In the following sections, we shall refer to model (1) estimated by equation (3) as the trigger TVP model tv-trig. The special case where the trigger variable z_t is the time dimension t will be referred to as model tv. For both types of model (tv and tv-trig) we shall use a normal kernel experiment with two types of cross-validation methods (described below) for the bandwidth selection and consider specifications with and without an MA component.

In our presentation the notion of similarity is defined in terms of the univariate trigger variable z_t and the symmetric kernel weighting scheme. This can be clearly extended to cases in which z_t is a vector that includes the most recent observations, capturing more information about the prevailing regime. Moreover, one can allow for asymmetric weighting schemes. This can be done, for instance, by skewed versions of the kernel function (e.g. the skewed normal distribution).

Although these are important extensions, the flexibility that these imply comes at the cost of increasing the computational complexity.

2.2. Similarity local averaging model

Giraitis *et al.* (2013) have exhaustively analysed the properties of a local averaging model. We now introduce a similarity-based extension of this model and discuss its use for economic prediction.

Following Giraitis *et al.* (2013), we assume that data are generated by a modified local averaging model. Now, locality is defined in terms of similarity of the current dependent variable with the past observations, rather than the time index variable t. In this specification, the trigger variable coincides with y_t. For this, the historical values of y_t that are more similar to its current value are weighted proportionally more than the less similar past values. Formally, we consider the model

$$y_t = \sum_{i=1}^{t-1} w_{i,\rho}\tilde{y}_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{IID}(0, \sigma_\varepsilon^2), \tag{4}$$

where $w_{i,\rho} = \rho^i / \Sigma_{j=1}^t \rho^j$, ρ is the tuning parameter with $0 < \rho < 1$, $\tilde{y}_t = y_t$ and $\{\tilde{y}_{t+1-i}\}_{i=2}^t$ are the ordered most similar observations to y_t (in the squared error sense), in the set $\{y_j\}_{j=1}^{t-1}$, with $\tilde{y}_{t-1}$ being the most similar to y_t (lowest squared error $(y_t - \tilde{y}_{t-1})^2$) and $\tilde{y}_1$ the least similar to y_t (highest squared error $(y_t - \tilde{y}_1)^2$). The tuning parameter ρ controls the relative magnitude of the weights w_i which can be calibrated by cross-validation, as we shall also discuss in the next section. For a given choice of ρ equal to $\hat{\rho}$, the forecast of y_{n+1} is

$$\hat{y}_{n+1} = \sum_{i=1}^{n} w_{i,\rho}\tilde{y}_{n+1-i}. \tag{5}$$

It is worth emphasizing the direct association of model (4) with our basic proposal given by equations (3) and (1). Accordingly, the trigger variable z_t in equation (3), which identifies the similar regimes, is now replaced by the lagged dependent variable y_{t-1}, indicating the most important observations in the history of y_t. Model (4) can be also interpreted as a similarity-based version of exponential smoothing. Additionally, when the kernel $k_{n_j,H}$ is set equal to $K(x) = \exp(-x)$, then equation (3) coincides with the exponentially weighted moving average model. This holds for $\rho = \exp(-1/H)$. In both cases, the weights could be made dependent on z_t also. Finally, equation (5) also coincides with an NN forecast, when using y_t as trigger variable.

In the following sections, we shall refer to model (4) as the similarity local averaging (SLA) model.

2.3. Similarity-based threshold regression

The SLA model that was presented in the previous section is solely based on the variable y_t, neglecting the information that might be provided by the covariates x_t. A related approach that can also account for the covariates is the threshold regression model (see for example Gonzalo and Pitarakis (2002)), which has an immediate similarity-based interpretation:

$$y_t = \tilde{x}_{t-1}\gamma + v_t, \qquad t = 1, \ldots, n, \quad v_t \sim \mathrm{IID}(0, \sigma_v^2), \tag{6}$$

where γ is a $2k \times 1$ vector of parameters, $\tilde{x}_t$ is a $1 \times 2k$ vector of explanatory variables defined as $\tilde{x}_t = (x_t I(z_t < \lambda_p), x_t\{1 - I(z_t < \lambda_p)\})$ and $I(A)$ is an indicator (dummy) variable, whose value is equal to 1 when the event A is true and 0 otherwise. The event A depends on the trigger variable z_t and the threshold parameter λ_p. Practically, this means that the effects of x_t are allowed to change depending on whether $z_t < \lambda_p$ (and consequently $I(z_t < \lambda_p) = 1$), or $z_t \geqslant \lambda_p$ (and $I(z_t \geqslant \lambda_p) = 1$). The value of the threshold parameter λ_p is essential, as it indicates the value of z_t that triggers the regime change. Empirically, it is sensible to assume that λ_p is the pth quantile of the filtered empirical distribution of z_t. In practice, the optimal λ_p can be also calibrated by using past data, through a cross-validation procedure. Given a value for λ_p, model (6) is easily estimated by OLS, and forecasts of y_{n+1} are given by

$$\hat{y}_{n+1} = \tilde{x}_n \hat{\gamma}_{\text{ols}}.$$

Model (6) can be also considered as a special case of our basic model proposal. In particular for a flat kernel of the form $K(x) = \{1 \text{ when } x < \lambda_p \text{ and } 0 \text{ otherwise}\}$, model (1) with estimator (3) nests the threshold regression model (6). Moreover model (6) can be also interpreted as a locally weighted regression.

In the following sections, we shall refer to model (6) as similarity-based threshold regression.

2.4. Theoretical considerations

The above models can be viewed within the context of non-parametric regression. If we assume a deterministic form for β_t, and stationarity and mixing conditions on x_t and u_t, then the theoretical properties of estimators such as estimator (3) follow readily from existing work. It is important to note that the work of Giraitis *et al.* (2013) relates to structural change, thus making the imposition of stationarity and mixing assumptions suspect, and allows for stochastic β_t, therefore requiring a separate theoretical analysis.

In our case, equation (3) can be viewed as a form of local linear regression estimator and standard theoretical results on consistency, rates and asymptotic normality, such as those provided in Pagan and Ullah (1999) and Robinson (1983), easily apply.

In particular, the following is a list of assumptions that are commonly encountered in the literature.

Assumption 1. $\beta(x)$ is a bounded and twice continuously differentiable function.

Assumption 2. (y_t, x_t, z_t) is a stationary α-mixing process with mixing coefficients α_k, such that $\Sigma_{k=n}^{\infty} \alpha_k^{1-2\theta} = O(n^{-1})$, $E|y_t|^{\theta} < \infty$, $E|x_t|^{\theta} < \infty$ and $E|z_t|^{\theta} < \infty$ for some $\theta > 2$.

Under these assumptions and also assuming technical regularity conditions relating to the kernel function $K(\cdot)$ that are satisfied by our set-up in the previous sections, we can obtain a host of standard useful asymptotic results for the estimates of β. Expressions for the bias and variance of the estimators can be obtained, and thereby rates and asymptotic normality can be established. In particular, a rate of order $(Th)^{1/2}$, for scalar x_t and z_t where the bandwidth h tends to 0, and asymptotic normality have been obtained in various references such as Robinson (1983) and Bierens (1987).

2.5. Selection of the tuning parameter

The similarity-based forecasting methods that were previously presented require the selection of some tuning parameters. Specifically, the trigger TVP model (3) depends on the bandwidth h, the similarity-based local averaging (4) on the parameter ρ and the similarity-based threshold regression (6) on λ_p. For all these choices we specify alternative cross-validation schemes, based on the mean-squared forecast error (MSFE).

Let δ denote a general tuning parameter, on which the one-step-ahead forecast $\hat{y}_{n+1|n,\delta}$ depends. In what we label as *end-of-sample cross-validation*, we calibrate δ by minimizing the MSFE over the last n_0 observations of the sample $\{y_t, z_t, x_t\}_{t=1}^n$:

$$\hat{\delta} = \arg\min_{\delta} \frac{1}{n_0} \sum_{t=n-n_0+1}^{n} (y_t - \hat{y}_{t|t-1,\delta})^2, \qquad \text{for } \delta \in [\delta_{\min}, \delta_{\max}]. \tag{7}$$

The parameter space $[\delta_{\min}, \delta_{\max}]$, over which the objective function in expression (7) is optimized, depends on the similarity-based approach at hand. For instance, in the trigger TVP model, we have that $\delta = h$. The upper bound of the parameter space of the bandwidth h, $h_{\max}$, is chosen to approximate the standard OLS estimator. The lower bound $h_{\min}$ is chosen such that when $h = h_{\min}$ a non-zero weight is attributed to a significant proportion of observations, to prevent computational issues. The actual values of $h_{\max}$ and $h_{\min}$ depend on the specific data at hand. In the similarity-based local averaging model, it is $\delta = \rho$, and $\rho_{\max} = 1$, whereas $\rho_{\min}$ should be greater than 0. Instead, in similarity-based threshold regression, it is $\delta = \lambda_p$, and λ_p is chosen by comparing various quantiles of the trigger variable $\{zt\}_{t=1}^n$.

To accommodate better the idea of similarity in the cross-validation scheme, we propose the following alternative, which we label *clustered cross-validation*. We focus the pseudo out-of-sample forecasting exercise on histories of data (blocks of observations) that are more similar to the current economic conditions. For this, we divide the data into blocks, and we search for blocks that are the most similar to the current block of observations. To cluster the data into blocks we use the trigger variable z_t. For this, suppose that data up to time j are denoted as $y^j = \{y_t\}_{t=1}^j$, $x^j = \{x_t\}_{t=1}^j$ and $z^j = \{z_t\}_{t=1}^j$. Then, we implement the following stepwise procedure.

(a) For each $j = n_1, \ldots, n$, we compute the distance

$$d_j^{m_0} = \sum_{i=1}^{m_0} (z_{n+1-i} - z_{j+1-i})^2. \tag{8}$$

(b) We use $d_j^{m_0}$ to order the blocks of data $\{y^j, x^j, z^j\}_{j=n_1}^n$, depending on how similar these are to the current regime, according to the trigger variable z_t. The quantity $d_j^{m_0}$ matches the last m_0 observations of z_t, i.e. $z_n, z_{n-1}, \ldots, z_{n+1-m_0}$, with other similar sequences in the data set $z^j = \{z_t\}_{t=1}^j$. Let $\{y^j, x^j, z^j\}_{j \in \Psi}$, with $\Psi = \{t_1, t_2, \ldots, t_{n-n_1+1}\}$, be the ordered histories, from the most similar (t_1) to the least (t_{n-n_1+1}).

