Time Varying Dimension Models

Joshua C.C. Chan
Australian National University

Gary Koop
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies

Rodney W. Strachan
Australian National University

January 3, 2012

Abstract: Time varying parameter (TVP) models have enjoyed an increasing popularity in empirical macroeconomics. However, TVP models are parameter-rich and risk over-fitting unless the dimension of the model is small. Motivated by this worry, this paper proposes several Time Varying Dimension (TVD) models where the dimension of the model can change over time, allowing for the model to automatically choose a more parsimonious TVP representation, or to switch between different parsimonious representations. Our TVD models all fall in the category of dynamic mixture models. We discuss the properties of these models and present methods for Bayesian inference. An application involving US inflation forecasting illustrates and compares the different TVD models. We find our TVD approaches exhibit better forecasting performance than many standard benchmarks and shrink towards parsimonious specifications.

Acknowledgements: Koop, Leon-Gonzalez and Strachan are Fellows of the Rimini Centre for Economic Analysis. We would like to thank the Economic and Social Research Council and the Australian Research Council for financial support under Grant RES-062-23-2646 and Grant DP0987170, respectively. Contact address: Gary Koop, Department of Economics, Strathclyde University, Glasgow, G4 0GE, United Kingdom, Gary.Koop@strath.ac.uk

1 Introduction

It is common for researchers to model variation in coefficients in time series models using state space methods. If, for $t = 1, \ldots, T$, y_t is an $n \times 1$ vector of observations on the dependent variables, Z_t is an $n \times m$ matrix of observations on explanatory variables and θ_t is an $m \times 1$ vector of states, then such a state space model can be written as:

$$\begin{aligned} y_t &= Z_t \theta_t + \varepsilon_t \\ \theta_{t+1} &= \theta_t + \eta_t, \end{aligned} \tag{1}$$

where ε_t is $N(0, H_t)$ and η_t is $N(0, Q_t)$. The errors, ε_t and η_t, are assumed to be independent (at all leads and lags and of each other). This framework can be used to estimate time-varying parameter (TVP) regression models, variants of which are commonly-used in macroeconomics (e.g., Groen, Paap and Ravazzolo, 2010, Koop and Korobilis, 2011). Furthermore, TVP-VARs (see among many others, Canova, 1993, Cogley and Sargent, 2005, D'Agostino, Gambetti and Giannone, 2009 and Primiceri, 2005) are obtained by letting Z_t contain deterministic terms and appropriate lags of the dependent variables, setting $Q_t = Q$ and giving H_t a multivariate stochastic volatility form.

Such TVP models allow for constant gradual evolution of parameters. However, they assume that the dimension of the model is constant over time in the sense that θ_t is always an $m \times 1$ vector of parameters. But there are several reasons for being interested in TVP models where the dimension of the state vector changes over time. Recent papers have found that the set of predictors for inflation can change over time or over the business cycle. For instance, Stock and Watson (2009) state that: "one of our key findings is that the performance of Phillips curve forecasts is episodic: there are times, such as the late 1990s, when Phillips curve forecasts improved upon using univariate forecasts, but there are other times (such as the mid-1990s) when a forecaster would have been better off using a univariate forecast" (page 100). A subsequent paper, Stock and Watson (2010), provides detailed evidence relating to a range of predictors for inflation and find that most of them improve forecast performance only in some episodes (e.g. it is only in economic downturns that forecast improvements occur when an unemployment recession gap variable is included). Indeed words like "episodic" and "instability" occur frequently in discussions of whether specific predictors help with forecasting inflation.

Similarly, macroeconomists are often interested in whether restrictions suggested by economic theory hold. For instance, Staiger, Stock and Watson (1997) show how, if the Phillips curve is vertical, a certain restriction is imposed on a particular regression involving inflation and unemployment. Koop, Leon-Gonzalez and Strachan (2010) investigate this restriction in a TVP regression model and find that the probability that it holds varies substantially over time. As another example, consider the VARs of Amato and Swanson (2001) where interest centres on Granger causality restrictions that imply that money has no predictive power for output or inflation. It is possible (and empirically likely) that restrictions such as these hold at some points in time but not others.

In cases such as those discussed above, the researcher would want to work with a TVP model, but where the parameters satisfy restrictions at certain points in time but not at others. To be precise, it is potentially desirable to develop a statistical approach which can formally model when (and if) explanatory variables enter or leave a regression model (or multivariate extension such as a VAR).

More broadly, there are many situations where the researcher is working with a model with high-dimensional parameter space and, worried about over-fitting problems, wishes to induce parsimony. In general, TVP models risk being over-parameterized. Allowing for the dimension of the model to change over time is potentially an effective way of reducing over-parameterization worries and ensuring shrinkage while minimizing the risk of model misspecification A simple example might be a wish to allow for lag length to change over time, then imposing different lag lengths at different points in time will lead to more precise estimates.

In short, there are many theoretical reasons for wanting to work with a time-varying dimension (TVD) model where restrictions which reduce the dimension of the model are imposed only at some points in time. The purpose of the present paper is to develop such a model.

The desire to work with a TVP model that falls into the familiar class of state space models, but allows for the dimension of the model to change over time motivates the present paper. To our knowledge, there are no existing papers in the econometric literature which consider this precise question. In the next section, the related literature will be discussed. Here we note that there are, as discussed above, many papers which allow parameters to change over time and adopt state space methods. However, this kind of papers does not allow for the dimension of the parameter space to change over time. Furthermore, in previous work (Koop, Leon-Gonzalez and Strachan, 2010), we have developed methods for calculating the probability that equality restrictions on states hold at any point in time (but without actually imposing the restrictions). Finally, there are some papers, such as Koop and Potter (2011), which develop methods for estimating state space models with inequality restrictions imposed. However, the aim of the present paper is different from all these approaches: we wish to develop methods for estimating models which impose equality restrictions on the states. In other words, the related econometric literature has considered the *testing* of *equality* restrictions on states in state space models and *estimation* of states under *inequality* restrictions. But the present paper is one of the few which considers *estimation* of state space models subject to *equality* restrictions on the states (where these restrictions may hold at some points in time but not others).

An advantage of the TVD models developed in this paper is that they are all dynamic mixture models (see, e.g., Gerlach, Carter and Kohn, 2000) and, thus, well-developed posterior simulation algorithms exist for estimating these models. Such models have proved popular in several areas of macroeconomics (e.g. Giordani, Kohn and van Dijk, 2007). We consider several new ways of implementing the dynamic mixture approach which lead to models which allow for time-variation in both the parameters and the dimension of the model. We investigate these methods in an empirical application involving forecasting US inflation.

2 Time Varying Dimension Models

The dynamic mixture model of Gerlach, Carter and Kohn (2000) is a very general type of state space model which can be used for many purposes. Gerlach, Carter and Kohn (2000) derive an efficient algorithm for posterior simulation in this model. Dynamic mixture models have been used for many purposes. For instance, Giordani, Kohn and van Dijk (2007) use them for modelling outliers and nonlinearities in economic time series models. Giordani and Kohn (2008) use them to model structural breaks and parameter change in univariate time series models and Koop, Leon-Gonzalez and Strachan (2009) use them to induce parsimony in TVP-VARs. All of these approaches, however, focus on parameter change. The contribution of the present paper lies in using the dynamic mixture model framework to allow for model change (in the sense that the dimension of the model can change over time).

