STATISTICS IN TRANSITION new series, June 2018
Vol. 19, No. 2, pp. 239–258, DOI 10.21307/stattrans-2018-014

MODIFIED RECURSIVE BAYESIAN ALGORITHM FOR ESTIMATING TIME-VARYING PARAMETERS IN DYNAMIC LINEAR MODELS

O. Olawale Awe[1], **A. Adedayo Adepoju**[2]

ABSTRACT

Estimation in Dynamic Linear Models (DLMs) with Fixed Parameters (FPs) has been faced with considerable limitations due to its inability to capture the dynamics of most time-varying phenomena in econometric studies. An attempt to address this limitation resulted in the use of Recursive Bayesian Algorithms (RBAs) which is also affected by increased computational problems in estimating the Evolution Variance (EV) of the time-varying parameters. In this paper, we propose a modified RBA for estimating TVPs in DLMs with reduced computational challenges.

Key words: discounted variance, dynamic models, granularity range, estimation algorithm.

1. Introduction

Generally speaking, a model is dynamic each time the variables (or parameters) are indexed by time or appear with different time lags (Ravines et al., 2006). In recent times, estimation of time-varying parameters in econometric models has become more relevant especially as the length of the observed time series increases and the series itself is subject to changes in the dynamic structure. Particular examples can be found in world economic time series where key monthly, quarterly or annual economic indicators are commonly available from the 1950s and cover periods of different economic conditions. For example, the periods of strong economic growth in the 1950s and 1960s, periods with oil crises in the 1970s, periods of major monetary policy changes in the 1980s, rapid changes of financial markets in the 1990s and the collapse of the financial and banking systems more recently (Doh and Connolly, 2013). Although, not all economic structures are subject to changes due to these developments, it is expected that the dynamic properties of longer time series require parameters that are allowed to change over time. Models with fixed parameters have been found to perform poorly for analysis of these kinds of data because basic econometric time series analysis lies in the possibility of finding a reasonable regularity in the phenomenon under study (Petris, 2010). In a dynamic economy, for instance, the relations between economic agents are subject to change. As the

[1]Department of Mathematical Sciences, Anchor University Lagos, Nigeria. E-mail: oawe@aul.edu.ng
[2]Department of Statistics, University of Ibadan, Nigeria. E-mail: pojuday@yahoo.com

knowledge of production techniques improved, as the means of transportation allow for long-distance trade, and as the society changed its preferences for certain goods or services, the structure of the economy varies accordingly. It is also natural to think that as time changes, information decays thereby necessitating the need for discounting the variance of the underlined evolution of the dynamic parameter as adopted in this work.

In a surprisingly short period of time, Markov chain Monte Carlo (MCMC) algorithms, especially the Metropolis-Hastings algorithm (Hastings, 1970) and the Gibbs sampling algorithm (Geman and Geman, 1984; Gelfand et al., 1990) have emerged as extremely popular tools for the analysis of complex statistical and econometric models. Bayesian analysis requires the evaluation of complex and often high-dimensional integrals in order to obtain posterior distributions for the unobserved quantities of interest in the model. In many such settings, alternative methodologies such as asymptotic approximation and non-recursive Monte Carlo algorithms are either infeasible or fail to provide sufficiently accurate results. Properly defined and implemented, MCMC methods enable the user to successively sample values from a Markov chain process. Important features of MCMC methods that enhance their applicability include their ability to reduce complex multidimensional problems to a sequence of much lower-dimensional ones. While MCMC algorithms allow an enormous expansion of the class of candidate models for a given data, they also suffer from a well-known and potentially serious problem: it is often difficult to decide when it is safe to terminate them and conclude their "convergence" (Zellner, 2009).

The algorithms for recursive estimation and Kalman filtering are being used increasingly in applied econometrics, but econometricians have been slower than other statisticians to explore them (Pollock, 2003). In recursive estimation, the knowledge about the parameters of a model is updated continuously as new measurements are collected. It is suitable in problems where the parameters have dynamic properties that make them change with time. Several measurements $y_1, y_2, ..., y_n$ are considered alongside their joint probability density function $f(\theta, y_1, y_2, y_3, ..., y_n)$. New measurements are received and estimation done one at a time. After measuring y_1, we construct an estimate $\hat{\theta}_i$, when y_{i+1} is received again, parameter $\hat{\theta}_i$ would be updated. This process continues recursively. To initiate the recursion, we need an initial estimate of the parameter θ and its variance-covariance matrix.

Another reason for the burgeoning popularity of the recursive approach in econometrics is the increased importance of numerical simulations in statistics and econometrics, hence, most computational algorithms rely on recursive methods . A significant breakthrough in the application of recursive methods in econometrics was achieved by several researchers including Cooley and Prescott (1973, 1976); Bertsekas (1976); Spear and Srivastava (1987); Blanchard and Fischer (1989); Abreu et al. (1990), Ng and Young (1990); Pollock (2003); Young (2011). Although, economic theory rarely provides a useful guide to distinguish between fixed and time-varying parameters, estimation of Dynamic Linear Models (DLMs) with Fixed Parameters (FPs) has been faced with considerable limitations due to its inability to capture the dynamics of most time-varying phenomena in econometric studies. This is because the classical linear regression model with fixed parameters presumes that the relationship between the explanatory and the explained variable remains constant through the estimation period. However, there are situations when this kind of assumption becomes unreasonable and totally non-implementable because assuming time-invariant parameters and variances turn out to be quite restrictive in capturing the evolution dynamics of most economic time series. For instance, business cycle dynamics and monetary policy in United States and other major economies of the world has changed substantially over the post-war period. In addition, the introduction of time-varying parameters in linear models can lead to different levels of complexities including the fact that they gain new evolution variance parameters which are also time-varying and, in turn, need to be estimated. In literature, the choice of the evolution variance (say Ω_t) has been found to be complex and usually difficult to characterize because of a number of practical problems associated with it which includes:

1. it varies with the measurement scale of regressor variables as specified in the observation equation of DLM.
2. it can be ambiguous i.e there may not be an optimal value of Ω_t suitable for all times.
3. it is often grossly mis-specified because most modellers have great difficulty in directly quantifying its variance and covariance elements.
4. the predictive performance of the dynamic linear model depends on the choice of the evolution variance Ω_t (West and Harrison, 1997).