(c) Finally, we find the δ that minimizes the MSFE over the n_0 most similar blocks of observations, $\{t_1, t_2, \ldots, t_{n_0}\}$, i.e.

$$\hat{\delta} = \arg\min_{\delta} \frac{1}{n_0} \sum_{t \in \{t_1, t_2, \ldots, t_{n_0}\}} (y_t - \hat{y}_{t|t-1,\delta})^2, \qquad \text{for } \delta \in [\delta_{\min}, \delta_{\max}]. \tag{9}$$

The values of n_1 and m_0 must also be chosen by the researcher. In practice, n_1 must be sufficiently large such that estimation from block $\{y^j, x^j, z^j\}_{j=n_1}$ does not pose numerical problems. A simple rule could be that model (1) can be reliably estimated with the data set $\{y^{n_1}, x^{n_1}, z^{n_1}\}$, for a sufficiently large value of H. Additionally, m_0 can be set equal to 1 or to a larger value, depending on the data at hand. In our empirical application and simulation experiments, we examine several values for m_0 as a check of robustness. A comparison example of the two cross-validation schemes is presented in the on-line appendix of the paper.

3. Alternative methods and forecast comparison

In this section we briefly review alternative time-varying forecasting models and forecast comparison criteria that will be used in later sections to assess the relative performance of our similarity-based methods.

3.1. Alternative time-varying forecasting methods

As discussed in Section 1, a common parametric time-varying approach assumes continuous evolution in the parameters of regression models; see for example Stock and Watson (1996) for an early forecasting application. The evolution of the parameter vector β_t is often specified as a multivariate random-walk process, with more general specifications feasible but more heavily parameterized. Hence, the model is

$$y_t = x_{t-1}\beta_{t-1} + \eta_{1t}, \qquad t = 1, \ldots, n, \tag{10}$$

$$\beta_t = \beta_{t-1} + \eta_{2t}, \tag{11}$$

$$\eta_{1t} \sim N(0, \sigma^2_{\eta_1}), \qquad \eta_{2t} \sim N(0, \Sigma_{\eta_2}),$$

where expression (10) is the measurement equation, equation (11) is the set of state equations, x_{t-1} is the vector of regressors, Σ_{η_2} is a diagonal matrix and η_{1t} and η_{2t} are independent error terms. In the following sections, we shall refer to model (10)–(11) as the TVP model.

An alternative specification that assumes abrupt parameter changes is the MS model, proposed by Hamilton (1989). We write an N-state MS model as

$$y_t = x_{t-1}\beta_{S_{t-1}} + \eta_t, \qquad \eta_t \sim \text{IID}N(0, \sigma^2), \tag{12}$$

$$\beta_{S_{t-1}} = \begin{cases} \beta_1 \text{ when } S_{t-1} = 1, \\ \beta_2 \text{ when } S_{t-1} = 2, \\ \qquad \vdots \\ \beta_N \text{ when } S_{t-1} = N \end{cases}$$

where S_{t-1} is the unobserved state variable, which is allowed to evolve stochastically according to a strictly stationary, homogeneous, first-order Markov chain with an $N \times N$ transition matrix $P = (p_{ij})$, with $p_{ij} = \Pr(S_t = i | S_{t-1} = j)$. In practice, the number of regimes is generally set at $N = 2$ or $N = 3$. In the following sections, we shall refer to model (12) as the MS model MS2AR or MS3AR, depending on the number of regimes.

Finally, as indicated by the recent literature (see for example Stock and Watson (2006)), we evaluate the constant parameter ARMA(p, q) model. This is defined as

$$y_t = x_{t-1}\beta + u_t, \qquad t = 1, \ldots, n, \quad u_t \sim \text{IID}(0, \sigma^2_u), \tag{13}$$

where $x_{t-1} = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p}, u_{t-1}, u_{t-2}, \ldots, u_{t-q})$, and β is a $p+q+1$ vector of parameters. When $q = 0$, model (13) is just an AR model of order p, AR(p). The ARMA orders p and q are typically set by optimizing an information criterion, like the Bayesian information criterion (BIC). Estimation of pure AR models can be done by OLS, whereas in the presence of an MA component we adopt the two-step procedure that was described in, for example, Dufour and Pelletier (2008).

All the methods that were presented above could be extended to allow for time variation in the variance. This could be important in particular for density forecasting, where a characterization of the uncertainty that is associated with the prediction is relevant.

Table 1 lists all the models that will be used in the forecasting exercises with simulated and actual data.

Table 1. Reference for the names of the methods in the following tables: description of the forecasting models

Name in the tables	*Explanation*
tv(n_0)-trig	AR model with trigger TVP and cross-validation over the last n_0 observations for optimal h; the kernel is normal; the cross-validation scheme is given in expression (7)
tv(n_0,m_0)-trig	AR model with trigger TVP and clustered cross-validation over the last n_0 observations for the optimal h; the kernel is normal; the clustered cross-validation scheme is given in equation (9), and m_0 is the clustering match parameter
tv(n_0)	AR model with TVP and cross-validation over the last n_0 observations for the optimal h; the kernel is normal; the cross-validation scheme is given in expression (7)
tv(n_0)-trig-arma(1,1)	ARMA model with trigger TVP and cross-validation over the last n_0 observations for optimal h; the kernel is normal; the cross-validation scheme is given in expression (7)
tv(n_0,m_0)-trig-arma(1,1)	ARMA model with trigger TVP and clustered cross-validation over the last n_0 observations for the optimal h; the kernel is normal; the clustered cross-validation scheme is given in expression (9), and m_0 is the clustering match parameter
SLA(n_0)	Similarity-based local averaging (see expression (6)) with cross-validation over the last n_0 observations for the optimal ρ; the cross-validation scheme is given in expression (7)
STR(n_0)	Similarity-based threshold regression (see expression (6)) of the AR model with cross-validation over the last n_0 observations for the optimal λ_p; the cross-validation scheme is given in expression (7)
STRc(n_0)	Similarity-based threshold regression (see expression (6)) of the AR model with cross-validation over the last n_0 observations for the optimal λ_p; the cross-validation scheme is given in expression (7); there is a break only on the constant parameter
TVP	TVP AR model; there is time variation on both the constant and the AR parameter of the process
TVPc	TVP AR model; there is time variation only on the constant of the process
MS2AR	Markov regime switching AR model with homogeneous variance and 2 states; there is regime switching on both the constant and the AR parameter of the process
MS2ARc	Markov regime switching AR model with homogeneous variance and 2 states; there is regime switching only on the constant of the process
MS3AR	Markov regime switching AR model with homogeneous variance and 3 states; there is regime switching on both the constant and the AR parameter of the process
MS3ARc	Markov regime switching AR model with homogeneous variance and 3 states; there is regime switching only on the constant of the process
ar(p)	AR model of order p
ar(bic)	AR model with p chosen by the BIC
arma(p,q)	ARMA average model of order p and q
arma(bic)	ARMA model with p and q chosen by the BIC
rw	The driftless random-walk model

3.2. Forecast comparison criteria

The forecasting performance of the alternative models is evaluated relatively to that of the benchmark: an AR(1) model, using the relative mean-squared forecast error rMSFE. For each model m and target series s, it is

$$\mathrm{rMSFE}_{(m,s)} = \frac{\sum_{t=t_0}^{T} (e_t^{(m,s)})^2}{\sum_{t=t_0}^{T} (e_t^{(\mathrm{AR}(1),s)})^2}, \tag{14}$$

where $e_t^{(m,s)} = y_t^{(s)} - \hat{y}_t^{(m,s)}$ is the one-step-ahead forecast error of model m for series s, and $e_t^{(\mathrm{AR}(1),s)} = y_t^{(s)} - \hat{y}_t^{(\mathrm{AR}(1),s)}$ is the counterpart for the benchmark AR(1) model. When $\mathrm{rMSFE}_{(m,s)}$ is less than 1, model m outperforms the benchmark AR(1) model for variable s. To assess the statistical significance of the MSFE differentials, we use the Diebold and Mariano (1995) test.

To assess whether the relative performance of a model is stable over time, we adopt a twofold approach. First, we compute the rMSFEs separately for the recession periods, as identified by National Bureau of Economic Research business cycle dating. Second, we implement the forecast fluctuation test that was developed by Giacomini and Rossi (2010). The forecast fluctuation test measures the relative, local forecasting performance for the two models. In contrast with the Diebold and Marino test, that measures the global performance over the forecasting horizon, the Giacomini and Rossi test concludes about the stability of the relative performance over the entire path of time. The test statistic is equivalent to the Diebold and Mariano statistic computed over rolling out-of-sample windows of size μ. In our empirical exercise we choose $\mu = 50$. Finally, to account for the comparison of many models, we have also considered the model confidence set (Hansen *et al.*, 2011). As the results are not conclusive, they are presented in the on-line appendix.

4. Simulation study

We carry out a Monte Carlo study to investigate the performance of our proposed similarity-based forecasting approaches in a controlled environment. We aim to study the overall forecasting performance and, more specifically, the working of the cross-validation approaches for estimating the tuning parameters.

4.1. Monte Carlo design

We consider two data-generating processes (DGPs). The first DGP accommodates the SLA model (see expression (4)) and it is specified as follows:

$$y_{t+1} = \sum_{i=1}^{t} w_i \tilde{y}_{t+1-i} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{15}$$

where $w_i = \rho^i / \Sigma_{j=1}^t w_j$ and $t = 1, \ldots, T$, with $T = 1000$. The $\{\tilde{y}_{t+1-i}\}_{i=1}^t$ are the ordered most similar observations to y_t (in the squared error sense), in the set $\{y_j\}_{j=1}^{t-1}$, with $\tilde{y}_{t-1}$ being the most similar to y_t (lowest squared error $(y_t - \tilde{y}_{t-1})^2$) and $\tilde{y}_1$ the less similar to y_t (highest squared error $(y_t - \tilde{y}_1)^2$). An initial sample and a burn-in period are needed for this scheme. For the initial sample we use 60 independent and identically distributed observations from $N(0, 1)$ and an additional burn-in period of 320 observations. After that period we store 1000 data points which are used for estimation and forecasting.

