

Article

Bayesian Analysis of Coefficient Instability in Dynamic Regressions

Emanuela Ciapanna and Marco Taboga *

Directorate General for Economics, Statistics and Research, Banca d'Italia, Via Nazionale 91, 00184 Roma, Italy

* Correspondence: marco.taboga@bancaditalia.it; Tel.: +39-06-4792-2355

Received: 2 May 2019; Accepted: 25 June 2019; Published: 28 June 2019



Abstract: This paper deals with instability in regression coefficients. We propose a Bayesian regression model with time-varying coefficients (TVC) that allows to jointly estimate the degree of instability and the time-path of the coefficients. Thanks to the computational tractability of the model and to the fact that it is fully automatic, we are able to run Monte Carlo experiments and analyze its finite-sample properties. We find that the estimation precision and the forecasting accuracy of the TVC model compare favorably to those of other methods commonly employed to deal with parameter instability. A distinguishing feature of the TVC model is its robustness to mis-specification: Its performance is also satisfactory when regression coefficients are stable or when they experience discrete structural breaks. As a demonstrative application, we used our TVC model to estimate the exposures of S&P 500 stocks to market-wide risk factors: We found that a vast majority of stocks had time-varying exposures and the TVC model helped to better forecast these exposures.

Keywords: coefficients' instability; TVC model; Bayesian regression; Monte Carlo experiments

JEL Classification: C11; C12; C22

There is widespread agreement that instability in regression coefficients represents a major challenge in empirical economics. In fact, many equilibrium relationships between economic variables are found to be unstable through time (e.g., [Stock and Watson 1996](#)).

An increasingly popular approach to this problem is to specify regression models with time-varying coefficients (TVC) and estimate the path of their variation (see e.g., [Doan et al. 1984](#); [Cogley and Sargent 2001](#); [Cogley et al. 2010](#); [D'Agostino et al. 2013](#); [Chan 2017](#))¹. Although the estimation of TVC models has been facilitated by advancements in Markov Chain Monte Carlo (MCMC) methods (e.g., [Carter and Kohn 1994](#) and [Chib and Greenberg 1995](#), and, for recent applications, [Mumtaz and Theodoridis 2018](#); [Casarin et al. 2019](#); [Pacífico 2019](#)), it often remains a complex task that requires a careful specification of priors and relies on computationally intensive numerical techniques. For a review and a discussion of the MCMC approach to the estimation of TVC models, see [Petrova \(2019\)](#).

In this paper, we propose a Bayesian TVC model that has low computational requirements and allows one to compute analytically the posterior probability that the regression is stable, the estimates

¹ An alternative approach is to specify and estimate regression models under the hypothesis of constant coefficients, and then test for the presence of structural breaks (e.g., [Chow 1960](#); [Brown et al. 1975](#); [Nyblom 1989](#)) and identify the breakpoints (e.g., [Andrews 1993](#); [Andrews et al. 1996](#); [Bai and Perron 1998](#)).

of the regression coefficients, and several other quantities of interest. Furthermore, it requires minimal input from the econometrician, in the sense that priors are specified automatically: In particular, the only inputs required from the econometrician are regressors and regressands, as in ordinary least squares (OLS) regressions with constant coefficients.

Thanks to the computational tractability of our TVC model, we are able to provide a Monte Carlo study of its finite sample properties (see [Kastner 2019](#) for another recent Monte Carlo study of a time-varying parameter model).

The main goal of the Monte Carlo experiments is to address the concerns of an applied econometrician who suspects that the coefficients of a regression might be unstable, does not know what form of instability to expect, and needs to decide what estimation strategy to adopt.

The first concern we address is loss of efficiency under the null of stability. Suppose that one's data has indeed been generated by a regression with constant coefficients; how much does the econometrician lose, in terms of estimation precision and forecasting accuracy, when they estimate the regression using the TVC model in place of OLS? Our results suggest that the losses from using the TVC model are generally quite small and are comparable to the losses found from using frequentist breakpoint detection procedures², such as [Bai and Perron's \(1998, 2003\)](#) sequential procedure and its model-averaging variant ([Pesaran and Timmerman 2007](#)).

Another concern is the robustness to mis-specification. Suppose one's data has been generated by a regression with few discrete structural breaks; how much does the econometrician lose from using the TVC model instead of standard frequentist procedures for breakpoint detection? The evidence from our Monte Carlo experiments indicates that as well as in this case, the estimation precision and the forecasting accuracy of the TVC model are comparable to those of standard frequentist procedures, the performance of the TVC model being slightly superior or slightly inferior depending on the sample size and on the design of the Monte Carlo experiments.

Finally, a third concern is efficiency: Even in the presence of frequently changing coefficients, does the TVC model provide better estimation precision and forecasting performance than other, possibly mis-specified, models? We found that it generally does and that this gain in efficiency can be quite large.

The Monte Carlo study is also complemented by a brief demonstration of how the TVC model can be applied to a real-world empirical problem. We considered a regression commonly employed to estimate how stock returns are related to market-wide risk factors. We found that the coefficients of this regression are unstable, with a high probability for a vast majority of the stocks included in the S&P 500 index (for applications of time-varying parameter models to stock returns, see also [Bitto and Frühwirth-Schnatter 2019](#) and [Kastner 2019](#)). We also found that the TVC model helps to better predict the exposures of these stocks to the risk factors.

Our model belongs to the family of Class I multi-process dynamic linear models defined by [West and Harrison \(1997\)](#). In our specification, there is a single mixing parameter that takes on finitely many values between 0 and 1. The parameter measures the stability of regression coefficients: If it equals 0, then the regression is stable (coefficients are constant); the closer it is to 1, the more unstable the coefficients are. We propose two measures of stability that can be derived analytically from the posterior distribution of the mixing parameter, one based on credible intervals and one based on posterior odds ratios. We analyzed the performance of a simple decision rule based on these measures of stability: Use OLS if they do not provide enough evidence of instability, otherwise use TVC. We found that such a rule performs well across

² For Monte Carlo studies of frequentist breakpoint detection methods see [Hansen \(2000\)](#) and [Bai and Perron \(2006\)](#).

different scenarios, leading to very small losses under the null of stability and being still able to produce satisfactory results when coefficients are indeed unstable.