An attempt to address these problems resulted in the use of Recursive Bayesian Algorithms (RBAs) which is also affected by increased computational problems in estimating the Evolution Variance (Ω_t) of the Time-Varying Parameters (TVPs). Consequently, researchers require a better way of structuring the evolution variance which previous studies have failed to effectively address. The aim of this study, therefore, was to modify an existing RBA of Fúquene et al. (2015) for estimating TVPs in DLMs. Discounting is proposed as an alternative way of coping with the system evolution variance of economic series in order to portray a clearer picture of the volatility of the parameters in the model under study over time.

The proposed recursive Bayesian estimation algorithm will be useful for proper choice of discount values to represent the evolution variance which is inevitable in order to address problems (2) and (3) above with reduced computational challenges.

2. Dynamic Linear Models with Time-Varying Parameters

It has been argued severally in literature that the parameters in econometric models cannot, in general, be expected to remain constant and hence Time-Varying Parameter (TVP) models should be considered in almost all circumstances (Soloviev

et al., 2011; Primiceri, 2005; Doh and Connolly, 2013). The difficulty in estimating such models is however often exacerbated by the fact that the econometrician would have only some idea regarding the most likely value that a parameter may assume (as indicated by, say, the Ordinary Least Squares (OLS) and maximum likelihood estimators); with a range of uncertainty surrounding this nominal value and consequently, misleading policy prescriptions are likely to arise from a straightforward optimization exercise based on such a set of nominal values especially in the presence of structural breaks in the underlying economic, technological, behavioural and institutional patterns. However, because discontinuities are a crucial feature of modern economic systems, there is the need to consider models with time-varying parameters. According to literature, such TVP models can be classified into three types:
First, the parameters can vary across subsets of observations within the sample but be non-stochastic. Examples of such models include the general systematically varying parameter model of Belsley and Kuti (1973) and a variety of switching regression models with either known joint points (see McZgee and Carleton (1970); Hinkley (1971); Goldfeld and Quandt (1973)) or unknown joint points of Gallant and Fuller (1973). A second class of models is where the parameters are stochastic, and are assumed to be generated by a stationary stochastic process. Examples of such models include the pure random coefficient model of Harvey and Phillips (1982) which includes the adaptive and varying-parameter regression models of Cooley and Prescott (1973) and the stochastically convergent parameter model of Rothenberg (1973). Finally, the third class of models consists of those where the stochastic parameters are generated by a process that is not stationary. These include the mixed estimation model of Cooper (1972), the Kalman filter model of Athans (1974), the stochastic variations model of Cooley and Prescott (1976), the systematically varying parameter model of Kalaba and Tesfatsion (1980) which was then extended to the flexible least squares (FLS) approach Kalaba and Tesfatsion (1988), the recursive and optimal control model of Ng and Young (1990) which have gained tremendous popularity in literature and become more relevant in recent times. Some of the rationale behind time-varying parameter models are documented in Sarris (1973). The archetypical (existing) dynamic linear model in literature with fixed variances has the following general form:

$$y_t = F'\beta_t + v_t \qquad v_t \sim N(0,V) \tag{1}$$

$$\beta_t = G\beta_{t-1} + w_t \qquad w_t \sim N_p(0,\Omega) \tag{2}$$

$$\beta_0 \sim N_p(m_0, C_0) \tag{3}$$

where y_t is a vector of dimension $m \times 1$

Equation (1) is known as the observation equation while equation (2) is a first order Markov process called the evolution equation.

The matrices F, V, G and Ω are known as the system matrices and contain non-random elements. If they do not depend deterministically on t, the model is time invariant, otherwise it is time varying. The initial state distribution is assumed to be Normally distributed with parameters m_0 and C_0 as shown in (3) where $E(v_t\beta_t') = 0$,$E(w_t\beta_t') = 0$ for $t = 1,...,T$.

The dynamic linear model with state space approach offers attractive features with respect to their generality, flexibility and transparency. The lack of publicly available software to estimate these models has been the main reason why only relatively few economic and finance related problems have been analyzed with dynamic linear models so far. Basically, the estimation of DLM involves three stages: prediction, filtering and smoothing. Prediction has to do with forecasting future values of the time-varying state parameters. Filtering makes the best estimate of the current values of the time-varying state parameter from the record of observations including the current observation. Smoothing involves making the best estimate of past values of the states given the record of observations.

3. Model Specification and Methodology

A typically difficult problem in econometrics is to formulate a stationary model that best resembles the model dictated by economic theory, but which does not pose serious problems of estimation (Chetty, 1971). This section, lays out the specified dynamic linear model proposed in this work. It also contains details of the developed recursive Bayesian algorithm. Remove (RBA) employed for the posterior estimation of the specified dynamic linear model in the presence of discounted evolution variance.