To generate realistic patterns of data, we set the parameters ρ and σ_ε at the estimates that are obtained when fitting model (15) to the US unemployment rate. Moreover, since in realtime forecasting the optimal ρ-parameter is allowed to change over time, we allow for varying values of ρ, $\{\rho_t\}_{t=1}^T$, set according to the values resulting from cross-validation in the US unemployment rate forecasting exercise. Specifically, we adopt 60 sequential estimates of ρ, from June 1st, 2005, until July 1st, 2014, and this sequence is then replicated a sufficient number of times to generate a total of $T + 380$ values for ρ. The last 240 data points of the series generated are used as the pseudo out-of-sample forecasting period.

Table 2. Forecasting exercise using simulated data of size $n = 1000$; forecasting performed over the last 240 observations: Monte Carlo simulation†

Model	*DGP1 (SLA)*	*DGP2 (MS2AR)*	*Model*	*DGP1 (SLA)*	*DGP2 (MS2AR)*
tv(6)-trig	0.95	0.91	SLA(true ρ)	0.74	—
tv(12)-trig	0.93	0.9	STR(6)	0.93	0.98
tv(6,1)-trig	0.93	0.89	STR(12)	0.96	0.99
tv(12,1)-trig	0.93	0.88	TVP	1	0.94
tv(6,2)-trig	0.94	0.89	TVPc	1	0.94
tv(12,2)-trig	0.93	0.88	MS2AR	0.91	0.82
tv(6)	0.98	0.95	MS2ARc	0.97	0.88
tv(12)	1	0.97	arma(bic)	0.93	1
SLA(6)	0.89	0.97	ar(bic)	1	1
SLA(12)	0.94	0.95	ARMA(1,1)	0.95	0.92

†The models are defined in Table 1. The figures are MSFEs relative to that of the AR(1) model, averaged over 300 replications.

In the on-line appendix of the paper (see Fig. A2), we report a sample from the generated series, as well as the set $\{\rho_t\}_{t=1}^T$ of tuning parameters that are used to generate the series. Intuitively, this process implies a transition between different regimes through the parameter ρ. As ρ approaches 1, the SLA model reduces to the simple sample average of the past data, denoting small persistence of the series. As ρ diminishes, the characteristics of the series differ, and in particular persistence increases.

In the second DGP, we generate data according to a two-regime MS AR model. For this, in expression (12) we set $S_t = \{S_1, S_2\}$, $x_{t-1} = [1, y_{t-1}]$, homogeneous variance σ and 2×2 transition matrix $P = (p_{ij})$, with $p_{ij} = \Pr(S_t = i | S_{t-1} = j)$. To generate realistic patterns of data, we also fit this model to the unemployment rate series and set the DGP parameters equal to the estimated values. Specifically, we have state 1 (S_1) parameters,

$$\beta_{11} = -0.044, \qquad \beta_{12} = -0.195, \qquad \sigma = 0.023;$$

state 2 (S_2) parameters,

$$\beta_{21} = 0.23, \qquad \beta_{22} = 0.0026, \qquad \sigma = 0.023;$$

transition matrix,

$$p_{12} = 0.016, \qquad p_{21} = 0.104$$

where $\beta_i = (\beta_{i1}, \beta_{i2})'$. The state variable $\{S_t\}_{t=1}^T$ is sampled once, and then it is kept fixed in all simulations. The length of the series generated is $T = 1000$, whereas we forecast one step ahead over the last 240 observations.

One issue that arises in the application of our approaches is the choice of the trigger variable z_t on the simulated samples (for example, see equations (3) and (6)). Since z_t is assumed to be an informative indicator of prevailing regime, a natural candidate is a smoothed transformation of the original target series. Hence, in our simulation experiments, for an initial value $P_0 = 100$ and a generated series $\{y_t\}_{t=1}^T$, we use the transformation $P_t = P_{t-1}(1 + y_{t-1})$ to generate $\{P_t\}_{t=1}^{T+1}$. Then, we set the trigger variable z_t as the smoothed growth rate of P_t, $z_t = (P_t - P_{t-3})/P_{t-3}$. This smoothed transformation is also considered later in the empirical applications, in addition to other choices for z_t.

For the two DGPs, we examine the one-step-ahead out-of-sample forecasting performance of the proposed similarity-based forecasting approaches; the MS model, the TVP model, AR(1), ARMA(1,1), AR(p) and ARMA(p,q) with p and q selected by the BIC. The models under comparison are listed in Table 2, with additional details for each model reported in Table 1. For each model and for each of 300 replications, we compute the MSFE relative to that of the AR(1) model, and then we average the relative MSFEs over the replications.

4.2. Simulation results

Out-of-sample forecasting results of the simulation exercise are presented in Table 2. Focusing on the first DGP (SLA), some interesting conclusions could be made. First, the infeasible SLA model with the true value of ρ is the best (rMSFE = 0.74) method. The second best is SLA with ρ selected according to cross-validation based on the last six observations (rMSFE = 0.89). The tv-trig models yield rMSFEs in the range 0.93–0.95, similarly to the similarity threshold model, with 0.93–0.96. The time-varying models with the time index t as trigger do not perform satisfactorily, with rMSFE in the range 0.98–1, and a similar finding holds for the TVP models. Instead, the MS models, which allow for abrupt changes in the parameters, perform quite well, in particular when allowing for changes in all the parameters, with an rMSFE of 0.91. ARMA models also perform remarkably well in this exercise, in particular, when the lag order is chosen

Table 3. Series used in the forecasting exercise†

Series name	*Short explanation*	*Stationarity transformation*
PAYEMS	All employees: total non-farm	First log-differences
RPI	Real personal income	First log-differences
FFR	Effective federal funds rate	First differences
MZMSL	MZM money stock	Second log-differences
UNRATE	Civilian unemployment rate	First differences
WPSID61	PPI: intermediate materials	Second log-differences
AVGHE	Average hourly earnings: goods producing	Second log-differences
M1SL	M1 money stock	Second log-differences
CUMFNS	Capacity utilization: manufacturing	First differences
CONSPI	Non-revolving consumer credit to personal income	First differences
INDPRO	Industrial production index	First log-differences
HOUST	Housing starts: total new privately owned	Logarithm
CPI	Consumer price index: all items	Second log-differences
TB3MS	3-month treasury bill	Second log-differences

†The stationarity transformation is as recommended in the FRED-MD database.

Table 4. Trigger variables used in the similarity-based forecasting methods

Same as the series to forecast
Quarter-on-quarter difference (or log-difference) of the forecast series
Year-on-year difference (or log-difference) of the forecast series
Oil price, month-on-month second log-differences
HOUST, housing starts: total new privately owned, logarithm
Fed rate, effective federal funds rate quarter-on-quarter differences

Table 5. MSFE results for the whole forecasting period (see Tables 1 and 3 for a description of the models and the series)†

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Whole sample forecasting														
ar(bic)	*0.7*	1.04	0.89	0.91	*0.87*	*0.94*	0.99	1.01	0.99	0.98	*0.91*	1	0.98	*0.87*
rw	1.12‡	1.46	1.75	4.01	2.24	1.56	1.22	2.02	1.87	5.19	1.59	2.28	2.68	0.99
TVP	0.7‡§	1	0.91	1	1	0.99	1	1	0.93‡§	1	1.07	1	1	0.87‡§
TVPc	0.7‡§	1	0.91	1	1	0.99	1	1	0.93‡§	1	1.07	1	1	0.87‡§
MS2AR	0.92‡§§	1.06	*0.87*§	0.94‡	1.05	0.98	1	1.21	0.94‡§	1	0.96‡§	1.07‡	0.99	1.01
MS2ARc	1	1	0.94‡§	1	1	1.01	1	1.01	0.92‡§	1	0.99	0.99‡	1	1
MS3AR	0.99	1.01	1‡	1.33	1.04	1.02	1.06	1	0.95‡§	1.7	1.03	1.01‡	1.15	1
MS3ARc	1	1	1.12	1.31	1.1‡	1.01	1	1	1.26‡	1.67	1	1.19	1.14	1
tv(6)	0.87‡§§	1.09	0.96	1.59	1.04	1.56	1.07	1.08‡	1.13*	1.43	1.09	1.42‡	1.04	1.03
tv(12)	0.83‡§	1.04	0.98	1.57	1.03	1.55	1.08*	1.04	1.01	1.41	1.11	1.43‡	1.01	1
SLA(6)	0.74‡§	1.06	0.98‡	1.38	1.07	1.04‡	1.11‡	1.21‡	1.01	1.63	0.98‡	1.2	1.1	1.01
SLA(12)	0.84‡§	1.01	0.97‡	1.33	0.99‡	1.04	1.06‡	1.07	1	1.62	0.97	1.18	1.06	1.02
Trigger variable: quarter-on-quarter difference of the forecasted series														
tv(6)-trig	0.75‡§	1.04‡	0.92‡	1.11	1.04	1.01	1.14‡	0.98	1.03‡	0.97	0.98	1.25‡	0.95	
tv(12)-trig	0.77‡§	1.09‡	0.9‡§	1.1	1.17	0.98	1.01‡	1‡	1.01‡	0.95§§	0.98‡	1.23	0.97	
tv(6,1)-trig	0.77‡§	1.03	0.91‡§§	0.95‡	1.02	0.98‡	*0.86*‡	0.89	0.97*	0.94§§	0.96‡	0.84‡§	*0.92*	
tv(12,1)-trig	0.77‡§	1.14‡	0.91§	0.94‡	0.97‡	0.97	0.91‡	0.9	0.98‡	0.94§	0.95‡§	*0.83*‡§	0.92	
tv(6,2)-trig	0.7‡§	1.38	0.91§	0.91‡	0.94§	0.99*	1‡	0.9	0.97‡§	*0.91*	0.96§	0.96‡	0.98	
tv(12,2)-trig	0.71‡§	1.5	0.9‡§	*0.89*‡	0.97	0.98*	1.11‡	*0.88*‡	0.97‡	0.91	0.96‡§	0.92‡	0.98	
STR(6)	0.75‡§	1.06	0.89§	0.98	0.95§	0.96‡§§	1.01‡	1.14	0.95§*	1	0.96‡§§	0.94‡§§	0.99	1
STRc(6)	0.78‡§	1.04‡	0.91§	0.96§	0.95§	0.94§	1‡	1.02	0.95§*	0.99	0.97‡§§	1.02	0.97	1
STR(12)	0.78‡§	1.05	0.89§	0.97	0.94§	0.97	1.03‡	1.12	0.97‡	0.95	0.97‡	0.91‡§	0.98§§	1‡
Trigger variable: year-on-year difference of the forecasted series														
tv(6)-trig	0.86‡§	0.94	0.94‡	1.26	1.11	1.04	1.58‡	1.29‡	1	1.1	1.03‡	1.05	2.21‡	
tv(12)-trig	0.84‡§	1.07	0.91‡§	1.15	1.03	1.03	1.55‡	1.27‡	1‡	1.02	0.99‡	1.06	2.2	
tv(6,1)-trig	1.01*	0.95	0.94‡§	1.27‡	0.98	1.01	0.99	1‡	0.98	1.06	1.01*	0.97‡	0.99	
tv(12,1)-trig	0.92‡	1.07	0.92‡§	1.06‡	0.97	1	1.02	0.99	0.97§§	1.02	1.03‡	1.02‡	0.99	
tv(6,2)-trig	0.95‡	1.05	0.95‡§§	1.12‡	1	1	1.48‡	0.95‡	0.97§§	1	0.96§§	0.95‡	0.98	
tv(12,2)-trig	0.93‡	1.04	0.93‡§	1.12	0.97	1	0.99	0.95	0.98*	1.01	0.98‡	0.99‡	0.99*	