An advantage of adopting a dynamic mixture framework is that efficient methods of posterior simulation are available and well-understood. Thus, our discussion of Bayesian inference in these models can be very brief. It is the structure and justification for how dynamic mixture methods can be used to produce TVD models that must be provided and this is what we do in this section.

2.1 Using the Dynamic Mixture Approach to Create a TVD Model

The dynamic mixture model of Gerlach, Carter and Kohn (2000) adds to (1) the assumption that any or all of the system matrices, Z_t, Q_t and H_t, depend on an $s \times 1$ vector K_t. Gerlach, Carter and Kohn (2000) discuss how this specification results in a mixtures of Normals representation for y_t and, hence, the terminology dynamic mixture model arises. The contribution of Gerlach, Carter and Kohn (2000) is to develop an efficient algorithm for posterior simulation for this class of models. The efficiency gains occur since the states are integrated out and $K = (K_1, \ldots, K_T)'$ is drawn unconditionally (i.e. not conditional on the states). A simple alternative algorithm would involve drawing from the posterior for K conditional on $\theta = (\theta_1', \ldots, \theta_T')'$ and then the posterior for θ conditional on K. Such a strategy can be shown to produce a chain of draws which is very slow to mix. The Gerlach, Carter and Kohn (2000) algorithm requires only that K_t be Markov (i.e. $p(K_t|K_{t-1}, \ldots, K_1) = p(K_t|K_{t-1})$) and is particularly simple if K_t is a discrete random variable.

In this paper, we consider three different ways K_t can enter the system matrices so as to yield a TVD model. We begin with a TVD model which adapts the approach of Gerlach, Carter and Kohn (2000) in a particular way such that θ_t remains an $m \times 1$ vector at all times, but there is a sense in which the dimension of the model can change over time. Since θ_t remains of full dimension at all times, our claim that the dimension of the model changes over time may sound odd. But we achieve our goal by allowing for explanatory variables to be included/excluded from the likelihood function depending on K_t. The basic idea can be illustrated quite simply in terms of (1). Suppose $Z_t = K_t z_t$ where z_t is an explanatory variable and $K_t \in \{0, 1\}$. If $K_t = 0$ then z_t does not enter the likelihood function and the coefficient θ_t does not enter the model. But if $K_s = 1$, then the coefficient θ_s does enter the model. Thus, the dimension of the model is different at time t than at time s.

An interesting and sensible implication of this specification can be seen by considering what happens if a coefficient is omitted from the model for h periods, but then is included again. That is, suppose we have $K_{t-1} = 1$,

$$K_t = K_{t+1} = \cdots = K_{t+h-1} = 0$$

but $K_{t+h} = 1$ and further assume $Q_t = Q$. Then (1) implies:

$$E(\theta_{t+h}) = \theta_{t-1}$$

but

$$var(\theta_{t+h}) = hQ.$$

In words, if an explanatory variable drops out of the model, but then reappears h periods later, then your best guess for its value is what it was when it was last in the model. However, the uncertainty associated with your best guess increases the longer the coefficient has been excluded from the model (since the variance increases with h).

It is worth stressing that, if $K_t = 0$ then θ_t does not enter the likelihood and, thus, it is not identified in the likelihood. However, because the state equation provides an informative hierarchical prior for θ_t, it will still have a proper posterior. To make this idea clear, let us revert to a general Bayesian framework. Suppose we have a model depending on a vector of parameters θ which are partitioned as $\theta = (\phi, \gamma)$. Suppose the prior is $p(\theta) = p(\phi, \gamma) = p(\gamma)\, p(\phi|\gamma)$ and the likelihood is $L(y|\theta)$. Now consider a second model which imposes the restriction that $\phi = 0$. Instead of directly imposing the restriction $\phi = 0$, consider what happens if we impose the restriction that ϕ does not enter the likelihood. That is, the likelihood for the second model is $L(y|\theta) = L(y|\gamma)$ and its posterior is

$$p(\theta|y) = \frac{L(y|\theta)\, p(\theta)}{\int L(y|\theta)\, p(\theta)\, d\theta} = \frac{L(y|\gamma)\, p(\gamma)}{\int L(y|\gamma)\, p(\theta)\, d\theta} p(\phi|\gamma) = p(\gamma|y)\, p(\phi|\gamma).$$

Since $p(\phi|\gamma)$ integrates to one (or assigns a point mass to $\phi = 0$) integrating $p(\theta|y)$ with respect to ϕ provides us with a valid posterior for the second model and the integral $\int L(y|\gamma)\, p(\theta)\, d\theta$ will result in the correct marginal likelihood. This is the strategy which underlies and justifies our approach.

To explain our second approach to TVD modelling, we return to our general notation for state space models given in (1). The state equation can be interpreted as a hierarchical prior for θ_{t+1}, expressing a prior belief that it is similar to θ_t. In the empirical macroeconomics literature (see, among many others, Ballabriga, Sebastian and Valles, 1999, Canova and Ciccarelli, 2004, and Canova, 2007), there is a desire to combine such prior information with prior information of other sorts (e.g. the Minnesota prior). This can be done by replacing (1) by

$$\begin{aligned} y_t &= Z_t \theta_t + \varepsilon_t \\ \theta_{t+1} &= M \theta_t + (I - M) \overline{\theta} + \eta_t, \end{aligned} \tag{2}$$

where M is an $m \times m$ matrix, $\overline{\theta}$ is an $m \times 1$ vector and η_t is $N(0, Q_t)$. For instance, Canova (2007) sets $\overline{\theta}$ and Q_t to have forms based on the Minnesota prior and sets $M = gI$ where g is a scalar. If $g = 1$, then the traditional TVP-VAR prior is obtained, but as g decreases we move towards the Minnesota prior.

In the case of the TVD model, alternative choices for M, $\overline{\theta}$ and Q_t suggest themselves. In particular, our second TVD model sets $\overline{\theta} = 0_m$, M becomes M_t which is a diagonal matrix with diagonal elements $K_{tj} \in \{0, 1\}$ and $Q_t = M_t Q$. This model has the property that, if $K_{jt} = 1$ then the j^{th} coefficient is evolving according to a random walk in standard TVP-regression fashion. But if $K_{jt} = 0$, then the j^{th} coefficient is set to zero, thus reducing the dimension of the model.

To understand the implications of this specification for K_t, consider the illustration above where $m = 1$ and, thus θ_t and K_t are scalars and see what happens if a coefficient is omitted from the model for h periods. That is, suppose we have $K_{t-1} = 1$,

$$K_t = K_{t+1} = \cdots = K_{t+h-1} = 0$$

but $K_{t+h} = 1$. In this case, (2) implies:

$$E(\theta_{t+h}) = \overline{\theta}$$

but

$$var(\theta_{t+h}) = Q.$$

In words, in contrast to our first TVD model, our second TVD model implies that, if a coefficient drops out of the model, but then reappears h periods later, then your best guess for its value is 0 and the uncertainty associated with your best guess is Q (regardless of how long the coefficient has been excluded from the model). Thus, there is more shrinkage in this model than in our first TVD model and (in contrast to the first TVD model) it will always be shrinkage towards zero.