A concrete mathematical formulation of the proposed dynamic linear model specification in this work takes the form of the two equations:

$$y_t = X_t\theta_t + v_t \qquad v_t \sim N(0,\varphi), \tag{4}$$

$$\theta_t = G_t\theta_{t-1} + w_t \qquad w_t \sim N(0,\Omega_t), \tag{5}$$

$$\theta_0 \sim N(m_0, C_0).$$

where equation (4) is known as the observation equation while equation (5) is the evolution equation. G_t is a known transition matrix of order $p \times p$ that determines how the observation and evolution equations evolve in time. Since each parameter at time t only depends on results from time $t-1$, the state parameters are time-varying and constitute a Markov chain. X_t is a matrix of observed time series of known order. θ_t is the time-varying parameter associated with the predictor matrix X_t. It is assumed that information decays arithmetically through the addition of future evolution error variance which we estimate with discount values. Parameters of interest to be estimated are the time-varying parameter θ_t, the error variances φ and Ω_t, and the one-step-ahead forecasts error f_t. φ is assumed to be distributed inverse-gamma a priori, while we estimate Ω_t via discounting method which is explained later in this section. The difference between this model and the one stated in Fuquene et al. (2013) is that the observational variance is presumed to be fixed and the evolution variance is estimated by the method of discounting unlike the use of Wishart prior which is common in literature. Also, in contrast to the Box-Jenkins methodology, which still plays an important role in time series analysis today, the specified dynamic linear model approach allows for structural analysis of univariate as well as multivariate problems without initial differencing or log transformation of the observed series. The different components of a series, such as trend and seasonality, as well as the effects of explanatory variables can be modelled explicitly. They do not have to be removed prior to the main analysis as is the case in the Box-Jenkins methodology.

3.1. Existing Recursive Bayesian Algorithm (RBA) and Gibbs Sampler for Estimating TVPs

The recursive Bayesian algorithms in literature usually takes the following form: Let $\Theta_t = [\theta_0, \theta_1, ..., \theta_t]$, θ_t is estimated from the conditional density $p(\Theta_t|y_T)$ which is denoted by

$$p(\Theta_t, y_T) = p(Y_T|\Theta_T)p(\Theta_T)$$

where $p(y_T|\Theta_T)$ and $p(\Theta_T)$ are given by

$$p(y_T|\Theta_T) = \prod_{t=1}^{T} p(y_t|\theta_t) \tag{6}$$

and

$$p(\Theta_T) = p(\theta_0)\prod_{t=1}^{T}(\theta_t|\theta_{t-1}), \tag{7}$$

where $p(y_t|\theta_t)$ and $p(\theta_t|\theta_{t-1})$ were derived from the observational and evolution equations (4) and (5) specified above to give

$$p(y_t|\theta_t) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\bigg(-\frac{1}{2\sigma^2}(y_t - x_t\theta_t)^2\bigg)$$

where $V = \sigma^2$,

$$p(\theta_t|\theta_{t-1}) = (2\pi)^{-\frac{k}{2}}|\Omega_t|^{-\frac{1}{2}} exp\bigg(-\frac{1}{2}(\theta_t - G_t\theta_{t-1})'\Omega_t^{-1}(\theta_t - G_t\theta_{t-1})\bigg)$$

The recursive algorithm alternatively compute the densities of the current and the future parameter θ conditional on all available observations. Using the notation $y_t = y_{1:t}$, the prediction equation is given by

$$p(\theta_{t+1}) = \int p(\theta_{t+1}, \theta_t|y_t)d\theta_t \tag{8}$$

$$= \int p(\theta_{t+1}|\theta_t)p(\theta_t|y_t)d\theta_t.$$

Applying Bayes' rule, the filtering equation gives

$$p(\theta_t|y_{1:t}) = \frac{p(y_t|\theta_t)p(\theta_t|\theta_{t-1}, y_{1:t-1})}{p(y_t|y_{1:t-1})} \tag{9}$$

$$\propto p(y_t|\theta_t)p(\theta_t|\theta_{t-1}, y_{1:t-1})$$

The denominator, $p(y_t|y_{1:t-1})$ is constant relative to θ_t and thereby ignored. These recursive equations were initialized with the density of the initial parameter

$$p(\theta_1|y_0) = p(\theta_1).$$

This posterior is then used to update the prior recursively until convergence is achieved. The forward filtering step is the standard Kalman filtering analysis to give $p(\theta_t|D_t)$ at each t, for $t = 1,...,n$. The backward sampling step uses the Markov's property to sample θ_T^* from $p(\theta_T|D_T)$ and then for $t = 1,...,T-1$, sample θ_t^* from $p(\theta_t|D_t, \theta_{t+1}^*)$ in order to generate samples from the posterior parameter structure. In particular, denote

$$p(\theta_0,...,\theta_T|y_T) = \prod_{t=0}^{T} p(\theta_t|\theta_{t+1},...,\theta_T, y_T)$$

and note that, by the Markov's property,

$$p(\theta_t|\theta_{t+1},...,\theta_t, y_t) = p(\theta_t|\theta_{t+1}, y_t) \tag{10}$$

and

$$p(\theta_t|y_t) = \int p(\theta_t, \theta_{t+1}|y_t)d\theta_{t+1}$$

$$= \int p(\theta_{t+1}|y_t)p(\theta_t|\theta_{t+1}, y_t)d\theta_{t+1}$$

$$= p(\theta_t|y_t)\int \frac{p(\theta_{t+1}|y_t)(p(\theta_{t+1}|\theta_t)d\theta_{t+1}}{p(\theta_{t+1}|y_t)} \tag{11}$$

which follows again from the recursive application of Bayes' rule and Markov property of θ_t.

Since the sampling is done from $t = T$ to $t = 0$, recursively, this procedure is referred to as recursive backward sampling.