(*continued*)

Table 5 (*continued*)

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Trigger variable: same as the forecasted series														
tv(6)-trig	0.98	1.57	0.94‡§	1.3	1.08	0.97	1.01	1.55	0.93‡§§	1.21	1.01	0.93‡	1.04	1.04
tv(12)-trig	0.97‡	1.03	0.94‡§	2.19‡	1.09	0.97‡	1	1.55	<u>*0.91*</u>‡§	1.23	0.98	0.93‡	1.02	1.04
tv(6,1)-trig	1‡	0.97	0.98	1.96‡	1.07	1.01	1.01	1.3	1	1.06	1.01	0.86‡§§	1.13	0.99
tv(12,1)-trig	0.98‡	0.98	0.97	0.98‡	1.08	1.01	0.99*	1.36	1‡	1.03	1	0.85‡§	1	1.01
tv(6,2)-trig	0.97‡	2.87	0.97‡	1.95‡	1.2	0.99	0.99	1.37	0.95§§	1.02	1	0.88‡§	1.05	0.98‡
tv(12,2)-trig	0.93‡§	1	0.96§§	1.17	1.01	0.98	0.98	1.39	0.95‡§	1.03	0.98	0.86‡§§	1.04‡	0.99
Trigger variable: oil price, 2nd difference, month on month														
tv(6)-trig	0.95‡§	0.98	1.02	1.05	0.99	1.04	1.03	0.97	1	1.01‡	1.01	1.06	0.99‡	1.02
tv(12)-trig	0.97‡	1.01‡	1.01‡	1.04	0.98	1.01	1.03	1	0.99	1.01	1	1.04	0.99‡	1.01
tv(6,1)-trig	0.99	1	1‡	1.02	0.97§	1.01‡	0.99	1.01	0.99	1.02	1.02	1.01	0.99	1.01
tv(12,1)-trig	0.99	1	1	1.06	0.96§	0.99	0.99	1.04	1	1.01	1.01	1.01	0.98	1.01
tv(6,2)-trig	1.01‡	1.02	1.02	1.03	0.96§	1.01‡	1.35	1.01	1	1.02	1.01	1.03	0.99‡	1.03
tv(12,2)-trig	0.98‡	0.98	1.01	1.03	0.95§	0.99	1	1.01	1	1	1	1.03	0.99‡	1
Trigger variable: housing starts, level														
tv(6)-trig	0.75‡§	1.24	0.88§	1.04	1.09	0.94‡§	1.04‡	1.14‡	1.04	1.2	0.93§	1.01‡	1.45	1.03
tv(12)-trig	0.79‡§	0.99	0.87§	1.04	1.08	0.96‡	1.01‡	1.08‡	1.03‡	1.19	0.98*	1	1.45	1.04
tv(6,1)-trig	0.88‡§§	1.12	0.95	1.05	1.11	1.01‡	1.09‡	1.89	1.01‡	1.03	1.03‡	1.04	1.05	1
tv(12,1)-trig	0.82‡§	1.09‡	0.95‡	1.1	1.03	0.98	1.07	1.13	1.04*	1.03	0.99	1	1.05*	1.01
tv(6,2)-trig	0.9‡	1.03	0.93	1.17	1.16	1.01‡	0.98	1.08	1.04‡	1.02	1.02	1.05‡	1.02‡	0.99‡
tv(12,2)-trig	0.88‡§§	1.01	0.94	1.1	1.01	0.98	1.09	1.04	1.03‡	1.02	1.02‡	1.02	1	0.99‡

(*continued overleaf*)

Table 5 (*continued*)

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Trigger variable: Fed funds rate, quarter on quarter														
tv(6)-trig	0.92	1.02	0.96§§	1.11	1.02	0.99		1.05	1	1.05	0.96‡§	1.06‡	1.01	1.1
tv(12)-trig	0.93	1.02	0.93§*	1.1	1.03	0.99		1.02	0.99	1.03	0.96§*	1.04‡	1	1.06
tv(6,1)-trig	0.99	0.99	1	1.02	0.98	1		1.14	0.98	1.03	0.98‡	0.99	0.99	1
tv(12,1)-trig	0.99	0.99	0.98	1.03	0.98	1.01‡		1.12	0.97	1.01	0.98	1	0.99	1
tv(6,2)-trig	1	*<u>0.91</u>*	1‡	1.04	1	0.98		1.04	1.04	1	0.98	1	1	1.01
tv(12,2)-trig	0.94	0.98	0.97‡§	1.01	0.98	0.97‡§		1.04	1	1.01	1	1.07	0.98	1.01

†The results are relative to the forecast MSFE of the AR(1) benchmark model. See Table 4 for a description of the trigger variables. In italics and underlined are the best performing methods, when these are better than the benchmark.
‡Statistically different forecasts from the AR(1) model at the 5% significance level according to the forecast fluctuation test.
§Statistically different forecasts from the AR(1) model at the 5% significance level according to the Diebold and Mariano test.
§§Statistically different forecasts from the AR(1) model at the 10% significance level according to the Diebold and Mariano test.
*Statistically different forecasts from the AR(1) model at the 10% significance level according to the forecast fluctuation test.

by the BIC (rMSFE = 0.93), confirming the relevance of including an MA component when it is suspected of parameter instability.

For the second DGP (MS2AR) the best performer is MS2AR, with an rMSFE of 0.82, whereas the rMSFE of MS2ARc is 0.88. The rMSFEs of the tv-trig models are in the range 0.88–0.91, which is better than those of SLA, 0.95–0.97, and the TVP model, 0.94. The use of the BIC for choosing the order of the ARMA model does not provide any forecasting gains over the benchmark in this DGP, but an ARMA(1,1) model has a decent performance, with an rMSFE of 0.92.

Note that, when we generate data from the two-regime MS model, the SLA(6) models fails to produce reliable forecasts, compared with the other models that were examined. We consider that this is reasonable behaviour, because SLA is favoured by DGPs with a smooth transition from one regime to the other rather than abrupt changes of regimes such as those considered in the MS model.

Focusing on cross-validation for the time-varying trigger models, in both experiments the differences between the two versions are small whereas the clustered approach seems to provide slight overall improvements over the end-of-sample scheme.

In summary, the similarity-based forecasting approaches work satisfactorily in both experiments, though the gains are not very large compared with MS specifications. The gains are larger with respect to the TVP model, but this is likely to be due to the fact that TVP models are not so suited to capture the kind of breaks that characterize both DGPs. Adding an MA component to AR models is also helpful in the presence of unmodelled parameter time variation.

5. Empirical application

5.1. Data

Our forecasting empirical analysis is performed on a set of key monthly US macroeconomic and financial variables, recursively over the sample 1961, month 1–2017, month 4. The out-of-sample forecasting evaluation is performed over the last 440 observations, i.e. from 1980, month 1, until 2017, month 4. Such a long sample enables us to have enough observations for estimation and out-of-sample evaluation, over periods of varied economic conditions.

To choose the target variables, we start from those considered by Guerron-Quintana and Zhong (2017), dropping those not available in the database FRED-MD (see McCracken and Ng (2016)) for the sample considered, and adding similar available series. We end up with 14 indicators (see Table 3 for details): employment, unemployment, earnings, real personal income, industrial production, capacity utilization, housing starts, federal funds rate, 3-month rate, money stock, consumer credit, CPI and PPI. All the data were downloaded from FRED-MD, and transformed as suggested by McCracken and Ng (2016); see Table 3 for additional details and Fig. A5 in the on-line appendix of the paper.

Regarding the trigger variable z_t, we consider two main possibilities. First, we use the target variable itself: a common choice in threshold and smooth transition models. Specifically, we consider either the variable itself, or its quarter-on-quarter or year-on-year differences (Table 4), which have often a smoother behaviour. Second, we use alternative macroeconomic indicators whose behaviour could affect the dynamics of the target variable of interest: (changes in) oil prices as a measure of external shocks, the federal funds rate as a measure of the monetary policy stance and housing starts as a leading indicator of economic conditions.