Some features of our model are borrowed from existing TVC models (in particular Doan et al. 1984; Stock and Watson 1996; Cogley and Sargent 2001). Furthermore, there are some similarities to the Dynamic Bayesian Model Averaging methods (DBMA; Hoeting et al. 1999; Raftery et al. 2010), which we discuss in Section 2. Other features of our model are completely novel. First of all, we propose an extension of Zellner's (1986) g -prior to dynamic linear models. Thanks to this extension, posterior probabilities and coefficient estimates are invariant to rescalings of the regressors³: This property is essential for obtaining a completely automatic specification of priors. Another novelty of the model is the use of an invariant geometrically-spaced support for the prior distribution of the mixing parameter. We argue that this characteristic of the prior allows the model to capture both very low and very high degrees of coefficients instability, while retaining a considerable parsimony. Our modeling choices have two main practical consequences: (1) The priors are specified in a completely automatic way so that the regressors and regressands are the only input required from the final user⁴, and (2) the computational burden of the model is minimum because analytical estimators are available both for the regression coefficients and for their degree of instability.

The paper is organized as follows: Section 1 introduces the model, Section 2 describes the specification of priors, Section 3 introduces the two measures of (in)stability, Section 4 reports the results of the Monte Carlo experiments, Section 5 contains the empirical application, Section 6 concludes the discussion, and the Appendix A contains proofs and other technical details.

1. The Bayesian Model

We consider a dynamic linear model (according to West and Harrison 1997) with time-varying regression coefficients:

$$y_t = x_t \beta_t + v_t \quad (1)$$

where x_t is a $1 \times k$ vector of observable explanatory variables, β_t is a $k \times 1$ vector of unobservable regression coefficients, and v_t is an i.i.d. disturbance with normal distribution having zero mean and variance V . Time is indexed by t and goes from 1 to T (T is the last observation in the sample).

The vector of coefficients β_t is assumed to evolve according to the following equation:

$$\beta_t = \beta_{t-1} + w_t \quad (2)$$

where w_t is an i.i.d. $k \times 1$ vector of disturbances having a multivariate normal distribution with mean of zero and covariance matrix W . w_t is also contemporaneously and serially independent of v_t . The random walk hypothesis in Equation (2), also adopted by Cogley and Sargent (2001) and Stock and Watson (1996), implies that changes in regression coefficients happen in an unpredictable fashion.

³ Before arriving to the specification of priors proposed in this paper, we tried several other specifications and found that the results can indeed be quite sensitive to rescalings if one chooses other priors.

⁴ A MATLAB function is made available on the internet at https://www.statlect.com/time_varying_regression.htm. The function can be called with the instruction:

$$\text{tvc}(y,X)$$

where y is a $T \times 1$ vector of observations on the dependent variable and X is a $T \times K$ matrix of regressors.

1.1. Notation

Let information available at time t be denoted by D_t . D_t is defined recursively by:

$$D_t = D_{t-1} \cup \{y_t, x_t\}$$

and D_0 contains prior information on the parameters of the model (to be specified below).

We denote by $(z|D_t)$ the distribution of a random vector z , given information at time t and by $p(z|D_t)$ its probability density (or mass) function.

If a random vector z has a multivariate normal distribution with mean m and covariance matrix S , given D_t , we write:

$$(z|D_t) \sim N(m, S)$$

If a random vector z has a multivariate Student's t distribution with mean m , scale matrix S , and n degrees of freedom, we write:

$$(z|D_t) \sim T(m, S, n)$$

and its density is parameterized as follows:

$$p(z|D_t) \propto \left[n + (z - m)^T S^{-1} (z - m) \right]^{\frac{k+n}{2}}$$

If z has a Gamma distribution with parameters V and n , we write:

$$(z|D_t) \sim G(V, n)$$

and its density is parameterized as follows:

$$p(z|D_t) = \frac{(Vn/2)^{n/2} z^{n/2-1} \exp(-Vnz/2)}{\Gamma(n/2)}$$

Finally, define $W^* = V^{-1}W$ and denote by X the design matrix,

$$X = [x_1^T \quad \dots \quad x_T^T]^T$$

1.2. Structure of Prior Information and Updating

In this subsection we state the main assumptions on the structure of prior information and derive the formulae for updating the priors analytically.

The first set of assumptions regards β_1 , the vector of regression coefficients at time $t = 1$, and V , the variance of the regression disturbances. We impose on β_1 and V a conjugate normal-inverse Gamma prior, i.e.,

- β_1 has a multivariate normal distribution conditional on V , with known mean $\hat{\beta}_{1,0}$, and covariance equal to $V \cdot F_{\beta,1,0}$ where $F_{\beta,1,0}$ is a known matrix;
- The reciprocal of V has a Gamma distribution, with known parameters \hat{V}_0 and n_0 .

The second set of assumptions regards W . We assume that W is proportional to the prior variance of β_1 . In particular, we assume that:

$$W = V\lambda(\theta) F_{\beta,1,0}$$

where θ is a random variable with finite support $R_\theta = \{\theta_1, \dots, \theta_q\} \subset [0, 1]$, $\theta_1 = 0$, and $\lambda(\theta)$ is strictly increasing in θ , such that $\lambda(\theta_1) = 0$. Hence, we assume that the innovations w_t to the regression coefficients have a variance proportional to the prior variance of the regression coefficients at time $t = 1$. The assumption that $W^* \propto F_{\beta,1,0}$ is made also by Doan et al. (1984) in their seminal paper on TVC models⁵.

The coefficient of proportionality $\lambda(\theta)$ is an (unknown) random variable whose posterior distribution can be computed analytically (see below). When $\theta = \theta_1$ (and $\lambda = 0$), the variance of w_t is zero and regression coefficients are stable. On the contrary, when $\theta \neq \theta_1$ (and $\lambda > 0$), w_t has a positive variance and the regression coefficients are unstable (i.e., they change through time). The higher θ is, the greater the variance of w_t is, and the more unstable the regression coefficients are.