In particular, Fuquene et al. (2013) proposed a dynamic linear model which is specified by a normal prior distribution for the p-dimensional state vector for macroeconomic modeling with prior $\theta_0)$ as follows: $\theta_0 \sim N_p(m_0, C_0)$ with the set of equations

$$y_t = F_t\theta_t + v_t, v_t \sim N_m(0, V_t) \tag{12}$$

$$\theta_t = G_t\theta_t + v_t, w_t \sim N_m(0, W_t) \tag{13}$$

with $t = 1 : T$ where F_t and G_t are known matrices of order $p \times p$ and $m \times p$, respectively. Let $\theta \sim Student - t(\mu, \tau, \nu)$ where ν is the degree of freedom, μ and τ are the location and scale parameters of the student-t density respectively. Then,

$$\pi(\theta|\tau^2) = \frac{k_1}{\tau}\left(1 + \frac{1}{\nu}\left(\frac{\theta - \mu}{\tau}\right)^2\right)^{\frac{-\nu+1}{2}} \tag{14}$$

where $\nu > 0$, $-\infty < \mu < \infty$, $-\infty < \theta < \infty$, and
$k1 = \Gamma\frac{\frac{(\nu+1)}{2}}{\Gamma\frac{(}{\nu}2)\sqrt{\nu\pi}}$ we have that $\pi(\theta) = \int_0^\infty \pi(\theta|\tau^2)\pi(\tau^2)d\tau^2$ Let $W_{t,i}$ denote the $i^t h$ diagonal element of the evolution variance $W_{t,i}$, $i = 1, ..., n$, the observation and evolution variances were given as $V_t^{-1} = \lambda_y w_y, t$
and $W_{t,i}^{-1} = \lambda\theta, iw\theta, t_i$

In order to obtain posterior inference on the time-varying parameter, θ_t, they used the recursive Forward Filtering Backward Sampling (FFBS) algorithm which proceeds as follows

1. Use the Kalman filter equations for (14) above.
   Let m_0 C_0 be known with $(\theta_0|D_0) \sim N(m_0, C_0)$ and $\theta_t|y_{1:t-1} \sim N(m_{t-1}, C_{t-1})$,
   The one-step predictive distribution of θ_t given $y_{1:t-1}$ is Gaussian i.e $\theta_t|D_{t-1} \sim N(a_t, R_t)$ with parameter $a_t = G_t m_{t-1}$, $R_t = G_t C_{t-1} G_t'$.
   The one step ahead predictive distribution of y_t given $y_{1:t-1}$ is Normally distributed as $(y_t|D_{t-1}) \sim N(f_t, Q_t)$ with parameters $f_t = F_t' a_t$, $Q_t = F_t' R_t F_t + V_t$.
   The filtering distribution of θ_t given $y_{1:t-1}$ is $(\theta_t|D_t) \sim N(m_t, C_t)$ with parameters $m_t = a_t + A_t e_t$, $C_t = R_t - A_t Q_t A_t'$ where $A_t = R_t F_t Q_t^{-1}$ and $e_t = y_t - f_t$.

2. At time $t = T$, sample θ_T from $N(\theta_T|m_t, C_T)$

3. For $t = T - 1$, sample θ_t from $N(\theta_t|m_t^*, C_t^*)$ where $m_t^* = m_t + b_t(\theta_{t+1} - a_{t+1})$ $C_t^* = C_t - b_t R_{t+1} b_t'$ where $b_t = C_t G_{t+1} R_{t+1}^{-1}$

This algorithm does not specify a block for the evolution variance of the time-varying parameter θ_t which is often difficult to characterize. Additionally, the algorithm presented in this work specifies a sub-algorithm for optimal selection of Average Granularity Range (AGR) of the discounting parameter λ which also plays an important role in determining convergence of the parameters. First, the observational variance φ, was specified as constant and estimated via Gibbs sampling as presented in the next section.

3.2. Recursive Estimation of Time-Varying Parameters in the Presence of Discounted Evolution Variance

In this section, we propose an algorithm to estimate the time-varying parameters in dynamic linear models in the presence of discounted evolution variance. This approach makes use of the Recursive Forward Filtering Backward Sampling algorithm within the Kalman filter framework to improve the efficiency of the adapted Gibbs sampler by discounting the evolution variance. The main idea of this procedure is to make use of Markov's property of the specified evolution equation so that

$$P(\theta_t|\theta_{t+1},D_t) = P(\theta_{t+1}|\theta_t,D_t) \tag{15}$$

where θ_t denotes the time-varying parameters at time t and $D_t = (y_1,...,y_t,x_1,...,x_t)$. Due to the Markovian structure of the time-varying parameter θ_t, it is estimated by computing the predictive and filtering distributions of θ_t recursively starting from the prior $\theta_0 \sim N(m_0,C_0)$. This recursive method allows us to draw the parameter vectors jointly. Consider a vector of unknown time-varying slope parameters $\theta_t = (\theta_1,...,\theta_p)$, the Gibbs sampling algorithm employed proceeds by sampling recursively the conditional posterior distribution where the most recent values of the conditioning parameters are used. Following the Bayesian paradigm, the specification of the model is complete only after specifying the prior distribution of all the unknown quantities of interest in the model. We assign a distribution to θ_t at time t=0, conditional on all the information available before any observation is made. Let D_0 be the set containing all this information, then the prior distribution is $\theta_0|D_0 \sim N(m_0,C_0)$ where m_0 and C_0 are known vector and matrix respectively. Next, an update is made for θ_1 and D_0 which is also normally distributed. Based on this update, the one step-ahead forecast follows from the conditional distribution $y|\theta_0,D_0$. Once the value of y_1 at time $t = 1$ is known, the posterior distribution of θ_1 is obtained recognizing that the information available at time $t = 1$ is $D_1 = y_1,D_0$. The inference is made in this recursive fashion for every time t. The Kalman filter was used to calculate the mean and variance of the parameter θ_t, given the observations D_t. It is a recursive algorithm because the current best estimate is updated whenever a new observation is obtained. This recursive Bayesian technique of model estimation can be stated in form of prediction, filtering and update equations. The prediction and update step requires a few basic calculations of which only the conditional means and variances of the filtering and prediction density is stored in each step of the iteration.
To describe the filtering procedure, let