5.2. Empirical results

The main empirical findings are reported in Tables 5–8. We report, for each variable, the MSFE

Table 6. MSFE results for the recession period (see Tables 1 and 3 for a description of the models and the series)†

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Recession period forecasting														
ar(bic)	0.51	0.95	0.57	0.88	0.97	0.89	1.05	1.15	1.03	0.93	0.87	1	1.16	0.93
rw	0.35‡	1.35	0.75	3.93	1.74	1.2	1.19	2.67	1.81	6.39	1.23	1.98	2.41	0.88
TVP	0.34‡	1	*0.45‡*	1	1	*0.8‡*	1	1	1.01	1	1.07‡	0.97	1	*0.88‡*
TVPc	0.34‡	1	0.45‡	1	1	0.8‡	1	1	1.01	1	1.07‡	0.97	1	0.88‡
MS2AR	0.71‡	1.02	0.6	0.87	1.03	0.99	0.99	1.12	0.91	1.02	0.93	0.86	1	1.01
MS2ARc	0.93§	1	0.77	1	1	0.97	1.01	1	0.9	0.99	0.96	0.92	1	1.03
MS3AR	0.91‡	1	0.85§	1.32	0.92	0.97	1.03	1.11	0.92§	1.82	0.91‡	1.02	1	1.03
MS3ARc	0.92‡	0.97	0.97	1.31	0.94	0.97	0.96	1.01	0.9§	1.81	0.92	0.99	1.02	1.03
tv-N-CV(6)	0.37‡	1.05	0.65	2.33	0.99	2.53	1.13	0.95	*0.87*	2.15	1.36	1.13	1.01‡	1.04
tv-N-CV(12)	0.39‡	1	0.64	2.33	1	2.56	1.14	0.96	0.94	2.13	1.4	1.18	0.98	1
SLA-CV(6)	*0.31‡*	1.21	0.67	1.33	1	1.22	1.09	1.12	0.87	1.89	1.15	1.08	0.98	0.89‡
SLA-(12)	0.34‡	0.99	0.74	1.32	*0.89*	1.23	*0.9*	1.08	0.91	1.89	1.17	1.08	0.97	0.92‡
Trigger variable: quarter-on-quarter difference of the forecasted series														
tv-N-CV(6)-trig	0.43‡	0.94	0.62	1.19	1.1	1.01	1.02	0.95	1.01	0.95	*0.82*	0.93	0.96	
tv-N-CV(12)-trig	0.46‡	1	0.59	1.19	1.52	0.98	0.99	1.16	1	0.94	0.97	0.94	0.97	
tv-N-CV(6,1)-trig	0.48‡	0.99	0.6	0.96	1.09‡	0.97‡	1	0.98	1.08§	0.92	0.93§	0.8	*0.8*	
tv-N-CV(12,1)-trig	0.45‡	0.96	0.62	0.93	0.96	0.96‡	1	0.99	1.02	0.97	0.97§	0.83	0.8	
tv-N-CV(6,2)-trig	0.46‡	0.96	0.66	0.89	0.94‡	0.96	1.05	1.04	1.04	0.8	0.99	*0.78*	1	
tv-N-CV(12,2)-trig	0.43‡	0.94	0.62§	*0.8*	0.98‡	0.96§	1.01	0.98	1.01	*0.79*	0.97	0.79	0.99	
STR-CV(6)	0.39‡	1.06	0.59	0.98	0.94‡	0.86‡	1	1.01	1.01	1.03	0.92‡	0.83	1§	0.99
STRc-CV(6)	0.46‡	1.09	0.63	0.95	0.93‡	0.82‡	0.98§	0.99	0.99	1.01	0.88‡	1.01	0.99§	0.98
STR-CV(12)	0.5‡	1	0.6	0.99	0.93	0.91‡	1.02	0.99	1.06	0.9	0.97§	0.89	1.01§	0.98
Trigger variable: year-on-year difference of the forecasted series														
tv-N-CV(6)-trig	0.55‡	0.94	0.72	1.05	1.12	1.07§	1.12	0.94	1.17	1.25	1.02§	1.18	4.54	
tv-N-CV(12)-trig	0.6‡	1.04	0.73	1.04	1.12	1.07§	1.14	*0.92*	1.18	1.06	1.03	1.18	4.57	
tv-N-CV(6,1)-trig	0.84‡	0.95	0.77‡	1.03	0.97	1.04‡	1.06	0.92	1	1.18	1.09§	1.02	0.99	
tv-N-CV(12,1)-trig	0.59‡	*0.92*	0.75‡	1	0.96	1.03‡	1.11	0.96	0.97	1.08	1.12	1.06	0.99	
tv-N-CV(6,2)-trig	0.68‡	0.93	0.83	0.99	1.03	1.04‡	1.11	0.99	0.92	1.01	0.89	1	0.99	
tv-N-CV(12,2)-trig	0.64‡	0.97	0.8	0.98	0.98	1.04‡	1.04	0.95	0.95	1.03	1.05	1	1	

(continued)

Table 6 (*continued*)

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Trigger variable: same as the forecasted series														
tv-N-CV(6)-trig	0.83	2.36	0.85	1.56	1.13	0.99	0.99	1.31	1	1.69	1.02	0.89	1.08	1.23
tv-N-CV(12)-trig	0.84	1.08	0.86	3.43	1.19	0.99	0.97	1.29	0.96	1.69	0.99	0.88	1.08	1.25
tv-N-CV(6,1)-trig	0.9	0.99	0.88‡	3.27	1	1.05	0.98	1.09	0.97	1.15	1.04	0.82	1.12	0.97
tv-N-CV(12,1)-trig	0.87	1.03	0.88	0.97	1	1.06	0.96	1.06	0.89	1.12	1.03	0.78	1.02	1.01
tv-N-CV(6,2)-trig	0.89	5.6	0.87	3.32	1.61	0.99	0.94	1.13	0.97	1.08	1.03	0.84	1.11	0.99
tv-N-CV(12,2)-trig	0.86	1.02	0.91	0.99	1.01	0.98	0.96	1.14	0.88	1.09	1	0.84	1.11	1.02
Trigger variable: oil price, 2nd difference, month on month														
tv-N-CV(6)-trig	0.87§	0.98	0.99	1.09	1.02	1.07	1.08	1.11	0.99	1	0.97	1.05	0.98	1.03
tv-N-CV(12)-trig	0.88	1	0.98	1.1	1	1.01	1.09	1.12	0.99	1	0.98	1.03	0.98	1.01
tv-N-CV(6,1)-trig	0.95	1.01	0.95	1.04	0.99	1	0.98	1.03	0.98	1.01	1.01	1.02	1.02	0.99
tv-N-CV(12,1)-trig	0.97	0.98	1	1.17	1	1	1	1.21‡	1	1	1.02	1.03	1.01	0.99
tv-N-CV(6,2)-trig	0.9	1.06	0.98	1.09	0.99§	0.98	0.97	1.01§	0.99	1.01	0.96	1	0.97	1
tv-N-CV(12,2)-trig	0.91	1	1.01	1.08	0.95§	0.96	1	1	1	1.02	1	1.02	0.96	0.99
Trigger variable: housing starts, level														
tv-N-CV(6)-trig	0.47‡	1.43	0.5‡	1.05	1.15	0.9‡	1.07	1.07	1.21	1.53	0.86‡	1.13	2.33	1.19
tv-N-CV(12)-trig	0.49‡	1.07	0.49§	1.04	1.14	0.86‡	0.98§	1	1.11	1.56	0.95‡	1.09	2.34	1.22
tv-N-CV(6,1)-trig	0.47‡	1.14	0.52	1.05§	1.3	0.92§	0.95	5.02	1.03	1.05	1.02§	1.01	1.13	1
tv-N-CV(12,1)-trig	0.47‡	1.12	0.54	1.2	1.08	0.84§	1.02	1.2	1.11	1.06	0.88‡	1.01	1.12	1.01
tv-N-CV(6,2)-trig	0.48‡	1.17	0.53§	1.34§	1.46	0.87	0.92	1.16	1.02	0.97	0.93	1.07	1	1.04
tv-N-CV(12,2)-trig	0.5§	1.07	0.49§	1.21§	1.01	0.82§	0.96	1.2	1.05	1.01	0.97	1.06	0.99	1.02
Trigger variable: Fed funds rate, quarter on quarter														
tv-N-CV(6)-trig	0.98‡	1.02	1.07	1.16	0.97	1.05		1.14	1.02	1.06	1.01	1.02	1	1.21
tv-N-CV(12)-trig	0.98§	1.04	1	1.16	0.99	1.01		1.06	1	1.04	0.98	1	0.98	1.13
tv-N-CV(6,1)-trig	0.99	0.98	1.03	1.02	0.99	1.08		1.03	0.99	1.03	1.09	1.02	1	1
tv-N-CV(12,1)-trig	1.01	0.98	1.03	1.03	0.98	1.11		1.03	1	1.03	1.07	1	1	0.99
tv-N-CV(6,2)-trig	0.99	1.04	1.02	1.04	0.97	1.04		1.04	1	1	1.06	1.01	1	1.02
tv-N-CV(12,2)-trig	0.97	1.05	0.98	0.99	0.97	1.02		1.01	1.01	1	1.07	1.01	0.96	1.02

†The results are relative to the forecast MSFE of the AR(1) benchmark model. See Table 4 for a description of the trigger variables. In italics and underlined are the best performing methods, when these are better than the benchmark.
‡Statistically different forecasts from the AR(1) model at the 5% significance level according to the Diebold and Mariano test.
§Statistically different forecasts from the AR(1) model at the 10% significance level according to the Diebold and Mariano test.