To justify our third approach to TVD modelling, we begin by discussing the TVP-SUR approach of Chib and Greenberg (1995) which has been used in empirical macroeconomics in papers such as Ciccarelli and Rebucci (2002). If we return to our general notation for state space models in (1), the model of Chib and Greenberg (1995) adds another layer to the hierarchical prior:

$$\begin{aligned} y_t &= Z_t \theta_t + \varepsilon_t \\ \theta_{t+1} &= M\beta_{t+1} + \eta_t, \\ \beta_{t+1} &= \beta_t + u_t. \end{aligned} \quad (3)$$

where the assumptions about the errors are described after (1) with the additional assumptions that u_t is i.i.d. $N(0, R)$ and u_t is independent of the other errors in the model. Note that β_t can potentially be of lower dimension than θ_t, which is another avenue the researcher can use to achieve parsimony. However, if M is a square matrix, the hierarchical prior in (3) expresses the conditional prior belief that

$$E(\theta_{t+1}|\theta_t) = M\beta_t$$

and, thus, is a combination of the random walk prior belief of the conventional TVP model with the prior beliefs contained in M. Our third TVD model can be constructed by specifying M and Q_t to be exactly as in our second TVD model.

To understand the properties of the third TVD model, we can consider the same example as used previously (where a coefficient drops out of the model for h periods and then re-enters it). Remember that, in this case, the first TVD model implied $E(\theta_{t+h}) = \theta_{t-1}$ and $var(\theta_{t+h}) = hQ$ while the second TVD model implied $E(\theta_{t+h}) = 0$ and $var(\theta_{t+h}) = Q$. The third TVD model can be seen to have properties closer to those of the first approach and yields $E(\theta_{t+h}) = \beta_{t-1}$ and $var(\theta_{t+h}) = hR + Q$ (if M is a square matrix).

The first and third TVD models, thus, can be seen to have similar properties. However, they differ in one important way. Remember that the first TVD model did not formally reduce the dimension of θ_t in that all of its elements were unrestricted (it constructed K_t in such a way so that some elements of θ_t did not enter the likelihood function). The third TVD model does formally reduce the dimension of θ_t since it allows for some of its elements at some points of time to be restricted to zero.

These three different TVD models can be implemented with any choice of K_t. However, the approach can become computationally demanding if the dimension of K_t is large. Consider a TVD regression model with p predictors. It is tempting to simply let K_t be a vector of p dummy variables controlling whether each regressor is included or excluded in the model at time t. With this approach there are 2^p values K_t could take and, since the Gerlach, Carter and Kohn (2000) algorithm involves evaluating the posterior for K_t at each of these values, the computational demands will be high unless p is small. In our forecasting exercise $p = 14$ and such an approach is computationally infeasible. Accordingly, the researcher will typically seek to restrict the dimension of K_t or the number of values each K_t can take. For instance, in an AR(d) model an unrestricted approach would lead to K_t taking on 2^d possible values. But if we define a hierarchical prior which restricts K_t such that lags appear sequentially, this reduces to d the number of possible values K_t can take.

In our forecasting exercise, we only consider models with no predictors, a single predictor or all p predictors. More precisely, the vector $K_t = (K_{1,t}, \ldots, K_{p,t})$ can only take values in $\mathcal{I}$, where

$$\mathcal{I} = \{(0, 0, \ldots, 0), (1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1), (1, 1, \ldots, 1)\}.$$

In other words, K_t can take on $p + 2$ values. In addition, we impose a Markov hierarchical prior which expresses the belief that, with probability c the model will stay with its current set of explanatory variables and with probability $1 - c$ it will switch to a new model. A priori, all of the $p + 1$ possible new models are equally likely. Thus we have:

$$\begin{aligned} \Pr(K_{t+1} = i \,|\, K_t = i) &= c, & i \in \mathcal{I} \\ \Pr(K_{t+1} = j \,|\, K_t = i) &= \frac{1 - c}{p + 1}, & i \neq j, \quad i, j \in \mathcal{I} \end{aligned}$$

for $t = 1, \ldots, T - 1$.

2.2 Comparison with the Existing Literature

The TVD approach falls into the growing literature which seeks to place restrictions on TVP regression or TVP-VAR models in order to decrease worries associated with over-parameterization problems. That is, VARs and regression models with many predictors will have many parameters to estimate. For the macroeconomist with fairly small data sets, this can lead to the problems associated with over-fitting and poor forecast performance. Time-varying parameter versions of these models suffer from such over-parameterization problems to an even greater degree. The simplest way to treat over-parameterization problems is to set some of the parameters to zero. However, conventional sequential hypothesis testing procedures can run into pre-testing problems. Furthermore, it may be empirically desirable to have a parameter being zero at some points in time, but not at others and traditional hypothesis testing procedures do not allow for this. In theory, TVP models, by allowing a coefficient to be estimated as being near zero at some points in time, but not others, should be able to allow for the dimension of the model to change over time, at least approximately. However, in practice, this approximation can be poor and use of over-parameterized TVP models can lead to poor forecast performance (see the forecasting results in this paper or Koop and Korobilis, 2011).

These considerations have led to a growing literature which works with models with many parameters, but shrinking some of them towards zero to ensure parsimony. In the VAR literature, Banbura, Giannone and Reichlin (2010) is an important recent contribution which uses a conventional Bayesian prior to obtain such shrinkage. De Mol, Giannone and Reichlin (2008) is a similar example applied to a regression model with many predictors. However, there is also an increasing number of papers which use hierarchical priors (i.e. where shrinkage is done using priors, but the hyperparameters of the priors depend on unknown parameters which are estimated from the data). Examples include the stochastic search variable selection approach of George, Sun and Ni (2008), the variable selection approach of Korobilis (2011), the use of the dynamic mixture approach to model parameter change in Koop, Leon-Gonzalez and Strachan (2009) and the related approach of Groen, Paap and Ravazzolo (2010). The TVD approach can also be interpreted as being a hierarchical prior.

However, there are few papers which deal with model change (i.e. where the model dimension can be reduced or expanded over time by setting time-varying coefficients to zero) as opposed to parameter change (which empirically can only poorly approximate model change by allowing coefficients to be estimated as being approximately zero). The most relevant papers use dynamic model averaging, DMA (see, e.g., Raftery et al, 2010 or Koop and Korobilis, 2011). DMA does model averaging over a set of restricted TVP regression models. This set of restricted models includes every sub-set of the potential regressors in a model. By allowing for the weights used in the model averaging to vary over time, DMA allows for model change in the sense that the weight attached to each restricted regression model varies over time. However, the computational demands of DMA means that approximate Bayesian methods are used. Our TVD approaches can be thought of an alternative way of dealing with this problem which does not resort to approximations.

In sum, our TVD approach adds to the growing literature on ways of ensuring parsimony in potentially over-parameterized TVP regression models. It does so in a different way from existing approaches (other than DMA) in that it allows for the model dimension to change over time (as opposed to simply imposing shrinkage on parameters). In our empirical work presented below, we investigate whether this property of TVD improves forecasting performance and find evidence that it does.

We have in mind that TVD could be a useful approach in cases where the researcher has a potentially high-dimensional parameter space, such as arises in regressions with many potential explanatory variables or VARs with many variables or long lag length. The researcher wishes to allow for time-variation in parameters and thus want to use a TVP model. However, in such cases, most of the parameters in the model are typically zero, at least at some points in time. The trouble is that the researcher does not know which parameters are zero and at what time periods they are zero. It is tempting to do a sequential hypothesis testing strategy to seek a more parsimonious specification in such a case. A trouble is that conventional hypothesis testing procedures are often criticized when working with high-dimensional parameter spaces due to pre-testing problems. More substantively, conventional hypothesis tests are designed for constant parameter models so as to test whether a restriction holds for all time periods or never. In the TVP context, we want to allow for a restriction holding at some points in time but not at others and conventional hypothesis testing procedures are poorly designed to address this. Our suggested strategy is to work with the TVP model with high-dimensional parameter space, but use TVD methods to impose restrictions (in a time-varying manner) on the potentially over-parameterized TVP model.