$$m_t = E(\theta_t|D_t) \tag{16}$$

be the optimal estimator of θ_t based on D_t and let

$$C_t = E((\theta_t - m_t)(\theta_t - m_t)'|D_t) \tag{17}$$

be the mean square error matrix of m_t. Let $\theta_{t-1}|y_{1:t-1} \sim N(m_{t-1},C_{t-1})$, where $y_{1:t-1}$ denote all observations up to time $t-1$. Then the one-step-ahead predictive density $\theta_t|y_{1:t-1}$ is Gaussian with parameters:

$$E(\theta_t|y_{1:t-1}) = m_{t-1} \equiv A_t(say) \tag{18}$$

$$Var(\theta_t|y_{1:t-1}) = C_{t-1} + \Omega_t \equiv R_t(say) \tag{19}$$

The one-step-ahead predictive density of $y_t|y_{1:t-1}$ is Gaussian with parameters:

$$f_t = E(y_t|y_{1:t-1}) = X_t A_t \tag{20}$$

$$Q_t = Var(y_t|y_{1:t-1}) = X_t R_t X_t^{'} + V \tag{21}$$

The filtering density of θ_t given $y_{1:t}$ is Gaussian with parameters:

$$m_t = E(\theta_t|y_{1:t}) = A_t + R_t X_t^{'} Q_t^{-1} e_t \tag{22}$$

$$C_t = Var(\theta_t|y_{1:t}) = R_t - R_t X_t^{'} Q_t^{-1} X_t R_t^{\prime} \tag{23}$$

where $e_t = y_t - f_t$ is the forecast error.

3.2.1 Posterior Estimation of Unknown Observational Variance (φ) with Independent Priors

In the simulation excercise for estimating the static observational variance, φ, the following Gibbs sampler of Nakajima et al. (2011) was adopted with slight modifications: Consider the linear equation which is the observational equation specified in (4) above

$$y_t = X_t \theta_t + v_t, v_t \sim N(0, \varphi), \tag{24}$$

let $\varphi \equiv \sigma^2$ and $\theta_t = \theta$

and assume a normal prior for the parameter θ and inverse gamma prior for the parameter σ^2, to sample from $\varphi|\theta$ we impose a gamma prior on φ^{-1} and derive the posterior hyperparameters. Let $\varphi^{-1} \sim Gamma(a_0, b_0)$, then

$$\varphi^{-1}|\theta \sim Gamma(a_0 + \frac{T}{2}, b_0 + \frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\theta_t)^2)$$

We start with

$$p(y|\theta, X) = (2\pi)^{-\frac{n}{2}} exp(-\frac{1}{2\sigma^2}(y - X\theta)'(y - X\theta)) \tag{25}$$

The priors are given as follows:

$$p(\theta, \varphi) = p(\theta)p(\varphi)$$

where

$$\theta \sim N(\mu_0, \varphi_0) \tag{26}$$

and

$$\varphi \sim IG(\nu_0, \tau_0) \tag{27}$$

μ_0 is the prior mean for θ and φ_0 is the prior variance- covariance matrix for θ

with

$$E(\varphi) = \frac{\tau_0}{v_0 - 1} \tag{28}$$

$$V(\varphi) = \frac{\tau_0^2}{(v_0-1)^2(v_0-2)} \tag{29}$$

We chose the form given in Gelman (2004) where v_0 and τ_0 are the shape and scale parameters respectively. Using Bayes rule to combine the priors (26) and (27) above with the likelihood and dropping all unrelated terms to the parameters of interest yields the following posterior kernels:

$$p(\theta,\varphi|y,X) \propto (\sigma^2)^{\frac{-n-2v_0-2}{2}} exp(-\frac{1}{2\sigma^2}(2\tau_0)) \\ \times exp\left(-\frac{1}{2}(\frac{1}{\sigma^2}(y-X\theta)'(y-X\theta)+(\theta-\mu_0)'\varphi_0^{-1}(\theta-\mu_0))\right) \tag{30}$$

First, we find the posterior density of θ, conditional on φ while treating φ as a constant.
This leaves us with the posterior kernel:

$$p(\theta|\varphi,y,X) \propto \\ exp(-\frac{1}{2}(\frac{1}{\varphi}(y-X\theta)^{'}(y-X\theta)+(\theta-\mu_0)^{'}(\varphi_0)^{-1}(\theta-\mu_0))). \tag{31}$$

Transformations

Let

$$\varphi_1 = (\varphi_0^{-1} + \frac{1}{\varphi}X'X)^{-1}$$

and

$$\mu_1 = \varphi_1(\varphi_0^{-1}\mu_0 + \frac{1}{\varphi}X'Xb) = \varphi_1(\varphi_0^{-1}\mu_0 + \frac{1}{\varphi}X'y)$$