The assumptions on the priors and the initial information are summarized as follows:

Summary 1. *The priors on the unknown parameters are:*

$$\begin{aligned} (\beta_1 | D_0, V, \theta) &\sim N(\widehat{\beta}_{1,0}, VF_{\beta,1,0}) \\ (1/V | D_0, \theta) &\sim G(\widehat{V}_0, n_0) \\ p(\theta_i | D_0) &= p_{0,i}, i = 1, \dots, q \end{aligned}$$

and the initial information set is:

$$D_0 = \{\widehat{\beta}_{1,0}, F_{\beta,1,0}, \widehat{V}_0, n_0, p_{0,1}, \dots, p_{0,q}\}$$

Given the above assumptions, the posterior distributions of the parameters of the regression can be calculated as follows:

Proposition 1. *Let priors and initial information be as in Summary 1. Let $p_{t,i} = p(\theta = \theta_i | D_t)$. Then:*

$$\begin{aligned} p(\beta_t | D_{t-1}) &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_{t-1}) p_{t-1,i} \\ p(y_t | D_{t-1}, x_t) &= \sum_{i=1}^q p(y_t | \theta = \theta_i, D_{t-1}, x_t) p_{t-1,i} \\ p(1/V | D_{t-1}) &= \sum_{i=1}^q p(1/V | \theta = \theta_i, D_{t-1}) p_{t-1,i} \\ p(\beta_t | D_t) &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_t) p_{t,i} \end{aligned}$$

⁵ However, in their model the coefficients β_t do not follow a random walk (they are mean reverting). They also use different priors: While we impose Zellner’s g-prior on β_1 (see Section 2), they impose the Minnesota prior.

where,

$$\begin{aligned}
 (\beta_t | \theta = \theta_i, D_{t-1}) &\sim T(\hat{\beta}_{t,t-1,i}, \hat{V}_{t-1,i} F_{\beta,t,t-1,i}, n_{t-1,i}) \\
 (y_t | \theta = \theta_i, D_{t-1}, x_t) &\sim T(\hat{y}_{t,t-1,i}, \hat{V}_{t-1,i} F_{y,t,t-1,i}, n_{t-1,i}) \\
 (1/V | \theta = \theta_i, D_{t-1}) &\sim G(\hat{V}_{t-1,i}, n_{t-1,i}) \\
 (\beta_t | \theta = \theta_i, D_t) &\sim T(\hat{\beta}_{t,t,i}, \hat{V}_{t,i} F_{\beta,t,t,i}, n_{t,i})
 \end{aligned}$$

The parameters of the above distributions are obtained recursively as:

$$\begin{aligned}
 \hat{\beta}_{t,t-1,i} &= \hat{\beta}_{t-1,t-1,i} & F_{\beta,t,t-1,i} &= F_{\beta,t-1,t-1,i} + \lambda(\theta_i) F_{\beta,1,0} \\
 \hat{y}_{t,t-1,i} &= x_t \hat{\beta}_{t,t-1,i} & F_{y,t,t-1,i} &= 1 + x_t F_{\beta,t,t-1,i} x_t^\top \\
 e_{t,i} &= y_t - \hat{y}_{t,t-1,i} & P_{t,i} &= F_{\beta,t,t-1,i} x_t^\top / F_{y,t,t-1,i} \\
 \hat{\beta}_{t,t,i} &= \hat{\beta}_{t,t-1,i} + P_{t,i} e_{t,i} & F_{\beta,t,t,i} &= F_{\beta,t,t-1,i} - P_{t,i} P_{t,i}^\top F_{y,t,t-1,i} \\
 n_{t,i} &= n_{t-1,i} + 1 & \hat{V}_{t,i} &= \frac{1}{n_{t,i}} \left(n_{t-1,i} \hat{V}_{t-1,i} + \frac{e_{t,i}^2}{F_{y,t,t-1,i}} \right)
 \end{aligned}$$

starting from the initial conditions $\hat{\beta}_{1,0,i} = \hat{\beta}_{1,0}$, $F_{\beta,1,0,i} = F_{\beta,1,0}$, $\hat{V}_{0,i} = \hat{V}_0$, and $n_{0,i} = n_0$. The mixing probabilities are obtained recursively as:

$$p_{t,i} = \frac{p_{t-1,i} p(y_t | \theta = \theta_i, D_{t-1}, x_t)}{\sum_{j=1}^q p_{t-1,j} p(y_t | \theta = \theta_j, D_{t-1}, x_t)}$$

starting from the prior probabilities $p_{0,1}, \dots, p_{0,q}$.

The updated mixing probabilities in the above proposition can be interpreted as posterior model probabilities, where a model is a TVC regression with fixed θ . Hence, for example, $p_{T,1}$ is the posterior probability of the regression model with stable coefficients ($\theta = 0$). A crucial property of the framework we propose is that posterior model probabilities are known analytically: They can be computed exactly, without resorting to simulations.

We note that the above recursive equations bear some similarities to those found in Dynamic Bayesian Model Averaging (DBMA; Hoeting et al. 1999; Raftery et al. 2010). The main difference is in the update of the covariance matrix of coefficients. In our model we have:

$$F_{\beta,t,t-1,i} = F_{\beta,t-1,t-1,i} + \lambda(\theta_i) F_{\beta,1,0}$$

while in DBMA, there is:

$$F_{\beta,t,t-1,i} = \lambda^{-1} F_{\beta,t-1,t-1,i}$$

where λ is a forgetting factor that was exogenously set at $\lambda = 0.99$ by Raftery et al. (2010). Furthermore, while in DBMA a forgetting factor is also applied to model probabilities $p_{t,i}$, the recursions for mixing probabilities in our model have no forgetting factor because we assume that θ stays constant through time. Applying a forgetting factor as in DBMA to the mixing probabilities would be a straightforward extension of our model. Finally, while Raftery et al. (2010) use a recursive moment estimator for $\hat{V}_{t,i}$, we derive it in a fully Bayesian way.

In the above proposition, the priors on the regression coefficients β_t in a generic time period t are updated using information received up to that same time t only. However, after observing the whole sample (up to time T), one might want to revise the priors on the regression coefficients β_i in previous

time periods ($t < T$), using the information subsequently received. This revision (usually referred to as smoothing) can be accomplished using the results of the following proposition:

Proposition 2. *Let priors and initial information be as in Summary 1. Then:*

$$p(\beta_{T-\tau} | D_T) = \sum_{i=1}^q p(\beta_{T-\tau} | \theta = \theta_i, D_T) p_{T,i}$$

where,

$$(\beta_{T-\tau} | \theta = \theta_i, D_T) \sim T(\widehat{\beta}_{T-\tau,T,i}, \widehat{V}_{T,i}, F_{\beta,T-\tau,T,i}, n_{T,i})$$