2.3 Posterior Computation in the TVD Models

The advantage of the TVD modelling framework outlined in this paper is that existing methods of posterior computation can be used to set up a fast and efficient Markov Chain Monte Carlo (MCMC) algorithm. Thus we can deal with computational issues quickly. For all our models, K is drawn using the algorithm described in Section 2 of Gerlach, Carter and Kohn (2000). Note that this algorithm draws K conditional on all the model parameters except for θ. The fact that θ is integrated out analytically greatly improves the efficiency of the algorithm. We draw θ (conditional on all the model parameters, including K) using the algorithm of Chan and

Jeliazkov (2009), although any of the standard algorithms for drawing states in state space models (e.g. Carter and Kohn, 1994 or Durbin and Koopman, 2002) could be used. All our models have stochastic volatility and to draw the volatilities and all related parameters we use the algorithm of Section 3 of Kim, Shephard and Chib (1998). The remaining parameters are the error variances in the state equations and the parameters characterizing the hierarchical prior for K which have textbook posteriors (see, e.g., Koop, 2003). Since all of these posterior conditional distributions draw on standard results, we do not reproduce them here, but refer the reader to the online appendix to this paper for details. The online appendix also includes MCMC convergence diagnostics.

3 Forecasting US Inflation

To investigate the properties of the TVD models, we use a TVD regression model and investigate how the various approaches work in an empirical exercise involving US inflation forecasting. The literature on inflation forecasting is a voluminous one. Here we note only that there have been many papers which use regression-based methods in recursive or rolling forecast exercises (e.g. Ang, Bekaert and Wei, 2007 and Stock and Watson, 2007, 2009) and that recently papers have been appearing using TVP models for forecasting (e.g. D'Agostino, Gambetti and Giannone, 2009) to try and account for parameter change and structural breaks. We compare our TVD models to a variety of forecasting procedures commonly used in the literature including constant coefficient models, structural break models and TVP models.

3.1 Overview of Modelling Choices and Forecast Metrics

All of our models include at least an intercept plus two lags of the dependent variable. We present results for forecasting one quarter ahead and one year ahead using the direct method of forecasting. Papers such as Stock and Watson (2007) emphasize the importance of correctly modelling time variation in the error variance and, accordingly, most of our models include stochastic volatility (although we also include some models without stochastic volatility for comparison). In the next section, we provide a list of the predictors we use.

The following is a list of the forecasting models used in this paper, along with their acronyms.

- TVD 1,2,3: the three versions of the TVD model.

- OLS: a constant coefficient model estimated via OLS with an intercept, three lags and all the predictors.
- OLS-AR: a constant coefficient model estimated via OLS with an intercept and two lags.
- OLS-AIC: a constant coefficient model estimated via OLS with an intercept and at most 4 lags. The lag length is selected by AIC recursively.
- OLS-F: a constant coefficient model estimated via OLS with an intercept, two lags and two factors constructed from the predictors using principal components.
- OLSroll, OLSroll-AR, OLSroll-AIC and OLSroll-F: rolling window (of size 40) versions of OLS, OLS-AR, OLS-AIC and OLS-F respectively.
- TVP: time-varying parameter model estimated via MCMC with an intercept, three lags and the predictors.
- TVP-AR: time-varying parameter model estimated via MCMC with an intercept and two lags.
- TVPSV and TVPSV-AR: same as TVP and TVP-AR but with stochastic volatility.
- UCSV: unobserved-components stochastic volatility model of Stock and Watson (2007) implemented as the TVPSV model with only an intercept.
- TVPXi: time-varying parameter model estimated via MCMC with an intercept, two lags and the ith regressor ($i = 1, .., 14$, ordered in the same manner as in the list in Section 3.2).
- TVPX1-X14: equally weighted average forecasts of TVPX1–TVPX14.
- TVPSVXi: same as TVPXi but with stochastic volatility.
- TVPSVX1-X14: equally weighted average forecasts of TVPSVX1–TVPSVX14.
- PPT-AR: the structural break model of Pesaran, Pettenuzzo and Timmerman (2006) on an AR(2) model.

Note that PPT-AR allows for structural breaks in both the AR(2) coefficients and the error variance. This model requires the selection of the number of breaks. In the context of a recursive forecasting exercise, we do this as in Bauwens, Koop, Korobilis and Rombouts (2011). See the online appendix for details.

When forecasting h periods ahead, our models provide us with $p\left(y_{\tau+h}|Data_{\tau}\right)$, the predictive density for $y_{\tau+h}$ using data available through time τ. The predictive density is evaluated for $\tau = \tau_0, \ldots, T-1$ where τ_0 is 1980Q1. Let $y^o_{\tau+h}$ be the observed value of $y_{\tau+h}$ as known in period $\tau + h$. Root mean squared forecast error and mean absolute forecast error are common measures of forecast performance. These are defined as:

$$RMSFE = \sqrt{\frac{\sum_{\tau=\tau_0}^{T-h}\left[y^o_{\tau+h} - E\left(y_{\tau+h}|Data_{\tau}\right)\right]^2}{T-\tau_0-h+1}}$$

and

$$MAFE = \frac{\sum_{\tau=\tau_0}^{T-h}\left|y^o_{\tau+h} - E\left(y_{\tau+h}|Data_{\tau}\right)\right|}{T-\tau_0-h+1}$$

RMSFE and MAFE only use the point forecasts and ignore the rest of the predictive distribution. For this reason, we also use the predictive likelihood to evaluate forecast performance. Note that a great advantage of predictive likelihoods is that they evaluate the forecasting performance of the entire predictive density. Predictive likelihoods are motivated and described in many places such as Geweke and Amisano (2011). The predictive likelihood is the predictive density for $y_{\tau+h}$ evaluated at the actual outcome $y^o_{\tau+h}$. We use the sum of log predictive likelihoods for forecast evaluation:

$$\sum_{\tau=\tau_0}^{T-h}\log\left[p\left(y_{\tau+h} = y^o_{\tau+h}|Data_{\tau}\right)\right].$$

Note that, if $\tau_0 = 0$ then this would be equivalent to the log of the marginal likelihood. Hence, the sum of log predictive likelihoods can also be interpreted as a measure similar to the log of the marginal likelihood, but made more robust by ignoring the initial $\tau_0 - 1$ observations in the sample (where prior sensitivity is most acute).

In our forecasting exercise, we present results from the individual models in the preceding list. However, we also do Bayesian model averaging (BMA) using products of predictive likelihoods. To be precise, we use TVD-BMA which is calculated using model averaging over the three versions of the

TVD model. These BMA weights vary over time using a window of ten years. That is, when forecasting y_{t+h} using information through time τ, we use weights proportional to $\prod_{t=\tau-h-40}^{\tau-h} p\left(y_{t+h}=y_{t+h}^{o}|Data_{t}\right)$ for each model.

We also present results of various standard tests of forecast performance. The null hypothesis of these tests is that a benchmark forecasting model (in our case, always TVD-BMA) predicts equally as well as a comparator (in our case, one of the models in the list above). These tests are based on point forecasts. Complete details are provided in the online appendix. Here we note that we use the three test statistics that are labelled S_1, S_2 and S_3 in Diebold and Mariano (1995). We also carry out the test of Giacomini and White (2006), which we label S_4.