Table 7. MSFE results for the whole forecasting period (see Tables 1 and 3 for a description of the models and the series)†

Model	*Result for the following series:*														
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Whole sample forecasting														
arma(1,1)	0.72	1.01	0.91	*0.75*	0.78	0.91	*0.97*	*0.97*	0.92	0.68	*0.92*	1	0.95	0.87
arma(bic)	1	1.1	1	0.76‡§	0.98‡	0.92§*	1.11	1.08	1	0.68‡§	0.93‡§	1	0.94	1
Trigger variable: quarter-on-quarter difference of the forecasted series														
tv(6)-trig-arma(1,1)	0.69§	1.07	0.86§	1.14‡	1.71‡	0.99*	1.09‡	1.2	0.94‡§	0.69‡§	1.05	0.99	*0.87*‡§§	
tv(12)-trig-arma(1,1)	0.68‡§	1.07	0.85‡§	1.11‡	1.7‡	0.97	1.09	1.14‡	0.92‡§	0.68‡§	0.97‡	0.98*	0.88§§	
tv(6,1)-trig-arma(1,1)	0.71‡§	1.08	0.84‡§	0.93	0.89‡§§	0.94§	1.2*	1.03*	0.94‡§	0.68‡§	0.95‡	0.86‡§	0.87‡§§	
tv(12,1)-trig-arma(1,1)	0.71‡§	1.21	*0.83*§*	0.95	0.88‡§	0.93§	1.22‡	1.06‡	0.92‡§	0.68‡§	0.94‡§	0.87‡§	0.91‡	
tv(6,2)-trig-arma(1,1)	0.71‡§	1.36	0.86‡§	0.98	0.87‡§§	0.93§	1.16‡	1.04	0.91‡§	0.68‡§	0.94‡§	0.94‡	0.91	
tv(12,2)-trig-arma(1,1)	0.71‡§	1.46‡	0.84‡§	0.84‡	0.85‡§	0.92§	1.38‡	1‡	0.92‡§	0.69‡§	0.92‡§	0.91‡§§	0.93‡	
Trigger variable: year-on-year difference of the forecasted series														
tv(6)-trig-arma(1,1)	0.76‡§	*0.99*	0.87§*	0.88‡§§	0.84‡§	0.98	1.86‡	1.53	1‡	0.72‡§	0.92§	1.09	2.72	
tv(12)-trig-arma(1,1)	0.72‡§	1	0.85§	0.85‡§	0.83‡§	0.98	2.48‡	1.48	0.98*	0.68‡§	0.92‡§	1.19	2.73	
tv(6,1)-trig-arma(1,1)	0.78‡§	1.14‡	0.85‡§	0.94	0.8‡§	0.92§	0.97‡	1.33	0.92§	0.7‡§	0.94‡§§	0.96‡	0.99	
tv(12,1)-trig-arma(1,1)	0.82‡§	1.11‡	0.85‡§	0.93‡	0.84‡§	0.94‡§	1.01*	1.31	0.92§	0.69‡§	0.93‡§	1.01	0.98‡	
tv(6,2)-trig-arma(1,1)	0.79‡§	1.04	0.85§*	0.91‡	0.84‡§	0.99	1.68‡	1.01	0.93‡§	0.67‡§	0.93‡§	0.93‡	1.02	
tv(12,2)-trig-arma(1,1)	0.78‡§	0.99	0.86§*	0.93‡	0.82‡§	0.93§*	0.97	1.23	0.94‡§	0.68‡§	0.92‡§	0.97‡	0.96	
Trigger variable: same as the forecasted series														
tv(6)-trig-arma(1,1)	0.85‡§	1.14	0.85‡§	1.32‡	0.95‡	0.92§	1.15‡	1.26	*0.87*‡§	0.69‡§	0.95‡§	1.76‡	0.92§§	0.92‡§
tv(12)-trig-arma(1,1)	0.84‡§	1.14	0.9‡§	0.92‡	0.94‡	0.93§*	1.03	1.29	0.89‡§	0.69‡§	0.93‡§	1.76‡	0.95*	0.93‡§§
tv(6,1)-trig-arma(1,1)	0.79‡§	1.04	0.88§*	1.03	0.92‡	0.93§	1.12	1.68	0.9‡§	0.68‡§	0.95‡§	*0.83*‡§	0.97‡	0.88‡§
tv(12,1)-trig-arma(1,1)	0.83‡§	1.06	0.88‡§	0.82‡§	0.92‡	0.92§	1.17‡	1.69	0.93‡§	0.66‡§	0.94§	0.83‡§	0.96	0.88‡§
tv(6,2)-trig-arma(1,1)	0.8‡§	1.09	0.88‡§	1.04‡	0.83‡§	1.07	1.06	1.96	0.91‡§	*0.65*‡§	0.95‡§§	0.83‡§	1.01‡	0.9‡§
tv(12,2)-trig-arma(1,1)	0.78‡§	1.09	0.89‡§	0.82‡§	0.84‡§	0.93‡§§	1.15	2.3	0.93‡§	0.66‡§	0.95‡§	0.83‡§	0.99	0.86‡§

(*continued*)

Table 7 (*continued*)

Model	*Result for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Trigger variable: oil price, 2nd difference, month on month														
tv(6)-trig-arma(1,1)	0.72‡§	1.06	0.9§	0.86‡§	0.8‡§	1.08	1.02	0.97	0.92‡§	0.68‡§	0.95‡	1.03	0.93	0.88‡§
tv(12)-trig-arma(1,1)	0.74‡§	1.04‡	0.89‡§	0.87‡§	0.8‡§	1.07	1.02	1.19	0.92‡§	0.69‡§	0.93‡§	1.04	0.92	0.88‡§
tv(6,1)-trig-arma(1,1)	0.72‡§	1.03	0.89§*	0.9‡§	0.84‡§	0.93‡§	1.04	1.01	0.9‡§	0.67‡§	0.94‡§	0.99	0.96	0.87‡§
tv(12,1)-trig-arma(1,1)	0.72‡§	1.02	0.88‡§	0.87‡§	0.82‡§	0.92‡§	1.04	1.08	0.91‡§	0.67‡§	0.93‡§	1.01	0.93	0.86‡§
tv(6,2)-trig-arma(1,1)	0.71‡§	1.05	0.9§	0.83‡§	0.77‡§	0.92§§	1.06	1.08	0.92‡§	0.7‡§	0.93§*	1	0.95	0.89‡§
tv(12,2)-trig-arma(1,1)	0.7‡§	1.02	0.9‡§	0.83‡§	*0.76*‡§	*0.9*§	1.02	1.08	0.92‡§	0.67‡§	0.93§	1	0.95‡	0.86‡§
Trigger variable: housing starts, level														
tv(6)-trig-arma(1,1)	0.73‡§	1.59	0.85§	1.13	0.9‡	1.01	1.17	1.02	1.01‡	0.76‡§	0.98‡	1.01‡	1.54	0.89‡§
tv(12)-trig-arma(1,1)	0.75‡§	1.06	0.85‡§	1.11	0.89‡	1.01	1.04	1.11	0.99‡	0.76‡§	0.94‡§§	1.02	1.54	0.91‡§
tv(6,1)-trig-arma(1,1)	0.72‡§	1.22	0.91‡§§	1.01	0.96‡	0.98	1.25‡	1.8	0.93‡§	0.69‡§	0.95‡§§	1.07	1‡	0.88‡§
tv(12,1)-trig-arma(1,1)	*0.67*‡§	1.21‡	0.89‡§	9.08‡	0.9‡§§	0.93‡§	1.29	1.05	0.94‡§	0.67‡§	0.93‡§	1.01	1.03‡	0.87‡§
tv(6,2)-trig-arma(1,1)	0.69‡§	1	0.88‡§	1.01‡	0.92‡	1.02	1.14	1.03	0.96‡§§	0.7‡§	0.99‡	1.08	0.89§*	0.88‡§
tv(12,2)-trig-arma(1,1)	0.68‡§	1.05	0.88§*	0.88‡§	0.86‡§	0.95‡§§	1.18	1.07	0.92‡§	0.71‡§	0.96‡	1.02	0.93*	0.87‡§
Trigger variable: Fed funds rate, quarter on quarter														
tv(6)-trig-arma(1,1)	0.69‡§	1.07	0.87‡§	0.89‡§	0.86‡§	0.94§§		2.29	0.93‡§	0.72‡§	0.93‡§	1.06	1.05	0.9‡§
tv(12)-trig-arma(1,1)	0.68‡§	1.04	0.85‡§	0.87‡§	0.83‡§	0.93§		2.3	0.89‡§	0.78‡§	0.92‡§	1.04	1.05‡	0.89‡§
tv(6,1)-trig-arma(1,1)	0.7‡§	1.02‡	0.91‡§	0.84‡§	0.8‡§	0.93§		1.09	0.89‡§	0.66‡§	0.92‡§	0.99*	0.94‡	*0.83*‡§
tv(12,1)-trig-arma(1,1)	0.7‡§	1.03	0.92‡§	0.8‡§	0.81‡§	0.94§		1.08	0.9‡§	0.67‡§	0.92‡§	0.98	0.94‡	0.84‡§
tv(6,2)-trig-arma(1,1)	0.72‡§	1	0.91‡§	0.82‡§	0.82‡§	0.93§		1.08	0.94‡§	0.67‡§	0.93‡§	0.99	0.95*	0.84‡§
tv(12,2)-trig-arma(1,1)	0.72‡§	1.05‡	0.92‡§	0.82‡§	0.81‡§	0.92§		1.07	0.91‡§	0.67‡§	0.93‡§	0.99	0.97*	0.85‡§

†The results are relative to the forecast MSFE of the AR(1) benchmark model. See Table 4 for a description of the trigger variables. In italics and underlined are the best performing methods, when those are better than the benchmark.
‡Statistically different forecasts from the AR(1) model at the 5% significance level according to the forecast fluctuation test.
§Statistically different forecasts from the AR(1) model at the 5% significance level according to the Diebold and Mariano test.
§§Statistically different forecasts from the AR(1) model at the 10% significance level according to the Diebold and Mariano test.
*Statistically different forecasts from the AR(1) model at the 10% significance level according to the forecast fluctuation test.