The mixing probabilities $p_{T,i}$ and the parameters $\widehat{V}_{T,i}$, $n_{T,i}$ are obtained from the recursions in Proposition 1 while the parameters $\widehat{\beta}_{T-\tau,T,i}$ and $F_{\beta,T-\tau,T,i}$ are obtained from the following backward recursions:

$$\begin{aligned} Q_{T-\tau,i} &= F_{\beta,T-\tau,T-\tau,i} (F_{\beta,T-\tau+1,T-\tau,i})^{-1} \\ \widehat{\beta}_{T-\tau,T,i} &= \widehat{\beta}_{T-\tau,T-\tau,i} + Q_{T-\tau,i} (\widehat{\beta}_{T-\tau+1,T,i} - \widehat{\beta}_{T-\tau+1,T-\tau,i}) \\ F_{\beta,T-\tau,T,i} &= F_{\beta,T-\tau,T-\tau,i} + Q_{T-\tau,i} (F_{\beta,T-\tau+1,T,i} - F_{\beta,T-\tau+1,T-\tau,i}) Q_{T-\tau,i}^T \end{aligned}$$

starting from $\tau = 1$ and taking as the final conditions the values $\widehat{\beta}_{T-1,T-1,i}$, $\widehat{\beta}_{T,T,i}$, $\widehat{\beta}_{T,T-1,i}$, $F_{\beta,T-1,T-1,i}$, $F_{\beta,T,T,i}$ and $F_{\beta,T,T-1,i}$ obtained from the recursions in Proposition 1.

Other important quantities of interest are known analytically, as shown by the following lemma:

Lemma 1. *The following quantities are also known analytically:*

$$\begin{aligned} E[\beta_t | D_s] &= \sum_{i=1}^q p_{s,i} E[\beta_t | \theta = \theta_i, D_s] \\ E[y_t | D_s] &= \sum_{i=1}^q p_{s,i} E[y_t | \theta = \theta_i, D_s] \\ \text{Var}[\beta_t | D_s] &= \sum_{i=1}^q p_{s,i} \text{Var}[\beta_t | \theta = \theta_i, D_s] + \sum_{i=1}^q p_{s,i} E[\beta_t | \theta = \theta_i, D_s] E[\beta_t^T | \theta = \theta_i, D_s] \\ &\quad - \left(\sum_{i=1}^q p_{s,i} E[\beta_t | \theta = \theta_i, D_s] \right) \left(\sum_{i=1}^q p_{s,i} E[\beta_t | \theta = \theta_i, D_s] \right)^T \\ \text{Var}[y_t | D_s] &= \sum_{i=1}^q p_{s,i} \text{Var}[y_t | \theta = \theta_i, D_s] + \sum_{i=1}^q p_{s,i} E[y_t | \theta = \theta_i, D_s]^2 \\ &\quad - \left(\sum_{i=1}^q p_{s,i} E[y_t | \theta = \theta_i, D_s] \right)^2 \end{aligned}$$

where $E[\beta_t | \theta = \theta_i, D_s]$, $E[y_t | D_s]$, $\text{Var}[\beta_t | D_s]$, and $\text{Var}[y_t | D_s]$ are calculated for each θ_i as in Propositions 1 and 2.

Thus, parameter estimates ($E[\beta_t | D_s]$) and predictions ($E[y_t | D_s]$) in any time period can be computed analytically and their variances are known in closed form. The probability distributions of β_t and y_t in a certain time period given information D_s are mixtures of Student's t distributions. Their quantiles are not known analytically, but they are easy to simulate by Monte Carlo methods. For example, if the

distribution of β_T conditional on D_T is the object of interest, one can set up a Monte Carlo experiment where each simulation is conducted in two steps: (1) Extract z from a uniform distribution on $[0, 1]$; find k^* such that $k^* = \arg \min \left\{ k : \sum_{i=1}^k p_{T,i} \geq z \right\}$, and (2) given k^* , extract β_T from the Student's t distribution $(\beta_T | \theta = \theta_{k^*}, D_T)$, which is given by Propositions 1 and 2. The empirical distribution of the Monte Carlo simulations of β_T thus obtained is an estimate of the distribution of β_T conditional on D_T .

2. The Specification of Priors

Our specification of priors aims to be:

1. *Objective*, in the sense that it does not require elicitation of subjective priors;
2. *Fully automatic*, in the sense that the model necessitates no inputs from the econometrician other than regressors and regressands, as in OLS regressions with constant coefficients.

The above goals are pursued by extending Zellner's (1986) g -prior to TVC models and by parameterizing $\lambda(\theta)$ in such a way that the support of θ is invariant (it does not need to be specified on a case-by-case basis).

2.1. The Prior Mean and Variance of the Coefficients

We use a version of Zellner's (1986) g -prior for the prior distribution of the regression coefficients at time $t = 1$. The prior mean is zero, corresponding to a prior belief of no predictability:

$$\hat{\beta}_{1,0} = 0 \tag{3}$$

while the prior covariance matrix is proportional to $(X^T X)^{-1}$:

$$F_{\beta,1,0} = g (X^T X)^{-1} \tag{4}$$

where g is a coefficient of proportionality.

Zellner's (1986) g -prior is widely used in model selection and model averaging problems similar to ours (we have a range of regression models featuring different degrees of instability) because it greatly reduces the sensitivity of posterior model probabilities to the specification of prior distributions (Fernandez et al. 2001), thus helping to keep the analysis as objective as possible. Furthermore, Zellner's (1986) g -prior has a straightforward interpretation: It can be interpreted as information provided by a conceptual sample having the same design matrix X as the current sample (Zellner 1986; George and McCulloch 1997; Smith and Kohn 1996).

To keep the prior relatively uninformative, we follow Kass and Wasserman (1995) and choose $g = T$ (see also Shively et al. 1999). Thus, the amount of prior information (in the Fisher sense) about the coefficients is equal to the average amount of information contained in one observation from the sample.

Given the assumption that $W^* = \lambda(\theta) F_{\beta,1,0}$ (see previous section), Zellner's prior (4) implies that the covariance matrix of w_t is also proportional to $(X^T X)^{-1}$:

$$W^* \propto (X^T X)^{-1} \tag{5}$$

The above proportionality has also been assumed in a TVC model by [Stock and Watson \(1996\)](#)⁶, who borrow it from [Nyblom \(1989\)](#). A similar hypothesis is also adopted by [Cogley and Sargent \(2001\)](#)⁷.

Equations (4) and (3), together with Equation (5), imply that all the coefficients β_t have zero prior mean and covariance proportional to $(X^T X)^{-1}$, given D_0 :

$$\begin{aligned} E[\beta_t | D_0, V, \theta] &= 0, t = 1 \dots, T, \forall \theta \\ \text{Var}[\beta_t | D_0, V, \theta] &= [T + (t - 1) \lambda(\theta)] V (X^T X)^{-1}, t = 1, \dots, T, \forall \theta \end{aligned}$$

As a consequence, the model has the desirable property that the posterior model probabilities are scale invariant in the covariates ([Nyblom 1989](#)).