All OLS methods are implemented in the standard non-Bayesian manner and require no prior (and no predictive likelihoods are obtained). In order to make sure all our approaches are as comparable as possible, our TVP regression models are exactly the same as our TVD models (including having the same prior for all common parameters) except that we set $K_{jt}=1$ for all t and for the j included in the relevant TVP model. For the TVP models with stochastic volatility we use the same stochastic volatility specification and prior as with the TVD models. For the homoskedastic version, the error variance has the same prior as that used for the initial volatility in the stochastic volatility model.

The precise details of our prior are given in the online appendix. Here we offer some general comments about prior elicitation in TVD models. Since these are state space models, we require priors on the initial conditions for the states as well as the initial conditions $K_{1,1},\ldots,K_{p,1}$ and the parameters in the state equations (i.e. state equation error variances and parameters in the hierarchical prior for K). In our experimentation with different priors (summarized in the prior sensitivity analysis in the online appendix), we find our forecasting results to be robust to the choice of prior. The results reported in this paper are for a subjectively elicited, but relatively noninformative prior. For instance, $K_{j,1}$ (for $j=1,\ldots,p$) is chosen to have a Bernoulli prior with $\Pr(K_{j,1}=1)=b_j$. We then use a Beta prior for b_j with hyperparameters chosen to imply $E\left(b_j\right)=0.5$ and a large prior variance. Thus, we are centring the prior over the noninformative choice that $K_{j,1}$ is equally likely to be zero or one, but attach a large prior variance to that choice. We also use $\theta_1 \sim N(0, 5\times I)$, thus shrinking the initial coefficients towards zero, but only slightly (since the prior variance is large). For the stochastic volatility part of the model, we make the same prior choices as in

Kim, Shephard and Chib (1998). In the TVP-VAR literature it is common to use training sample priors (e.g. Cogley and Sargent, 2005 and Primiceri, 2005). As discussed in the online appendix, we have also used a training sample prior and find results to be virtually the same as for the relatively non-informative prior used in the body of the paper.

3.2 Data

In this paper we use real time quarterly data so that all our forecasts are made using versions of the variables available at the time the forecast is made. We provide results for core inflation as measured by the Personal Consumption Expenditure (PCE) deflator for 1962Q1 through 2008Q3. If P_t is the PCE deflator, then we measure inflation as $100 \times \log(P_{t+h}/P_t)$ when forecasting h periods ahead.

As predictors, authors such as Stock and Watson (1999) consider measures of real activity including the unemployment rate. Various other predictors (e.g. cost variables, the growth of the money supply, the slope of term structure, etc.) are suggested by economic theory. Finally, authors such as Ang, Bekaert and Wei (2007) have found surveys of inflation expectations to be useful predictors. These considerations suggest the following list of potential predictors which we use in this paper.

- UNEMP: unemployment rate.
- CONS: the percentage change in real personal consumption expenditures.
- GDP: the percentage change in real GDP.
- HSTARTS: the log of housing starts (total new privately owned housing units).
- EMPLOY: the percentage change in employment (All Employees: Total Private Industries, seasonally adjusted).
- PMI: the change in the Institute of Supply Management (Manufacturing): Purchasing Manager's Composite Index.
- TBILL: three month Treasury bill (secondary market) rate.
- SPREAD: the spread between the 10 year and 3 month Treasury bill rates.

- DJIA: the percentage change in the Dow Jones Industrial Average.
- MONEY: the percentage change in the money supply (M1).
- INFEXP: University of Michigan survey of inflation expectations.
- COMPRICE: the change in the commodities price index (NAPM commodities price index).
- VENDOR: the change in the NAPM vendor deliveries index.
- y_{t-3}: the third lag of the dependent variable.

This set of variables is a wide one reflecting the major theoretical explanations of inflation as well as variables which have found to be useful in forecasting inflation in other studies. The third lag of the dependent variable is included so that the model can, if warranted, choose a longer lag length than the benchmark two lags that are always included. Most of the variables were obtained from the "Real-Time Data Set for Macroeconomists" database of the Philadelphia Federal Reserve Bank. The exceptions to this are PMI, TBILL, SPREAD, DJIA, COMPRICE, INFEXP and VENDOR which were obtained from the FRED database of the Federal Reserve Bank of St. Louis.

3.3 Results

Tables 1 and 2 present the results of our forecasting exercise for one quarter and one year ahead forecasts, respectively. Predictive likelihoods, MAFEs and RMSFEs are telling a very similar story and it is one which says that the TVD models forecast very well. The main methods that occasionally forecast better are parsimonious TVP regression models which include only one regressor. For instance one year ahead, a TVP regression model using two AR lags and housing starts as regressors forecasts slightly better than the TVD models. However, a priori, a researcher in this field would not know which regressor to include (e.g. housing starts might not come to mind as being the logical regressor to include and the more logical choice of the unemployment rate does not yield a good forecast performance) and it might have been difficult to discover the fact that this was a good forecasting model using traditional model selection procedures. An alternative to the use of TVD models would be to do sequential hypothesis testing procedures to try and select which regressors to include in a forecasting model. However, even in a constant coefficient model, pre-testing problems would make this a risky strategy. In TVP regression models, such problems would worsen. Furthermore, the TVD model allows for a regressor to be included at some points in time, but excluded at others which is not possible with a conventional testing strategy. In sum, TVD models always are among the top forecasting models in Tables 1 and 2. Even in the cases where they are not the very best, it is hard to imagine a simple strategy that the researcher could use to reliably find the best forecasting model among the choices we consider. The best alternative appears to be simple averaging of parsimonious TVP models. In the remainder of this section we expand on these points.

TVD methods consistently forecast better than any of the OLS methods we consider. At the quarterly forecast horizon, forecast gains are small but at the annual horizon they are much larger. This holds true for simple AR forecasts, OLS methods using many predictors and factor methods. It also holds true regardless of whether we use rolling or expanding windows of data to produce the OLS estimates. In general, we are finding evidence that constant coefficient models (even if estimated using rolling windows) do not forecast as well as TVD models which explicitly allow for parameter change and change in model dimension over time.

The tables also show that non-parsimonious TVP models forecast very poorly as well. TVP regression models which include all the 14 predictors forecast poorly in our application. In theory, one might expect such a TVP model to be able to approximate a TVD model (i.e. the coefficients in the TVP model could evolve to be close to zero for a particular predictor and, thus, it could drop out of the model ensuring a dimension reduction). In practice, this is not happening and TVD models are forecasting better than TVP models.

TVD is also forecasting better than the popular structural break model of Pesaran, Pettenuzzo and Timmerman (2006). The tables only present results for an AR version of this structural break model. Including all the predictors leads to much worse forecast performance.

Our three TVD models exhibit similar forecast performance. Forecast metrics based on point forecasts indicate TVD1 is the best, whereas predictive likelihoods indicate TVD2. However, overall there is some evidence that use of BMA is beneficial in improving forecast performance since TVD-BMA exhibits strong forecast performance by both metrics.

Another class of popular forecasting models are parsimonious TVP models such as TVP-AR models. For instance, the popular UCSV model of Stock and Watson (2007) is a TVP regression model with only a time-varying intercept (and stochastic volatility). The UCSV and TVPSV-AR model does forecast quite well, although overall TVD-BMA forecasts slightly better (see, in particular, the predictive likelihoods for one-quarter ahead forecasts).