Table 8. MSFE results for the recession period (see Tables 1 and 3 for a description of the models and the series)†

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Recession period forecasting														
arma(1,1)	0.51	1.01	0.66	0.77	0.92	0.83	1.02	*0.8*	0.93	0.76	0.85	1	1.2	0.95
arma(bic)	1	1.04	1	0.77	1.12	0.86‡	1.01	1.14	1	0.76§	0.81	1	1.15	1
Trigger variable: quarter-on-quarter difference of the forecasted series														
tv-N-CV(6)-trig-arma(1,1)	0.46‡	1.02	0.62	0.6	3.6	0.9	1.07	0.96	0.96	0.64§	0.85	0.94	0.89	
tv-N-CV(12)-trig-arma(1,1)	0.46‡	0.97	0.54	*0.59*	3.6	0.95	1.04	0.95	0.92	0.66§	0.85	0.93	0.92	
tv-N-CV(6,1)-trig-arma(1,1)	0.43‡	0.96	0.53‡	1.05	1.03‡	0.87‡	1.06	0.99	0.85§	0.64§	0.89	0.82	*0.85*	
tv-N-CV(12,1)-trig-arma(1,1)	0.45‡	0.96	0.52‡	1.09	0.99‡	0.87‡	1.06	1.15	*0.84*	0.65	0.91§	0.85	0.99	
tv-N-CV(6,2)-trig-arma(1,1)	0.44‡	0.94§	0.57	1.18	0.95‡	0.83	1.08	0.99	0.89	0.63§	0.94§	*0.8*	0.87	
tv-N-CV(12,2)-trig-arma(1,1)	0.45‡	*0.93*‡	0.51§	0.81	0.93‡	0.81‡	1.14	0.95	0.87	0.7§	0.88	0.83	0.98	
Trigger variable: year-on-year difference of the forecasted series														
tv-N-CV(6)-trig-arma(1,1)	0.45‡	1.06	0.59	0.81	0.98	0.98‡	1.12	1.38	1.11	0.72§	0.82‡	1.37	6.34‡	
tv-N-CV(12)-trig-arma(1,1)	0.52‡	1.03	0.6	0.82	1	0.99§	1.2	1.08	1.08	0.66§	0.87	1.78	6.37	
tv-N-CV(6,1)-trig-arma(1,1)	0.42‡	0.94§	0.62§	1.01	0.93	0.86‡	1.03	1.02	0.96	0.73§	0.94‡	0.94	1.26	
tv-N-CV(12,1)-trig-arma(1,1)	0.44‡	0.93‡	0.57	1	0.96§	0.9‡	1.17	1.03	0.95	0.71§	0.87‡	1.06	1.24	
tv-N-CV(6,2)-trig-arma(1,1)	0.43‡	0.98	0.61	1.02	1§	0.99‡	1.49	1.04	0.9	0.67§	0.85‡	0.85	1.36	
tv-N-CV(12,2)-trig-arma(1,1)	0.42‡	0.97	0.62	1.03	0.98§	0.89‡	1.02	1.04	0.93§	0.68§	0.88‡	0.91	1.18	
Trigger variable: same as the forecasted series														
tv-N-CV(6)-trig-arma(1,1)	0.53‡	1.02	0.63	1.85	1.21	0.85	1.12	1.38	0.91	0.68	0.88	0.86	1.03	1.11
tv-N-CV(12)-trig-arma(1,1)	0.53‡	1.01	0.68	0.95	1.19	0.88	1.04	1.23	0.92	0.65§	0.87‡	0.88	1.12	1.12
tv-N-CV(6,1)-trig-arma(1,1)	0.59‡	1	0.64‡	1.32	0.94	0.89	1.15	0.98	0.94	0.69	0.94	0.85	1.2	0.94
tv-N-CV(12,1)-trig-arma(1,1)	0.59‡	1.05	0.63§	0.84	0.95	0.87	1.16	1	0.9	0.64§	0.92	0.85	1.22	0.94
tv-N-CV(6,2)-trig-arma(1,1)	0.56‡	1.01	0.68	1.34	0.99	1.29	1.08	0.99	0.93	0.65§	0.99	0.86	1.23	1.01
tv-N-CV(12,2)-trig-arma(1,1)	0.52‡	1	0.66	0.82	1.03	0.84	1.1	1.07	0.91	*0.62*§	0.95	0.9	1.27§	0.93

(*continued*)

Table 8 (*continued*)

Model	*Results for the following series:*													
	PAYEMS	*TB3MS*	*UNRATE*	*M1SL*	*CPI*	*INDPRO*	*FFR*	*MZMSL*	*CONSPI*	*AVGHE*	*CUMFNS*	*RPI*	*WPSID*	*HOUST*
Trigger variable: oil price, 2nd difference, month on month														
tv-N-CV(6)-trig-arma(1,1)	0.54‡	1.06	0.75	0.92	0.81	1.18‡	1.04	1.09	0.94	0.65§	0.81‡	1.01	1.04‡	1
tv-N-CV(12)-trig-arma(1,1)	0.57‡	1.06	0.74	0.92	0.81	1.17‡	1.03	1.8	0.94	0.65§	0.85‡	1.08	1.05‡	1
tv-N-CV(6,1)-trig-arma(1,1)	0.53‡	1.04	0.69§	1.01	1	0.83§	1.03	1.04	0.89	0.65§	0.87§	1	1.11	0.92
tv-N-CV(12,1)-trig-arma(1,1)	0.55‡	0.99	0.69§	0.95	0.97	0.86	1.03	1.18	0.91§	0.65§	0.88§	1.05	1.1	0.92
tv-N-CV(6,2)-trig-arma(1,1)	0.49‡	1.05	0.72	0.89	0.8	0.8‡	1.01	1.04	0.91§	0.65§	*0.76*‡	0.99	1.09	0.93
tv-N-CV(12,2)-trig-arma(1,1)	0.48‡	0.98	0.73‡	0.85	*0.79*	*0.78*	1.06	1.06	0.92§	0.66§	0.83	0.98	1.13	0.93
Trigger variable: housing starts, level														
tv-N-CV(6)-trig-arma(1,1)	0.4‡	2.3	0.49‡	1.48	1.15	0.96‡	1.05	1.06	1.18	0.84‡	0.78‡	1.04	2.79	0.99
tv-N-CV(12)-trig-arma(1,1)	0.42‡	1.16	*0.47*‡	1.48	1.15	0.94‡	1.04	1.2	1.06	0.85§	0.85‡	1.04	2.82	1.01
tv-N-CV(6,1)-trig-arma(1,1)	0.36‡	1.38	0.5§	1.2	1.33‡	0.82‡	1.04	4.86	0.99	0.66	0.88§	1.01	1.27	0.96
tv-N-CV(12,1)-trig-arma(1,1)	*0.34*‡	1.4	0.56	20.62	1.16‡	0.81‡	1.05	0.93	0.93	0.63§	0.79‡	0.99	1.35	0.93
tv-N-CV(6,2)-trig-arma(1,1)	0.34‡	1.04	0.57	1.25	1.18‡	0.95	1.26	1.02	0.93	0.71§	0.91	1.09	1.02	0.96
tv-N-CV(12,2)-trig-arma(1,1)	0.34‡	1.02	0.52	0.93	1.05‡	0.87§	1.04	0.97	0.91	0.73‡	0.92	1.03	1.05	0.98
Trigger variable: Fed funds rate, quarter on quarter														
tv-N-CV(6)-trig-arma(1,1)	0.48‡	1.04	0.74	0.92	1.02§		1.07	1.07	0.9	0.71‡	0.87§	0.99	1.26	1.09
tv-N-CV(12)-trig-arma(1,1)	0.5‡	1	0.71	0.9	1‡		1.04	1.11	0.91	0.93‡	0.87	0.97	1.24	1.09
tv-N-CV(6,1)-trig-arma(1,1)	0.52‡	0.96	0.76	0.88	0.97‡		1.06	1	0.92	0.66§	0.92	0.98	1.17	0.93
tv-N-CV(12,1)-trig-arma(1,1)	0.53‡	0.97	0.76	0.8	0.96‡		1.06	1.01	0.91§	0.67§	0.9	0.97	1.17	0.95
tv-N-CV(6,2)-trig-arma(1,1)	0.56‡	0.93	0.77‡	0.82	0.97‡		1.08	0.98	0.9§	0.65§	0.89	0.97	1.08	*0.91*§§
tv-N-CV(12,2)-trig-arma(1,1)	0.53‡	1.05§	0.81§	0.83	0.96‡		1.14	0.97	0.91	0.65§	0.9	0.97	1.14	0.93

†The results are relative to the forecast MSFE of the AR(1) benchmark model. See Table 4 for a description of the trigger variables. In italics and underlined are the best performing methods, when these are better than the benchmark.
‡Statistically different forecasts from the AR(1) model at the 5% significance level according to the Diebold and Mariano test.
§Statistically different forecasts from the AR(1) model at the 10% significance level according to the Diebold and Mariano test.

for a range of models relative to that of an AR(1) benchmark model. Additionally we indicate statistical significance at the 5% and 10% levels for the Diebold and Mariano, and Giacomeni and Rosso tests, against the AR(1) model. For ease of exposition, Tables 5–8 report the best method in italics and underlined, when this is better than the benchmark. In Tables 5 and 7 the (relative) MSFEs are computed over the whole evaluation period (1980, month 1–2017, month 4), whereas in Tables 6 and 8 MSFEs are reported only for the recession periods as these are defined by the National Bureau of Economic Research. In Tables 5 and 6 the regressors in all models are the lagged value of the dependent variable y_{t-1} and an intercept. This is to make the differences with respect to the benchmark AR(1) model dependent only on parameter time variation, modelled either via the three similarity approaches (tv-trig, SLA and the similarity threshold model), or MS, or continuous time variation (TVP). We recall here that SLA can be also interpreted as an NN forecast, and the similarity threshold model as a locally weighted regression forecast. To allow for more complex dynamics, Tables 7 and 8 present tv-trig ARMA(1,1) models. It is well known that fixed parameter ARMA(1,1) models typically forecast very well macroeconomic and financial indicators, and we expect that extending them with our proposals can provide further benefits. Additionally, as also shown in our empirical exercise, the ARMA(1,1) model is almost always better than the ARMA(p,q) model, with p and q selected by the BIC. For the similarity approaches we also experiment with the two cross-validation criteria that were described in Section 2, i.e. end-of-sample and clustered cross-validation. The precise specification of the models that are reported in Tables 5–8 can be inferred from the model classification in Table 1, whereas the names for the series are explained in Table 3.

Some comments can be made based on the empirical results. First, starting with Table 5, one of the similarity approaches produces the lowest MSFE for 10 out of the 14 variables under analysis (PAYEMS, TB3MS, UNRATE, M1SL, INDPRO, FFR, MZMSL, CONSPI, AVGHE, RPI and WPSID). Yet, for PAYEMS and INDPRO the performance is the same as that of AR(BIC), and this model is competitive for several other variables, probably because a larger number of AR lags can try to capture the spurious persistence that is generated by unmodelled parameter changes. Moreover, for UNRATE our similarity proposals perform equivalently to that of the best procedure (MS2AR). More generally, and in line with the simulation results, the MSFE-gains from the similarity approaches are not large, though they are quite systematic. It is also worth mentioning that the quarter-on-quarter difference of the target series is the best performing trigger variable in most cases. Moreover, for this trigger variable, clustered cross-validation is slightly but systematically better than end-of-sample cross-validation.