2.2. The Variance Parameters \hat{V}_0 and n_0

In objective Bayesian analyses, the prior usually assigned to V in conjunction with [Zellner's \(1986\)](#) g -prior (e.g., [Liang et al. 2008](#)) is the improper prior:

$$p(V | D_0, \theta) \propto V^{-1}$$

With this choice, the updating equations in Proposition 1 would have to be replaced with a different set of updating equations until the first non-zero observation of y_t ([West and Harrison 1997](#)). Furthermore, the updating of posterior probabilities would be slightly more complicated. To avoid the subtleties involved in using an improper prior, we adopt a simpler procedure that yields almost identical results in reasonably sized samples. We use the first observation in the sample (denote it by y_0) to form the prior on V :

$$\begin{aligned} \hat{V}_0 &= y_0^2 \\ n_0 &= 1 \end{aligned}$$

We then discard the observation and start the updating in Proposition 1 from the subsequent observation. If the first observation is zero ($y_0 = 0$) we discard it and use the next to form the prior (or repeat until we find the next non-zero observation).

2.3. The Mixing Parameter θ and the Prior Mixing Probabilities $p_{0,i}$

When relaxing the assumption that the matrix W^* is known, we have assumed that:

$$W^* = \lambda(\theta) F_{\beta,1,0}$$

where θ is a random variable with finite support $R_\theta = \{\theta_1, \dots, \theta_q\} \subset [0, 1]$, $\theta_1 = 0$ and $\lambda(\theta)$ is strictly increasing in θ , such that $\lambda(\theta_1) = 0$. We now propose a specification of the function $\lambda(\theta)$ that satisfies the above requirements and allows for an intuitive interpretation of the parameter θ , while also facilitating the specification of a prior distribution for θ .

⁶ However, they assume that $F_{\beta,1,0}$ is proportional to the identity matrix, while we assume that also $F_{\beta,1,0}$ is proportional to $(X^T X)^{-1}$. Furthermore, they do not estimate V . Their analysis is focused on the one-step-ahead predictions of y_t , which can be computed without knowing V . They approach the estimation of λ in a number of different ways, but none of them allows one to derive analytically a posterior distribution for λ .

⁷ In their model the prior covariance of w_t is proportional to $(X^T X)^{-1}$, but X is the design matrix of a pre-sample not used for the estimation of the model.

First, note that:

$$y_t = x_t \beta_{t-1} + x_t w_t + v_t$$

Hence, given λ and the initial information D_0 , the variance generated by innovations at time t is:

$$\text{Var} [x_t w_t + v_t | D_0, \lambda] = \lambda \widehat{V}_0 x_t F_{\beta,1,0} x_t^\top + \widehat{V}_0$$

We assume that θ is the fraction of this variance generated on average by innovations to the coefficients:

$$\theta = \frac{\frac{1}{T} \sum_{t=1}^T \text{Var} [x_t w_t | D_0, \lambda]}{\frac{1}{T} \sum_{t=1}^T \text{Var} [x_t w_t + v_t | D_0, \lambda]}$$

When θ satisfies the above property, it is immediate to prove that:

$$\lambda (\theta) = \frac{\theta}{\omega - \theta \omega}$$

where,

$$\omega = \frac{1}{T} \sum_{t=1}^T x_t F_{\beta,1,0} x_t^\top$$

The function $\lambda (\theta)$ is strictly increasing in θ , such that $\lambda (0) = 0$, as required. Hence, when $\theta = 0$, the regression has stable coefficients. Furthermore, by an appropriate choice of θ , any degree of coefficient instability can be reproduced (when θ tends to 1, λ approaches infinity).

The support of θ is a geometrically spaced grid, consisting of q points:

$$R_\theta = \left\{ 0, \theta_{\max} c^{q-2}, \theta_{\max} c^{q-1}, \dots, \theta_{\max} c^1, \theta_{\max} \right\}$$

where $0 \leq \theta_{\max} < 1$ and $0 < c < 1$. θ_{\max} cannot be chosen to be exactly equal to 1 (in which case $\lambda (\theta_{\max}) = \infty$), but it can be set equal to any number arbitrarily close to 1.

If the grid is considered as an approximation of a finer set of points (possibly a continuum), the geometric spacing ensures that the maximum percentage round-off error is constant on all subintervals $[\theta_i, \theta_{i+1}]$ ($1 < i < q$). $\frac{1-c}{2}$ is the constant that bounds the percentage round-off error.

Using a geometrically spaced grid is the natural choice when the order of magnitude of a parameter is unknown (e.g., [Guerre and Lavergne 2005](#), [Horowitz and Spokoiny 2001](#); [Lepski et al. 1997](#), and [Spokoiny 2001](#)). In our case, it allows one to simultaneously consider both regressions that are very close to being stable and regressions that are far from being stable, without requiring too fine a grid⁸.

Assuming prior ignorance on the order of magnitude of θ , we assign equal probability to each point in the grid:

$$p (\theta_i | D_0) = q^{-1}, i = 1, \dots, q$$

Note that, given the above choices, the prior on θ and its support are invariant, in the sense that they do not depend on any specific characteristic of the data to be analyzed, but they depend only on the maximum percentage round-off error $\frac{1-c}{2}$. As a consequence, they allow the specification of priors to remain fully automatic.

⁸ For example, in the empirical part of the paper, setting $q = 100$ and $c = 0.9$, we are able to simultaneously consider 5 different orders of magnitude of instability. With the same number of points q and an arithmetic grid, we would have been able to consider only 2 orders.

3. Measures of (In)Stability

After computing the posterior distribution of θ , a researcher might naturally ask: How much evidence did the data provide against the hypothesis of stability? Here, we discuss some possible ways to answer this question.

The crudest way to evaluate instability is to look at the posterior probability that $\theta = 0$. The closer to 1 this probability is, the more evidence of stability we have. However, a low posterior probability that $\theta = 0$ does not necessarily constitute overwhelming evidence of instability. It might simply be the case that the sample is not large enough to satisfactorily discriminate, a posteriori, between stable and unstable regressions: In such cases, even if the true regression is stable, unstable regressions might be assigned posterior probabilities that are only marginally lower than the probability of the stable one. Furthermore, if R_θ contains a great number of points, it can happen that the posterior probability that $\theta = 0$ is close to zero, but still much higher than the posterior probability of all the other points.