Tables 1 and 2 also indicate the importance of allowing for stochastic volatility. This is not so clear in terms of point forecasts, where homoskedastic and heteroskedastic versions of a model tend to have similar MAFEs and RMSFEs. However, predictive likelihoods in many cases, increase substantially when stochastic volatility is added to a model.

Tables 1 and 2 also present results for the four hypothesis tests of equal predictive performance described above (see also the online appendix). Remember that these are implemented so that each model is compared to the TVD-BMA model. Critical values for the test statistics S_1, S_2 and S_3 can be obtained from the standard normal distribution with positive values for test statistics indicate that TVD-BMA is forecasting better than the comparator model. Critical values for S_4 are obtained from the $\chi^2(3)$ distribution.

Results from these tests are largely supportive of our previous conclusions. That is, the value of these test statistics almost always indicates TVD-BMA is forecasting better and it is often the case that this forecast improvement is statistically significant. For instance, the hypothesis of equal predictability between TVP models containing all the regressors and TVD-BMA is always rejected. In most cases, the same conclusion holds for the OLS methods. Tests of equal predictability between TVD-BMA and parsimonious TVP models yield weaker results. Often it is the case that TVD-BMA forecasts better than a particular parsimonious TVP model at the 5% level of significance, but it is more common for the test statistic to be positive but insignificant at the 5% level. Although it is worth noting that there are many cases where TVD-BMA would forecast significantly better if we used a 10% level of significance.

Tables 1 and 2 establish that, overall, the TVD approaches do tend to forecast better than many commonly used benchmarks. However, they relate to average forecast performance from 1980Q1 through the end of the sample. In order to see whether there are any interesting fluctuations in forecast performance over time, Figures 1 through 4 present rolling sums of log predictive likelihoods and square roots of rolling averages of forecast errors squared for one-quarter ahead and one-year ahead forecasts. To keep the figures readable, these figures omit many models such as OLS results and results for all the TVP models containing a single regressor. Note that we use a rolling window of 40 observations which means that the figures start in 1990 (since our forecasts start in 1980).

The most striking thing about all these figures is the deterioration in forecast performance around the time of the financial crisis. Unsurprisingly,

Table 1: Measures of one quarter ahead forecast performance

	Forecast Performance			Test Statistics			
Model	RMSFE	MAFE	sum of log pre-like	S_1	S_2	S_3	S_4
TVD-BMA	0.428	0.311	-54.68	–	–	–	–
TVD1	0.424	0.305	-58.23	-0.850	0.928	0.055	7.217
TVD2	0.430	0.313	-57.91	0.552	0.186	0.926	0.850
TVD3	0.422	0.308	-61.70	-1.269	-0.186	-0.595	23.860
OLS	0.439	0.321	–	0.470	0.371	0.691	–
OLS-AR	0.448	0.321	–	2.052	1.300	1.463	–
OLS-AIC	0.430	0.309	–	0.151	0.371	0.543	–
OLS-F	0.455	0.333	–	2.694	1.857	2.397	–
OLSroll	0.486	0.365	–	2.627	3.343	2.953	–
OLSroll-AR	0.442	0.332	–	1.048	1.300	2.198	–
OLSroll-AIC	0.443	0.328	–	1.066	1.486	1.501	–
OLSroll-F	0.455	0.343	–	1.771	2.600	3.055	–
PPT-AR	0.443	0.315	-61.47	1.447	0.928	2.096	4.582
UCSV	0.438	0.315	-83.72	1.102	-0.186	0.970	1.401
TVP	0.475	0.353	-107.32	2.131	2.043	2.066	6.119
TVPSV	0.476	0.354	-87.12	2.175	2.043	2.096	7.546
TVP-AR	0.430	0.316	-76.11	0.475	1.857	1.672	3.977
TVPSV-AR	0.428	0.313	-55.67	-0.120	0.928	1.047	1.657
TVPX1	0.438	0.318	-73.20	1.290	1.300	1.306	2.425
TVPX2	0.435	0.322	-78.31	1.160	2.043	2.105	6.171
TVPX3	0.433	0.318	-77.48	0.873	2.043	1.879	2.431
TVPX4	0.428	0.314	-77.63	-0.056	-0.186	-0.025	1.331
TVPX5	0.436	0.321	-76.87	1.281	2.414	2.716	5.637
TVPX6	0.437	0.323	-80.75	1.028	3.157	2.072	2.855
TVPX7	0.448	0.330	-83.24	1.395	1.300	1.793	4.796
TVPX8	0.433	0.314	-74.29	0.646	0.928	1.006	1.063
TVPX9	0.473	0.340	-87.11	2.074	1.486	1.196	5.331
TVPX10	0.442	0.324	-80.46	1.991	0.371	1.642	4.732
TVPX11	0.441	0.321	-69.65	1.572	2.785	2.179	3.000
TVPX12	0.461	0.331	-85.63	2.225	1.671	1.656	5.161
TVPX13	0.436	0.321	-80.94	0.797	1.486	1.617	2.659
TVPX14	0.418	0.302	-71.94	-1.265	1.114	0.083	2.405
TVPSVX1	0.434	0.315	-56.85	0.742	0.371	0.780	0.575
TVPSVX2	0.430	0.318	-55.79	0.290	1.486	1.653	5.431
TVPSVX3	0.431	0.314	-56.24	0.391	2.228	1.705	0.576
TVPSVX4	0.425	0.311	-54.14	-0.499	0.000	-0.129	1.433
TVPSVX5	0.432	0.317	-57.00	0.736	2.600	2.311	4.016
TVPSVX6	0.432	0.317	-58.71	0.424	2.043	1.744	1.094
TVPSVX7	0.446	0.329	-55.67	1.226	1.486	1.804	8.222
TVPSVX8	0.432	0.312	-56.48	0.478	0.928	0.887	0.393
TVPSVX9	0.467	0.337	-60.10	1.916	2.043	1.507	3.996
TVPSVX10	0.438	0.322	-58.53	1.458	1.857	1.760	3.554
TVPSVX11	0.434	0.316	-55.45	0.934	1.114	1.639	1.283
TVPSVX12	0.455	0.324	-65.22	1.941	1.486	1.380	4.692
TVPSVX13	0.433	0.318	-59.09	0.553	1.486	1.141	0.911
TVPSVX14	0.415	0.297	-51.78	-1.595	0.000	-1.008	2.814
TVPX1-14	0.432	0.317	-69.32	0.796	2.228	1.981	3.553
TVPSVX1-14	0.429	0.314	-53.87	0.105	1.671	1.452	4.082