Then from (3.34),

$$\frac{1}{\varphi}(y-X\theta)'(y-X\theta)+(\theta-\mu_0)^{'}\varphi_0^{-1}(\theta-\mu_0)$$

$$
\begin{aligned}
&= \frac{1}{\varphi}y'y+\theta'\frac{1}{\varphi}X'X\theta-\frac{1}{\varphi}y'X\theta-\theta'\frac{1}{\varphi}X'y+\theta'\varphi_0^{-1}\theta \\
&- \mu_0^{'}\varphi_0^{-1}\theta-\theta^{'}\varphi_0^{-1}\mu_0+\mu_0^{'}\varphi_0^{-1}\mu_0 \\
&= \theta'(\varphi_0^{-1}+\frac{1}{\varphi}X^{'}X)\theta-\theta^{'}(\varphi_0^{-1}\mu_0+\frac{1}{\varphi}X^{'}y) \\
&- (\mu_0^{'}\varphi_0^{-1}+\frac{1}{\varphi}y'X)\theta+\frac{1}{\varphi}y^{'}y+\mu_0^{'}\varphi_0^{-1}\mu_0
\end{aligned}
$$

$$
\begin{aligned}
&= \theta^{'}\varphi_1^{-1}\theta-\theta^{'}\varphi_1^{-1}\mu_1-\mu_1^{'}\varphi_1^{-1}\theta \\
&+ \mu_1^{'}\varphi_1^{-1}\mu_1-\mu_1^{'}\varphi_1^{-1}\mu_1+\frac{1}{\varphi}y'y+\mu_0^{'}\varphi_0^{-1}\mu_0 \\
&= (\theta-\mu_1)'\varphi_1^{-1}(\theta-\mu_1)-\mu_1'\varphi_1^{-1}\mu_1\varphi_1^{-1}\mu_1+\frac{1}{\varphi}y'y \\
&+ \mu_0'\varphi_0^{-1}\mu_0.
\end{aligned}
$$

Therefore, the conditional posterior kernel in (31) above can be written as :

$$
p(\theta|\varphi,y,X)\propto
$$

$$
exp(-\frac{1}{2}(\theta-\mu_1)'\varphi_1^{-1}(\theta-\mu_1)exp(-\frac{1}{2}(\frac{1}{\varphi}y'y+\mu_0'\varphi_0^{-1}\mu_0-\mu_1'\varphi_1^{-1}\mu_1)) \tag{32}
$$

Since none of the terms in the second exponent include θ, we simplify the full conditional distribution in (32) to

$$
p(\theta|\varphi,y,X)\propto exp(-\frac{1}{2}(\theta-\mu_1)'\varphi_1^{-1}(\theta-\mu_1)) \tag{33}
$$

Therefore, we have again, the kernel of a multivariate normal density, and we can say that

$$
\theta|\varphi,y,X\sim N(\mu_1,\varphi_1)
$$

where

$$
\varphi_1=(\varphi_0^{-1}+\frac{1}{\varphi}X'X)^{-1}
$$

and

$$
\mu_1=\varphi_1(\varphi_0^{-1}\mu_0+\frac{1}{\varphi}X'y)
$$

to sample from.

Posterior Inference on φ

In order to derive the conditional posterior density for φ, we return to our original expression for the joint posterior given in (30). Ignoring terms that are not related to φ, we have :

$$p(\varphi|\theta,y,X) \propto (\varphi)^{\frac{-n-2v_0-2}{2}} exp(-\frac{1}{2\varphi}(2\tau_0+(y-X\theta)'(y-X\theta))) \quad (34)$$

Comparing this expression with the kernel of the inverse gamma prior specified in (29) above, we have the kernel of another inverse gamma density: Hence

$$\varphi|\theta,y,X \sim IG(v_1,\tau_1) \quad (35)$$

where

$$v_1 = \frac{2v_0+n}{2}$$

and

$$\tau_1 = \frac{2\tau_0+(y-X\theta)'(y-X\theta)}{2}$$

3.3. Estimation of Evolution Variance (Ω_t) with Discount Values

Consider the evolution equation in (5) above,

$$\theta_t = G_t\theta_{t-1}+w_t, w_t \sim N(0,\Omega_t) \quad (36)$$

where Ω_t is the evolution variance and other parameters are as defined earlier. Let

$$\begin{aligned} V(\theta_{t-1}|D_{t-1}) &= V(G_t\theta_{t-1}|D_{t-1}) \\ &= G_tC_{t-1}G_t' \\ &= C_{t-1} \end{aligned}$$

so that

$$V(\theta_t|D_{t-1}) = C_{t-1}+\Omega_t$$

The prior distribution for θ_{t-1} is

$$\theta_{t-1}|D_{t-1} \sim N(m_{t-1},C_{t-1})$$

where $D_{t-1} = (y_1,y_2,...,y_{t-1})$ and the prior distribution for θ_t is

$$\theta_t|D_{t-1} \sim N(m_{t-1},Q_t)$$

where

$$Q_t = C_{t-1}+\Omega_t$$

Therefore,

$$\Omega_t = Q_t - C_{t-1} \quad (37)$$

We introduce the discount factor as a quantity λ such that

$$Q_t = C_{t-1}/\lambda \tag{38}$$

can be interpreted as the percentage of information that passes from time $t-1$ to t.

Therefore, we select the discounting grid $\lambda \in [0.01, 0.99]$. We next develop a sub-algorithm to select optimal granularities of the discount values λ which enable us to conlude the convergence of the model.