We propose two measures of stability to help circumvent the above shortcomings. The first one is based on credible intervals (e.g., Robert 2007). The second one is based on posterior odds ratios.

Define the higher posterior probability set H_θ as follows:

$$H_\theta = \{\theta_i \in R_\theta : p(\theta = \theta_i | D_T) > p(\theta = 0 | D_T)\}$$

i.e., H_θ contains all points of R_θ having higher posterior probability than $\theta = 0$ (remember that $\theta = 0$ means that regression coefficients are stable). Define Π as follows:

$$\Pi = 1 - \frac{\sum_{\theta_i \in H_\theta} p(\theta = \theta_i | D_T)}{\sum_{\theta_i \neq 0} p(\theta = \theta_i | D_T)}$$

where we adopt the convention $0/0 = 0$.

When $\Pi = 1$, $\theta = 0$ is a mode of the posterior distribution of θ : We attach to the hypothesis of stability a posterior probability that is at least as high as the posterior probability of any alternative hypothesis of instability. On the contrary, when $\Pi = 0$, the posterior probability assigned to the hypothesis of stability is so low that all unstable models are more likely than the stable one, a posteriori. In the intermediate cases ($0 < \Pi < 1$), Π provides a measure of how far the hypothesis of stability is from being the most likely hypothesis (the lower Π , the less likely stability is).

The second measure we propose is based on the probability of the posterior mode of θ . Define:

$$p^* = \max(\{p(\theta = \theta_i | D_T) : \theta_i \in R_\theta\})$$

p^* is the probability of (one of) the mode(s) of the posterior distribution of θ , i.e., the probability of the most likely value(s) of θ .

Using p^* , we construct our second measure of stability as a posterior odds ratio:

$$\pi = \frac{p(\theta = 0 | D_T)}{p^*}$$

As with the previously proposed measure, when $\pi = 1$, $\theta = 0$ is a mode of the posterior distribution of θ and stability is the most likely hypothesis, a posteriori. On the contrary, the closer π is to zero, the less likely stability is, when compared with the most likely hypothesis. For example, when $\pi = 1/10$, there is an unstable regression that is 10 times more likely than the stable one.

Both measures of stability (Π and π) can be used to make decisions. For example, one can fix a threshold τ and decide to reject the hypothesis of stability if the measure of stability is below the

threshold ($\Pi < \tau$ or $\pi < \tau$). In that case that Π is used, the procedure can be assimilated to a frequentist test of hypothesis, where $1 - \tau$ represents the level of confidence. Π can be interpreted as a sort of Bayesian p -value (e.g., Robert 2007): The lower Π is, the higher is the confidence with which we can reject the hypothesis of stability⁹. In that case that π is used, one can resort to Jeffreys's (1961) scale to qualitatively assess the strength of the evidence against the hypothesis of stability (e.g., substantial evidence if $\frac{1}{3} \leq \pi < \frac{1}{10}$, strong evidence if $\frac{1}{10} \leq \pi < \frac{1}{30}$, very strong evidence if $\frac{1}{30} \leq \pi < \frac{1}{100}$).

In the next section we explore the consequences of using these decision rules to decide whether to estimate a regression by OLS or by TVC.

4. Monte Carlo Evidence

4.1. Performance When the Data Generating Process (DGP) Is a Stable Regression

In this subsection we present the results of a set of Monte Carlo simulations aimed at evaluating how much efficiency is lost when a stable regression is estimated with our TVC model. We compare the forecasting performance and the estimation precision of the TVC model with those of OLS and of a standard frequentist procedure used to identify breakpoints and estimate regression coefficients in the presence of structural breaks. In particular, we consider the performance of Bai and Perron's (1998, 2003) sequential procedure, as implemented by Pesaran and Timmerman (2002, 2007).

For our Monte Carlo experiments, we adapt a design that has already been employed in the literature on parameter instability (Hansen 2000).

The design is as follows:

- *Data generating process:* y_t is generated according to:

$$y_t = \rho y_{t-1} + u_{t-1} + v_t$$

where $y_0 = 0$, $u_t \sim T(0, 1, 5)$ i.i.d., $v_t \sim N(0, 1)$ i.i.d., and u_t and v_t are serially and cross-sectionally independent;

- *Estimated equations:* Two equations are estimated. In the first case, a constant and the first lags of y_t and u_t are included in the set of regressors; hence, the estimated model is (1), where:

$$x_t = \begin{bmatrix} 1 & y_{t-1} & u_{t-1} \end{bmatrix}$$

In the second case, a constant and the first three lags of y_t and u_t are included in the set of regressors; hence, the estimated model is (1), where:

$$x_t = \begin{bmatrix} 1 & y_{t-1} & y_{t-2} & y_{t-3} & u_{t-1} & u_{t-2} & u_{t-3} \end{bmatrix}$$

- *Parameters of the design:* Simulations are conducted for three different sample sizes ($T = 100, 200, 500$), four different values of the autoregressive coefficient ($\rho = 0, 0.50, 0.80, 0.99$), and the two estimated equations detailed above, for a total of 24 experiments.

⁹ Note, however, that the parallelism can be misleading, as Bayesian p -values have only a frequentist validity in special cases. Ghosh and Mukerjee (1993); Mukerjee and Dey (1993); Datta and Ghosh (1995), and Datta (1996) provide conditions that priors have to satisfy in order for Bayesian p -values to have also frequentist validity.