Table 2: Measures of one year ahead forecast performance

	Forecast Performance			Test Statistics			
Model	RMSFE	MAFE	sum of log pre-like	S_1	S_2	S_3	S_4
TVD-BMA	0.469	0.365	-77.84	–	–	–	–
TVD1	0.471	0.360	-79.23	-0.850	-0.557	-0.482	1.608
TVD2	0.487	0.382	-76.64	0.552	1.671	1.455	5.942
TVD3	0.498	0.393	-86.65	-1.269	1.857	2.328	13.924
OLS	1.319	1.046	–	0.470	6.871	7.780	–
OLS-AR	1.193	0.958	–	2.052	6.128	7.460	–
OLS-AIC	1.193	0.957	–	0.151	6.128	7.460	–
OLS-F	1.215	0.969	–	2.694	5.942	7.394	–
OLSroll	1.435	1.141	–	2.627	7.242	8.155	–
OLSroll-AR	1.226	0.987	–	1.048	6.871	7.576	–
OLSroll-AIC	1.289	1.034	–	1.066	6.871	7.675	–
OLSroll-F	1.148	0.932	–	1.771	6.685	7.394	–
PPT-AR	1.199	0.952	-191.20	1.447	6.685	3.758	73.771
UCSV	0.505	0.366	-84.54	1.102	0.928	0.510	3.413
TVP	0.603	0.471	-108.22	2.131	3.157	3.851	16.242
TVPSV	0.601	0.468	-108.44	2.175	2.971	3.758	15.734
TVP-AR	0.480	0.369	-74.69	0.475	0.371	0.455	3.771
TVPSV-AR	0.477	0.367	-74.18	-0.120	-0.186	-0.017	3.237
TVPX1	0.488	0.374	-79.27	1.290	0.557	1.066	2.909
TVPX2	0.477	0.372	-75.64	1.160	-0.743	-0.551	1.345
TVPX3	0.482	0.375	-76.45	0.873	1.114	0.915	2.188
TVPX4	0.466	0.363	-70.33	-0.056	-1.486	-1.088	4.477
TVPX5	0.469	0.358	-74.89	1.281	-1.486	-1.127	2.359
TVPX6	0.496	0.386	-80.42	1.028	0.000	1.044	2.579
TVPX7	0.493	0.377	-77.59	1.395	-0.186	-0.259	2.065
TVPX8	0.485	0.369	-76.24	0.646	0.000	-0.410	2.462
TVPX9	0.499	0.384	-81.41	2.074	1.486	1.339	5.401
TVPX10	0.479	0.370	-74.86	1.991	-0.186	-0.705	1.687
TVPX11	0.488	0.377	-76.43	1.572	0.371	0.270	2.741
TVPX12	0.505	0.388	-82.30	2.225	0.928	0.763	2.860
TVPX13	0.490	0.378	-77.91	0.797	0.557	0.981	4.835
TVPX14	0.475	0.368	-74.51	-1.265	0.186	-0.047	1.715
TVPSVX1	0.485	0.372	-78.52	0.742	0.743	0.744	2.631
TVPSVX2	0.478	0.371	-74.98	0.290	-0.557	-0.375	2.026
TVPSVX3	0.479	0.372	-75.12	0.391	0.928	0.708	2.437
TVPSVX4	0.462	0.361	-69.10	-0.499	-2.043	-1.523	5.591
TVPSVX5	0.470	0.359	-74.57	0.736	-1.114	-0.937	2.819
TVPSVX6	0.499	0.388	-79.44	0.424	0.557	1.069	3.661
TVPSVX7	0.482	0.372	-76.06	1.226	-1.114	-0.204	1.993
TVPSVX8	0.483	0.368	-75.91	0.478	-0.557	-0.534	2.194
TVPSVX9	0.498	0.383	-80.63	1.916	1.114	1.320	4.417
TVPSVX10	0.480	0.369	-73.99	1.458	-0.371	-1.011	2.674
TVPSVX11	0.481	0.372	-75.17	0.934	0.000	-0.008	2.339
TVPSVX12	0.504	0.388	-81.35	1.941	0.371	0.639	2.553
TVPSVX13	0.485	0.374	-76.92	0.553	0.557	0.849	3.183
TVPSVX14	0.477	0.368	-74.11	-1.595	-0.186	-0.292	2.639
TVPX1-14	0.476	0.368	-75.20	0.796	0.186	-0.215	1.797
TVPSVX1-14	0.474	0.366	-74.22	0.105	0.000	-0.207	2.426

Figure 1: Root mean squared forecast errors for quarterly forecasts with a rolling window of size 40.

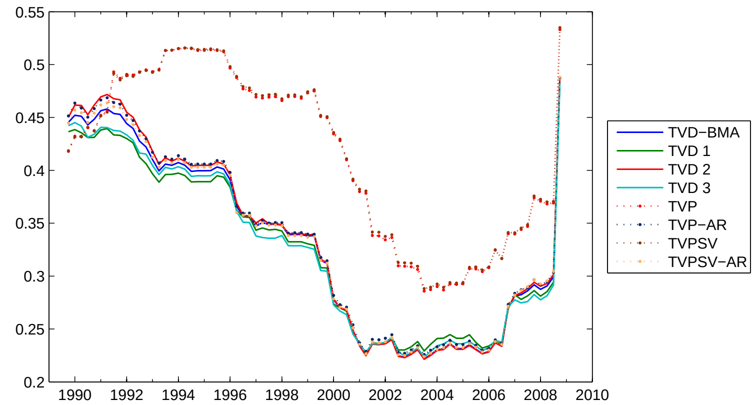


Figure 2: Root mean squared forecast errors for annual forecasts with a rolling window of size 40.

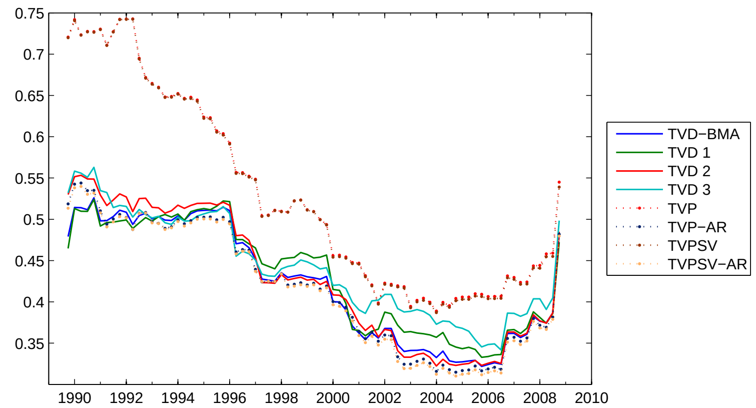


Figure 3: Average log predictive likelihoods for quarterly forecasts with a rolling window of size 40.

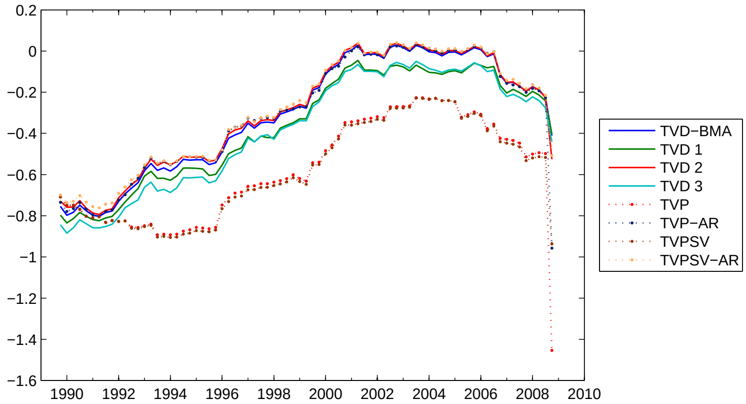
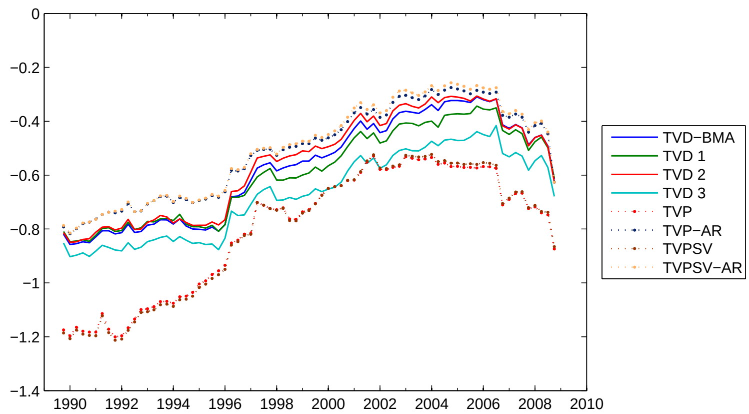


Figure 4: Average log predictive likelihoods for annual forecasts with a rolling window of size 40.



this occurs with every forecasting method. However, this deterioration is much less for the TVD methods than for some of the other methods. For instance, in Figure 3, the TVD-BMA line does drop with the financial crisis, however it drops much more for the TVP models with all the predictors and the more parsimonious TVP-AR models.