4. Parsimonious Model Selection Algorithm (PMSA) for Optimal Model and Discount Value Selection

Since the choice of the evolution variance determines the forcasting performance of DLM, a sub-algorithm for optimal discount value selection with Mean Squared Prediction Error was developed as follows:

1. Init: i=0
2. Let $\lambda_i \in [0.01..0.99]$
3. Compute Ω_t in the DLM with λ_i
4. Estimate one-step ahead predictive density of the specified Bayesian DLM
5. Compute concurrent MSPE of DLM in 3 and cross-validate with the discount value of Ω_{ti}
6. Set $i = i+1$
7. Is the current MSPE lower than the previous one?
8. If No, Go To 6
9. If Yes, Go To 10
10. Stop: Pick the current discount value and DLM as the best.

4.1. Convergence Diagnostics

The convergence diagnostics of Geweke (1993) was used to compare values in the early part of the Markov chain to those in the latter part of the chain in order to detect failure of convergence. The statistic is constructed as follows: Two sub-sequences of the Markov chain θ are taken out, with $\theta_1^t : t = 1, ..., n_1$ and $\theta_2^t : t = n_a, ..., n$ where $1 \leq n_1 \leq n_a < n$.
Let $n_2 = n - n_a + 1$ and define $\bar{\theta}_1 = \frac{1}{n_1}\sum_{t=1}^{n_1}\theta^t$ and $\bar{\theta}_2 = \frac{1}{n_2}\sum_{t=na}^{n}\theta^t$. Geweke test statistics was used to test whether the mean estimates have converged by comparing means from the early and latter part of the Markov chain. Assuming the ratios

$\frac{n_1}{n}$ and $\frac{n_2}{n}$ are fixed, $\frac{n_1+n_2}{n} < 1$, then the following statistic converges to standard normal distribution as n approaches ∞ we have

$$Z_n = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\hat{s_1}(\theta)/n_1 + \hat{s}_2(\theta)/n_2}} \tag{39}$$

where $\hat{s_1}(\theta)$ and $\hat{s_2}(\theta)$ respresent spectral density estimates at zero frequencies. This is a two-sided test and large absolute value $Z-score$ indicates rejection of the null hypothesis of non-stationarity. Effective sample size relates to autocorrelation and measures mixing of the Markov chain. Most often, much discrepancy between the effective sample size and the simulation sample size indicates poor mixing. Effective Sample Size (ESS) is defined as

$$ESS = \frac{n}{\eta} = \frac{n}{1+2\sum_{k=1}^{\infty}\rho_k(\theta)} \tag{40}$$

where n is the total sample size and $\rho_k(\theta)$ is the autocorrelation at lag k for θ. The quantity η is autocorrelation time. The Bayesian process for estimating it is to first find a cut off point k after which the autocorrelations are very close to zero and then sum all the ρ_k to that point. The cut-off point k is such that $\rho_k < 0.01$ or $\rho_k < 2s_k$ where s_k is the standard deviation defined as

$$s_k = 2\sqrt{(\frac{1}{n}(1+2\sum_{j=1}^{k-1}\rho_k^2(\theta)))} \tag{41}$$

In this method, the Lowest Average Granularity Range (AGR) of λ required for convergence and for minimum Mean Squared Prediction Error (MSPE) would be used to determine optimal performance of the DLMs.

4.2. The Modified Recursive Bayesian Algorithm

In summary, the modified recursive Bayesian algorithm for estimating time-varying parameter proceeds as follows:

1. Sample from $p(\theta_T|D_T)$ using the filtering density in section 3.2 . This distribution is assumed to be Normally distributed with parameter $N(h_t,H_t)$ where:

$$h_t = m_t + C_t G_t' R_{t+1}^{-1}(\theta_{t+1} - a_{t+1}) \tag{42}$$

$$H_t = C_t - C_t G_t' R_{t+1}^{-1} G_t C_t' \tag{43}$$

2. Sample from $p(\theta_{T-1}|\theta_T, D_T)$.

3. For the filtering algorithm to run, estimate φ using the Gibbs sampler in section 3.2.1 .

4. Given $(\theta_t|D_t)$, obtain $\Omega_t = C_t(1-\lambda)/\lambda$ via the discounting method in section 3.3

5. Proceed by sampling recursively in this manner for $t+1, t+2$, ...
6. Use the sub-algorithm in section 3.4 to determine AGR required for convergence and when to stop sampling.
7. Sample from $p(\theta_0, ..., \theta_T | D_T)$.
8. Starting from the final density sampled in equation (7) above, the smoothing recursion proceeds backwards in time, using the previously computed filtering and prediction densities.
9. Employ the convergence diagnostics discussed in section 3.4.1 to detect failure or otherwise of convergence of the Markov chain.
10. Use λ and minimum MSPE to assess the performance of the modified algorithm for DLM with FPs and TVPs for various sample sizes.

5. Conclusion

A sound theoretical exposition of how recursive Bayesian algorithms can be employed to model dynamic relationships over time in the presence of discounted evolution variance constituted a major portion of this paper. The modelling of change in the context of widely established concepts in econometrics was addressed by proposing a conceptually implementable Recursive Bayesian Algorithm (RBA) for estimating of time-varying slope parameters (θ_t) in dynamic linear model in the presence of discounted evolution variance. A fast and efficient sub-algorithm for optimal discount value and model selection was also proposed, to determine the average granularities of discount values required for convergence in estimation of time-varying parameters. Future studies will explore the application of this algorithm to simulated and real financial, economic and environmental time series data.

REFERENCES

ABREU, D., PEARCE, D., STACCHETTI, E., (1990). Toward a theory of discounted repeated games with imperfect monitoring. Econometrica, Journal of the Econometric Society, pp. 1041–1063.

ATHANS, M., (1974). The importance of kalman filtering methods for economic systems, In Annals of Economic and Social Measurement, Vol. 3, No. 1, pp. 49–64. NBER.