Each Monte Carlo experiment consists of 10,000 simulations. The loss in estimation precision is evaluated by comparing the estimate of the coefficient vector at time T (denote it by $\tilde{\beta}_T$) with its true value. We consider seven different estimates:

- Model averaging (TVC-MA) estimates, where:

$$\tilde{\beta}_T = E[\beta_T | D_T] = \sum_{i=1}^q p_{T,i} E[\beta_T | D_T, \theta = \theta_i]$$

- Model selection (TVC-MS) estimates, where:

$$\tilde{\beta}_T = E[\beta_T | D_T, \theta = \theta_{j^*}]$$

and

$$j^* = \arg \max_j p_{T,j}$$

i.e., only the model with the highest posterior probability is used to make predictions;

- Estimates obtained from the regression model with stable coefficients when $\Pi \geq 0.1$ and from model averaging when $\Pi < 0.1$ (denoted by TVC- Π):

$$\tilde{\beta}_T = \begin{cases} E[\beta_T | D_T, \theta = 0] & \text{if } \Pi \geq 0.1 \\ E[\beta_T | D_T] & \text{if } \Pi < 0.1 \end{cases}$$

i.e., coefficients are estimated with the TVC model only if there is enough evidence of instability ($\Pi < 0.1$); otherwise, the standard OLS regression is used. This is intended to reproduce the outcomes of a decision rule whereby the econometrician uses the TVC model only if the TVC model itself provides enough evidence that OLS is inadequate;

- Estimates obtained from the regression model with stable coefficients when $\pi \geq 0.1$ and from model averaging when $\pi < 0.1$ (denoted by TVC- π):

$$\tilde{\beta}_T = \begin{cases} E[\beta_T | D_T, \theta = 0] & \text{if } \pi \geq 0.1 \\ E[\beta_T | D_T] & \text{if } \pi < 0.1 \end{cases}$$

This estimator is similar to the previous one, but π is used in place of Π to decide whether there is enough evidence of instability;

- Estimates obtained from the regression model with stable coefficients (OLS):

$$\tilde{\beta}_T = E[\beta_T | D_T, \theta = 0]$$

- OLS estimates obtained from [Bai and Perron's \(1998, 2003\)](#) sequential¹⁰ procedure (denoted by BP), using the Schwarz information criterion (SIC) criterion to choose the number of breakpoints ([Pesaran](#)

¹⁰ We estimate the breakpoint dates sequentially rather than simultaneously to achieve a reasonable computational speed in our Monte Carlo simulations. Denote by v_S the number of breakpoints estimated by the sequential procedure and by v_σ the number estimated by the simultaneous procedure. Given that we are using the SIC criterion to choose the number of points, if $v_\sigma \leq 1$, then $v_S = v_\sigma$; otherwise, if $v_\sigma > 1$, then $v_S \leq v_\sigma$. Therefore, in our Monte Carlo simulations (where the true number of breakpoints is either 0 or 1), the sequential procedure provides a better estimate of the number of breakpoints than the simultaneous procedure.

and Timmerman 2002 and 2007). If $\tilde{\tau}$ is the last estimated breakpoint date in the sample, then $\tilde{\beta}_T$ is the OLS estimate of β_T obtained by using all the sample points from $\tilde{\tau}$ to T ;

- Estimates obtained from Pesaran and Timmerman’s (2007) model-averaging procedure (denoted by BP-MA): The location of the last breakpoint is estimated with Bai and Perron’s procedure (as in the point above); if $\tilde{\tau}$ is the last estimated breakpoint date in the sample, then:

$$\tilde{\beta}_T = \sum_{\tau=1}^{\tilde{\tau}} w_{\tau} \tilde{\beta}_{T,\tau}$$

where $\tilde{\beta}_{T,\tau}$ is the OLS estimate of β_T obtained using all the sample points from τ to T ; w_{τ} is a weight proportional to the inverse of the mean squared prediction error committed when using only the sample points from τ onwards to estimate the regression and predict y_t ($\tau + k + 1 \leq t \leq T$).

The Monte Carlo replications are used to estimate the mean squared error (MSE) of the coefficient estimates:

$$MSE_j^{\beta} = E \left[\left\| \beta_T - \tilde{\beta}_T \right\|^2 \right]$$

where $\| \cdot \|$ is the Euclidean norm and $j = \text{TVC-MA, TVC-MS, TVC-II, TVC-}\pi, \text{OLS, BP, BP-MA}$ depending on which of the above methods has been used to estimate β_T .

The two parameters regulating the granularity of the grid for θ are chosen as follows: $q = 100$ and $c = 0.9$. To avoid degeneracies, rather than setting $\theta_{\max} = 1$ (the theoretical upper bound on θ), we chose a value that is numerically close to 1 ($\theta_{\max} = 0.999$). Thus, the roundoff error is bounded at 5 per cent and the model is able to detect both very high and very low degrees of instability (as low as $\theta \simeq 3 \cdot 10^{-5}$).

Panel A of Table 1 reports the Monte Carlo estimates of MSE_j^{β} for the case in which x_t includes only the first lags of y_t and u_t . Not surprisingly, the smallest MSE is in all cases achieved by the OLS estimates. There are significant differences between the case in which the autoregressive component is very persistent ($\rho = 0.99$) and the other cases ($\rho = 0, 0.50, 0.80$). In the latter cases, the TVC- π coefficient estimates are those that yield the smallest increase in MSE with respect to OLS (in most cases under 5 per cent). The performance of BP-MA is the second best, being only slightly inferior to that of TVC- π , but slightly superior to that of TVC-II. $MSE_{\text{TVC-MA}}^{\beta}$ and $MSE_{\text{TVC-MS}}^{\beta}$ are roughly between 20 per cent and 60 per cent higher than $MSE_{\text{TVC-OLS}}^{\beta}$, while $MSE_{\text{TVC-BP}}^{\beta}$ is on average equal to several multiples of $MSE_{\text{TVC-OLS}}^{\beta}$. Qualitatively speaking, the loss in precision from using TVC-II, TVC- π , and BP-MA is almost negligible, while there is a severe loss using BP and a moderate loss using TVC-MA and TVC-MS. In the case in which $\rho = 0.99$, results are very different: On average, MSE_{TVC}^{β} (all four kinds of TVC) and MSE_{BP}^{β} become almost two orders of magnitude greater than MSE_{OLS}^{β} , while $MSE_{\text{BP-MA}}^{\beta}$ remains comparable to MSE_{OLS}^{β} (although there is a worsening with respect to the case of low persistence).