At the quarterly forecast horizon, the forecasting superiority of TVD improvements in forecast performance only appears after the early 1990s. In fact, there is a period in the 1980s and early 1990s that the over-parameterized TVP models (which include all the regressors) forecast better than the other models. However, later in the sample there is a clear deterioration in forecast performance of TVP and TVPSV. At the annual forecast horizon, this pattern is not found. The TVP regression models forecast poorly from the very beginning of our forecast period.

4 Conclusions

In this paper, we have presented a battery of theoretical and empirical arguments for the potential benefits of TVD models. Like TVP models, TVD models allow for the values of the parameters to change over time. Unlike TVP models, they also allow for the dimension of the parameter vector to change over time. Given the potential benefits of a TVD framework, the task is to build specific TVD models. This task was taken up in Section 2 of this paper where three different TVD models were developed. These models all are dynamic mixture models and, thus, have the enormous benefit that we can draw on existing methods of posterior computation developed in Gerlach, Carter and Kohn (2000).

An empirical illustration involving forecasting US inflation illustrated the feasibility and desirability of the TVD approach.

References

Amato, J. and Swanson, N. (2001), "The Real-time Predictive Content of Money for Output," *Journal of Monetary Economics*, 48, 3-24.

Ang, A. Bekaert, G. and Wei, M. (2007), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?," *Journal of Monetary Economics* 54, 1163-1212.

Ballabriga, F., Sebastian, M. and Valles, J. (1999), "European Asymmetries," *Journal of International Economics*, 48, 233-253.

Banbura, M., Giannone, D. and Reichlin, L. (2010), "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics*, 25, 71-92.

Bauwens, L., Koop, G., Korobilis, D. and Rombouts, J. (2011). "The Contribution of Structural Break Models to Forecasting Macroeconomic Time Series," Rimini Centre for Economic Analysis, Working Paper 11-38.

Canova, F. (1993), "Modelling and Forecasting Exchange Rates Using a Bayesian Time Varying Coefficient Model," *Journal of Economic Dynamics and Control*, 17, 233-262.

Canova, F. (2007), *Methods for Applied Macroeconomic Research*, Princeton: Princeton University Press.

Canova, F. and Ciccarelli, M. (2004), "Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model," *Journal of Econometrics*, 120, 327-359.

Carter, C. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Chan, J.C.C. and Jeliazkov, I. (2009), "Efficient Simulation and Integrated Likelihood Estimation in State Space Models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101-120.

Chib, S. and Greenberg, E. (1995), "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-varying Parameter Models," *Journal of Econometrics*, 68, 339-360.

Ciccarelli, M. and Rebucci, A. (2002), "The Transmission Mechanism of European Monetary Policy: Is There Heterogeneity? Is it Changing Over Time?," International Monetary Fund working paper, WP 02/54.

Cogley, T. and Sargent, T. (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the post WWII U.S.," *Review of Economic Dynamics*, 8, 262-302.

D'Agostino, A., Gambetti, L. and Giannone, D. (2009), "Macroeconomic Forecasting and Structural Change," ECARES working paper 2009-020.

De Mol, C., Giannone, D. and Reichlin, L. (2008), "Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components?" *Journal of Econometrics*, 146, 318-328.

Diebold, F. and Mariano, R. (1995), "Comparing Predictive Accuracy," *Journal of Business Economics and Statistics*, 13, 134-144.

Durbin, J. and Koopman, S. (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603-616.

George, E., Sun, D. and Ni, S. (2008), "Bayesian Stochastic Search for VAR Model Restrictions," *Journal of Econometrics*, 142, 553-580.

Gerlach, R., Carter, C. and Kohn, R. (2000), "Efficient Bayesian Inference in Dynamic Mixture Models," *Journal of the American Statistical Association*, 95, 819-828.

Geweke, J. and Amisano, G. (2011), "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," *Journal of Applied Econometrics,* 26, 1-29.

Giacomini, A., and White, H. (2006), "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578

Giordani, P. and Kohn, R. (2008), "Efficient Bayesian Inference for Multiple Change-point and Mixture Innovation Models," *Journal of Business and Economic Statistics*, 12, 66-77.

Giordani, P., Kohn, R. and van Dijk, D. (2007), "A Unified Approach to Nonlinearity, Structural Change and Outliers," *Journal of Econometrics*, 137, 112-133.

Groen, J., Paap, R. and Ravazzolo, F. (2010), "Real-time Inflation Forecasting in a Changing World," *Federal Reserve Bank of New York Staff Report Number 388*.

Kim, S., Shephard, N. and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-93.

Koop, G. (2003), *Bayesian Econometrics*, Chichester: Wiley.

Koop, G. and Korobilis, D. (2011), "Forecasting Inflation using Dynamic Model Averaging," *International Economic Review*, forthcoming.

Koop, G., León-González, R. and Strachan R.W. (2009), "On the Evolution of the Monetary Policy Transmission Mechanism," *Journal of Economic Dynamics and Control*, 33, 997–1017.

Koop, G., León-González, R. and Strachan R.W. (2010), "Dynamic Probabilities of Restrictions in State Space Models: An Application to the Phillips Curve," *Journal of Business and Economic Statistics*, 28, 370-379.

Koop, G. and Potter, S. (2011), "Time Varying VARs with Inequality Restrictions," *Journal of Economic Dynamics and Control*, forthcoming.

Korobilis, D. (2011), "VAR Forecasting using Bayesian Variable Selection," *Journal of Applied Econometrics*, forthcoming.

Pesaran, M.H., Pettenuzzo, D. and Timmerman, A. (2006), "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057-1084.

Primiceri. G. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821-852.

Raftery, A., Karny, M., Andrysek, J. and Ettler, P. (2010), "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52, 52-66.

Staiger, D., Stock, J. and Watson, M. (1997), "The NAIRU, Unemployment and Monetary Policy," *Journal of Economic Perspectives,* 11, 33-49.

Stock, J. and Watson, M. (2007), "Why has US Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 39, 3-33.

Stock, J. and Watson, M. (2009), "Phillips Curve Inflation Forecasts," chapter 3, pages 99-202 in Jeffrey Fuhrer, Yolanda Kodrzycki, Jane Little, and Giovanni Olivei (eds.), Understanding Inflation and the Implications for Monetary Policy. Cambridge: MIT Press.

Stock, J. and Watson, M. (2010), "Modeling Inflation After the Crisis," National Bureau of Economic Research Working Paper #16488.