BELSLEY, D. A., KUTI, E., (1973). Time-varying parameter structures: An overview, In Annals of Economic and Social Measurement, Volume 2, number 4, pp. 375–379. NBER.

BERTSEKAS, D. P., (1976). Dynamic programing and stochastic control, Mathematics in Science and Engineering, Academic Press, New York.

BLANCHARD, O. J., FISCHER, S., (1989). Lectures on macroeconomics, MIT press.

CHETTY, V. K., (1971). Estimation of solowŠs distributed lag models, Econometrica: Journal of the Econometric Society, pp. 99–117.

COOLEY, T. F., PRESCOTT, E. C., (1973). Systematic (non-random) variation models: varying parameter regression: a theory and some applications, In Annals of Economic and Social Measurement, Vol. 2, No. 4, pp. 463–473. NBER.

COOLEY, T. F., PRESCOTT, E. C., (1976). Estimation in the presence of stochastic parameter variation, Econometrica: journal of the Econometric Society, pp. 167–184.

COOPER, R. L., (1972). The predictive performance of quarterly econometric models of the united states, In Econometric Models of Cyclical Behavior, Vol. 1 and 2, pp. 813–947. NBER.

DOH, T., CONNOLLY, M., (2013). The state space representation and estimation of a time- varying parameter VAR with stochastic volatility, Springer.

FUQUENE, J., ALVAREZ, M., PERICCHI, L., (2013). A robust Bayesian dynamic linear model to detect abrupt changes in an economic time series: The case of Puerto Rico, arXiv preprint arXiv:1303.6073.

FÚQUENE, J., M. ÁLVAREZ, PERICCHI, L. R., (2015). A robust Bayesian dynamic linear model for Latin-American economic time series: Şthe Mexico and Puerto Rico casesŤ, Latin American Economic Review 24 (1), pp. 1–17.

GALLANT, A. R., FULLER, W. A., (1973). Fitting segmented polynomial regression models whose join points have to be estimated, Journal of the American Statistical Associa- tion 68 (341), pp. 144–147.

GELFAND, A. E., HILLS, S. E., RACINE-POON, A., SMITH, A. F., (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, Journal of the American Statistical Association 85 (412), pp. 972–985.

GELMAN, A., (2004). Parameterization and Bayesian modeling, Journal of the American Statistical Association 99 (466).

GEMAN, S., GEMAN, D., (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images, Pattern Analysis and Machine Intelligence, IEEE Trans- actions on (6), pp. 721–741.

GEWEKE, J., (1993). Bayesian treatment of the independent student-t linear model, Journal of Applied Econometrics 8(S1), pp. 19–40.

GOLDFELD, S. M., QUANDT, R. E., (1973). A Markov model for switching regressions, Journal of econometrics 1 (1), pp. 3–15.

HARVEY, A., PHILLIPS, G., (1982). The estimation of regression models with time-varying parameters, In Games, economic dynamics, and time series analysis, pp. 306–321. Springer.

HASTINGS, W. K., (1970). Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1), pp. 97–109.

HINKLEY, D. V., (1971). Inference in two-phase regression. Journal of the American Statistical Association 66 (336), 736–743.

KALABA, R., TESFATSION, L., (1980). A least-squares model specification test for a class of dynamic nonlinear economic models with systematically varying parameters, Journal of Optimization Theory and Applications 32 (4), pp. 537–567.

KALABA, R., TESFATSION, L., (1988). The flexible least squares approach to time-varying linear regression. Journal of Economic Dynamics and Control 12(1), 43–48.

MCZGEE, V. E., CARLETON, W. T., (1970). Piecewise regression. Journal of the American Statistical Association 65 (331), pp.1109–1124.

NAKAJIMA, J., KASUYA, M., WATANABE, T., (2011). Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy, Journal of the Japanese and International Economies 25 (3), pp. 225–245.

NG, C., YOUNG, P. C., (1990). Recursive estimation and forecasting of non-stationary time series, Journal of Forecasting 9 (2), pp. 173–204.

PETRIS, G., (2010). An r package for dynamic linear models, Journal of Statistical Soft- ware 36(12), pp. 1–16.

POLLOCK, D., (2003). Recursive estimation in econometrics, Computational statistics data analysis 44 (1), pp. 37–75.

PRIMICERI, G. E., (2005). Time varying structural vector autoregressions and monetary policy, The Review of Economic Studies 72 (3), pp. 821–852.

RAVINES, R., SCHMIDT, A. M, MIGON, H. S, (2006). Revisiting distributed lag models through a Bayesian perspective, Applied Stochastic Models in Business and Industry 22 (2), pp. 193–210.

ROTHENBERG, T. J., (1973). Efficient estimation with a priori information, New Haven: Yale University Press.

SARRIS, A. H., (1973). A Bayesian approach to estimation of time-varying regression co-efficients, In Annals of Economic and Social Measurement, Vol. 2, No. 4, pp.497–520. NBER.

SOLOVIEV, V., SAPTSIN V., CHABANENKO, D., (2011). Markov chains application to the financial-economic time series prediction, arXiv preprint arXiv:1111.5254.

SPEAR, S. E., S. SRIVASTAVA, (1987). On repeated moral hazard with discounting, The Review of Economic Studies 54 (4), pp. 599–617.

WEST, M., HARRISON, P., (1997). Bayesian Forecasting and Dynamic Models (2nd ed.), New York: Springer-verlag.

YOUNG, P. C., (2011). Recursive estimation and time-series analysis: An introduction for the student and practitioner, Springer.

ZELLNER, A., (2009). Bayesian econometrics: past, present, and future. Advances in Econometrics 23, pp. 11–60.