Second, from Table 6, the forecasting gains from the use of parameter time variation in the AR(1) model generally increases during recessionary periods, much more so for real variables such as PAYEMS, UNRATE and INDPRO than for nominal variables and interest rates. In terms of ranking of the various types of models, similarity approaches remain best for 10 out of 14 variables, but now SLA is best for four of the 11 variables (PAYEMS, CPI, FFR and CONSPI) and tv-trig for seven of them. The TVP model becomes best instead for the three remaining variables (UNRATE, INDPRO and HOUST). Moreover, the quarter-on-quarter difference of the target series remains overall the best performing trigger variable, with HOUST as second best, which highlights the importance of this variable during recessionary periods. Again, clustered cross-validation is generally better than end-of-sample cross-validation for the trig models.

Third, from Table 7, adding an MA(1) component to the constant parameter (benchmark) AR(1) model is helpful for 12 of the 14 variables and comparable for the remaining two variables. BIC-based lag selection is basically never helpful in the ARMA case. The ARMA(1,1) model has the same or lower MSFE than the ARMA(BIC) model for all variables, whereas it lowered the MSFE for 11 of the 14 variables in the case of a pure AR model (see Table 5). This suggests that, once an MA component has been included in the model, longer AR lags are no longer needed to capture unmodelled parameter breaks. Extending the ARMA models with the similarity-based, specifically trig, approaches provides advantages over the fixed parameter ARMA model for 10 out of the 14 variables. Comparing tv-trig ARMA models with the AR(1) benchmark shows that our proposals perform better in 12 out of the 14 cases that were examined. This leads us to believe that extending ARMA models with our approaches can impact the out-of-sample forecast. The most significant improvement for our proposals over the ARMA models is considered for the series PAYEMS, UNRATE, CONSPI, RPI and WPSID. Now, the preferred trigger variable and the cross-validation method depend more on the series of interest. Yet, for the trig models there are in general small gains from adding the MA component, probably because they already take into consideration parameter time variation. The three variables where the gains from adding the MA term are sizable are M1SL, CPI and AVGHE.

Finally, from Table 8, the forecasting gains from using an ARMA(1,1) instead of an AR(1) model are generally larger during recessionary periods, but much more so for real variables such as PAYEMS, UNRATE and INDPRO. One of the tv-trig ARMA models is better than the benchmark (and the fixed parameter ARMA models) for 12 out of 14 variables but, as noted before, the gains from adding the MA component in tv-trig methods are small.

It becomes clear that in practice the forecaster faces several modelling choices. Based on our forecasting exercises, our baseline recommendation is to use a smoothed transformation of the forecasted variable as a trigger indicator, with the clustered cross-validation scheme to choose the bandwidth parameter. This seems a robust and easy-to-implement, short-term forecasting approach. Alternatively, the researcher can always use a preselected window of data to set these specification choices optimally.

Overall our methods can provide slight, but systematic, improvements over the benchmarks considered. Nevertheless, we need to highlight that these are not always statistically different from the AR(1) model. For instance in Table 5 our proposed and best performing methods are statistically different from the AR(1) forecasts for five out of 14 cases, in Table 6, in one case, in Table 7, in nine cases, and, in Table 8, in nine cases. This is an indication that, although we can establish improvements in terms of RMSE in many cases, it is sometimes difficult to establish statistical significance.

Our empirical exercise also highlights the importance of parameter changes for modelling and forecasting macroeconomic variables. The forecasting performance of the fixed parameter AR(1) model can be generally improved by either adding more lags or an MA component, which both capture additional persistence that can be either real or due to unaccounted parameter changes. Evidence in favour of the latter option is provided by the better forecasting performance of models with TVPs. Within this class, our newly proposed similarity-based methods behave well, and the trigger-based models particularly well, even without an MA component, probably because of their flexible non-parametric accounting of parameter evolution, that filters snapshots (periods) of data which are more suitable for estimation and forecasting under the economic conditions maintained. Yet, a careful choice of the trigger variable is required, combined with cross-validation for the selection of the tuning parameters.

6. Conclusions

In this paper we propose similarity-based approaches for macroeconomic forecasting. The basic idea is to overweight periods that are similar to the current period and to downweight the rest of the sample, when estimating the model parameters, to be used later to construct the forecast. The weighting is based on the behaviour of trigger variables, combined with a non-parametric kernel estimator. Lags of the dependent variable or other exogenous variables can be used as triggers, possibly after some smoothing to amplify the signal.

Although our approach is related to existing methods such as NNs, we provide considerable extensions and refinements, such as combining similarity and time variation modelling, and introduce cross-validation methods to select tuning parameters.

Further, we assess the forecasting performance of our proposals both in Monte Carlo experiments and in an empirical application based on a set of key US macroeconomic and financial indicators, also in comparison with common competing time-varying specifications, such as MS and TVP models.

Overall, the similarity approach is promising, even though the forecasting gains with respect to existing methods are not uniform across variables and require a careful specification search, including the choice of the proper trigger variable and tuning parameters. Yet, with a careful specification we can almost always do better, in an MSFE sense, than ARMA or competing TVP models.

Acknowledgements

We thank the Joint Editor two referees, Todd Clark, Frank Diebold, Pablo Guerron Quintana, Mike McCracken and Molin Zhong for helpful comments on a previous draft.

The research work was supported by the Hellenic Foundation for Research and Innovation under the ‘First call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant’ (project 11320601).

References

Bierens, H. J. (1987) Kernel estimators of regression functions. In *Advances in Econometrics: Fifth World Congress* (ed. T. F. Bewley). Cambridge: Cambridge University Press.

Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.

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

Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *J. Am. Statist. Ass.*, **74.**, 829–836.

Cleveland, W. S. and Devlin, S. J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Statist. Ass.*, **83**, 596–610.

Cogley, T. and Sargent, T. J. (2005) The conquest of US inflation: learning and robustness to model uncertainty. *Rev. Econ. Dyn.*, **8**, 528–563.

Dendramis, Y., Kapetanios, G. and Tzavalis, E. (2015) Shifts in volatility driven by large stock market shocks. *J. Econ. Dyn. Control*, **55**, 130–147.

Diebold, F. X. and Mariano, R. S. (1995) Comparing predictive accuracy. *J. Bus. Econ. Statist.*, **13**, 253–263.

Diebold, F. X. and Nason, J. A. (1990) Nonparametric exchange rate prediction? *J. Int. Econ.*, **28**, 315–332.

Dufour, J. M. and Pelletier, D. (2008) Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application. *Discussion Paper*. McGill University, Montreal.

Eklund, J., Kapetanios, G. and Price, S. (2010) Forecasting in the presence of recent structural change. *Working Paper 406*. Bank of England, London.

Ferrara, L., Marcellino, M. and Mogliani, M. (2015) Macroeconomic forecasting during the Great Recession: the return of non-linearity? *Int. J. Forecast.*, **31**, 664–679.

Ghysels, E. and Marcellino, M. (2018) *Applied Economic Forecasting using Time Series Methods*. Oxford: Oxford University Press.

Giacomini, R. and Rossi, B. (2010) Forecast comparisons in unstable environments. *J. Appl. Econmetr.*, **25**, 595–620.

Gilboa, I., Lieberman, O. and Schmeidler, D. (2011) A similarity-based approach to prediction. *J. Econmetr.*, **162**, 124–131.

Giraitis, L., Kapetanios, G. and Price, S. (2013) Adaptive forecasting in the presence of recent and ongoing structural change. *J. Econmetr.*, **177**, 153–170.
Giraitis, L., Kapetanios, G. and Yates, T. (2018) Inference on multivariate heteroscedastic time varying random coefficient models. *J. Time Ser. Anal.*, **39**, 129–149.
Gonzalo, J. and Pitarakis, J. Y. (2002) Estimation and model selection based inference in single and multiple threshold models. *J. Econmetr.*, **110**, 319–352.
Guerron-Quintana, P. and Zhong, M. (2017) Macroeconomic forecasting in times of crises. *Working Paper*.
Hamilton, J. (1989) A new approach to the economic analysis of non-stationary time series and business cycle. *Econometrica*, **57**, 357–384.
Hansen, P. R., Lunde, A. and Nason, J. M. (2011) The model confidence set. *Econometrica*, **79**, 453–497.
Hendry, D. F. (2000) On detectable and non-detectable structural change. *Struct. Change Econ. Dyn.*, **11**, 45–65.
Kapetanios, G. and Tzavalis, E. (2010) Modeling structural breaks in economic relationships using large shocks. *J. Econ. Dyn. Control*, **34**, 417–436.
McCracken, M. W. and Ng, S. (2016) FRED-MD: a monthly database for macroeconomic research. *J. Bus. Econ. Statist.*, **34**, 574–589.
Medeiros, M. C., Vasconcelos, G., Veiga, A. and Zilberman, E. (2019) Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *J. Bus. Econ. Statist.*, to be published.
Nicholls, D. F. and Pagan, A. R. (1985) Varying coefficient regression. *Handb. Statist.*, **5**, 413–449.
Pagan, A. and Ullah, A. (1999) *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
Pesaran, M. H., Pick, A. and Pranovich, M. (2013) Optimal forecasts in the presence of structural breaks. *J. Econmetr.*, **177**, 134–152.
Pesaran, M. H., Pick, A. and Timmermann, A. (2011) Variable selection, estimation and inference for multiperiod forecasting models. *J. Econmetr.*, **164**, 173–187.
Potter, S. (2011) The failure to forecast the great recession. *Blog*. Liberty Street Economics.
Primicery, G. (2005) Time varying structural vector autoregressions and monetary policy. *Rev. Econ. Stud.*, **72**, 821–852.
Robinson, M. (1983) Non parametric estimators for time series. *J. Time Ser. Anal.*, **4**, 185–207.
Stock, J. H. and Watson, M. (1996) Evidence on structural instability in macroeconomic time series relations. *J. Bus. Econ. Statist.*, **14**, 11–30.
Stock, M. M. J. H. and Watson, M. W. (2006) A comparision of direct and iterated multistep AR methods for forecasting macroeconomic time series. *J. Econmetr.*, **135**, 499–526.
Terasvirta, T. (1998) Modeling economic relationships with smooth transition regressions. In *Handbook of Applied Economic Statistics* (eds A. Ullah and D. E. Giles). New York: Dekker.
Tong, H. (1990) *Non-linear Time Series: a Dynamical System Approach*. Oxford: Oxford University Press.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Online appendix to “A similarity-based approach for macroeconomic forecasting”’.