The unsatisfactory performance of the TVC and BP estimates in the case of high persistence can arguably be explained by an identification problem. In the unit root case, the regression generating the data is:

$$y_t = y_{t-1} + u_{t-1} + v_t$$

Table 1. Estimation precision when coefficients are stable—Monte Carlo evidence—MSE of coefficient estimates.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.0257	0.0286	0.0239	0.0215	0.0205	0.0540	0.0221
T = 200	0.0138	0.0142	0.0121	0.0106	0.0102	0.0479	0.0108
T = 500	0.0066	0.0056	0.0049	0.0043	0.0040	0.0070	0.0043
$\rho = 0.50$							
T = 100	0.0271	0.0303	0.0249	0.0219	0.0207	0.0730	0.0227
T = 200	0.0136	0.0141	0.0119	0.0104	0.0099	0.0329	0.0106
T = 500	0.0064	0.0055	0.0048	0.0041	0.0038	0.0132	0.0039
$\rho = 0.80$							
T = 100	0.0308	0.0352	0.0278	0.0234	0.0219	0.1537	0.0272
T = 200	0.0141	0.0145	0.0121	0.0103	0.0097	0.0853	0.0105
T = 500	0.0062	0.0053	0.0046	0.0039	0.0037	0.0179	0.0038
$\rho = 0.99$							
T = 100	4.7933	5.3498	4.7233	4.7344	0.0666	1.9222	0.1145
T = 200	2.5390	2.7297	2.5072	2.5057	0.0244	0.7189	0.0367
T = 500	0.4448	0.4792	0.4301	0.4287	0.0062	0.1338	0.0072
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.0885	0.0964	0.0860	0.0804	0.0785	0.0827	0.0795
T = 200	0.0433	0.0453	0.0414	0.0385	0.0376	0.0379	0.0378
T = 500	0.0186	0.0175	0.0162	0.0150	0.0146	0.0146	0.0146
$\rho = 0.50$							
T = 100	0.0953	0.1033	0.0920	0.0863	0.0842	0.0905	0.0850
T = 200	0.0457	0.0478	0.0434	0.0406	0.0398	0.0399	0.0398
T = 500	0.0194	0.0184	0.0170	0.0159	0.0155	0.0157	0.0156
$\rho = 0.80$							
T = 100	0.1049	0.1134	0.1017	0.0959	0.0940	0.1024	0.0947
T = 200	0.0511	0.0532	0.0484	0.0454	0.0442	0.0447	0.0443
T = 500	0.0215	0.0203	0.0188	0.0174	0.0169	0.0169	0.0169
$\rho = 0.99$							
T = 100	0.7213	0.9599	0.6244	0.6713	0.1505	0.1830	0.1531
T = 200	0.2619	0.3052	0.2317	0.2425	0.0632	0.0867	0.0636
T = 500	0.0327	0.0310	0.0279	0.0264	0.0213	0.0213	0.0213

For any $\alpha < 1$, it can be rewritten as:

$$y_t = \mu_t + \alpha y_{t-1} + u_{t-1} + v_t$$

where $\mu_t = (1 - \alpha) y_{t-1}$ is an intercept following a random walk. Furthermore, its innovations $(\mu_t - \mu_{t-1})$ are contemporaneously independent of the innovations v_t . Therefore, if the estimated equation includes a constant and time-varying coefficients are not ruled out, it is not possible to identify whether the regression has a unit root and stable coefficients, or has a stationary autoregressive component and

a time-varying intercept¹¹. When ρ is near unity, identification is possible, but it will presumably be weak, giving rise to very imprecise estimates of the coefficients and of their degree of stability. Note that the two equivalent (and unidentified) representations above obviously yield the same one-step-ahead forecasts of y_t . Therefore, if our conjecture that this weak identification problem is affecting our results is correct, we should find that the out-of-sample forecasts of y_t produced by the TVC model are not as unsatisfactory as its coefficient estimates. This is exactly what we find and document in the last part of this subsection.

Panel B of Table 1 reports the Monte Carlo estimates of MSE_j^β for the case in which x_t includes three lags of y_t and u_t . In the case of low persistence, the BP-MA estimates are those that achieve the smallest increase in MSE with respect to the OLS estimates (on average below 2 per cent). The performance of the TVC- π estimates is only slightly inferior (around a 3 percent increase in MSE with respect to OLS). All the other estimates (TVC-MA, TVC-MS, TVC-II, and BP) are somewhat less efficient, but their MSEs seldom exceed those of the OLS estimates by more than 30 per cent. As far as the highly persistent case ($\rho = 0.99$) is concerned, we again observe a degradation in the performance of the TVC and (to a lesser extent) of the BP estimates. However, the degradation is less severe than the one observed in the case of fewer regressors. Intuitively, adding more regressors (even if their coefficients are 0) helps to alleviate the identification problem discussed before, because the added regressors have stable coefficients and hence help to pin down the stable representation of the regression.

The loss in forecasting performance is evaluated using a single out-of-sample prediction for each replication. In each replication, $T + 1$ observations are generated, the first T are used to update the priors, the vector of regressors x_{T+1} is used to predict y_{T+1} , and the prediction (denote it by \tilde{y}_{T+1}) is compared to the actual value y_{T+1} . As for coefficient estimates, we consider seven different predictions:

- Model averaging (TVC-MA) predictions, where:

$$\tilde{y}_{T+1} = E [y_{T+1} | D_T, x_{T+1}] = \sum_{i=1}^q p_{T,i} E [y_{T+1} | D_T, x_{T+1}, \theta = \theta_i]$$

- Model selection (TVC-MS) predictions, where:

$$\tilde{y}_{T+1} = E [y_{T+1} | D_T, x_{T+1}, \theta = \theta_{j^*}]$$

- Predictions generated by the regression model with stable coefficients when $\Pi \geq 0.1$ and by model averaging when $\Pi < 0.1$ (denoted by TVC-II):

$$\tilde{y}_{T+1} = \begin{cases} E [y_{T+1} | D_T, x_{T+1}, \theta = 0] & \text{if } \Pi \geq 0.1 \\ E [y_{T+1} | D_T, x_{T+1}] & \text{if } \Pi < 0.1 \end{cases}$$

- Predictions generated by the regression model with stable coefficients when $\pi \geq 0.1$ and by model averaging when $\pi < 0.1$ (denoted by TVC- π):

$$\tilde{y}_{T+1} = \begin{cases} E [y_{T+1} | D_T, x_{T+1}, \theta = 0] & \text{if } \pi \geq 0.1 \\ E [y_{T+1} | D_T, x_{T+1}] & \text{if } \pi < 0.1 \end{cases}$$

- Predictions generated by the regression model with stable coefficients (OLS):

$$\tilde{y}_{T+1} = E [y_{T+1} | D_T, x_{T+1}, \theta = 0]$$

¹¹ This identification problem is discussed in a very similar context by Hatanaka and Yamada (1999) and Perron and Zhu (2005).

- Predictions obtained from Bai and Perron’s sequential procedure (BP); if $\tilde{\beta}_T$ is the BP estimate of β_T (see above), then:

$$\tilde{y}_{T+1} = x_{T+1}\tilde{\beta}_T$$

- Predictions obtained from Pesaran and Timmerman’s (2007) model-averaging procedure (BP-MA); if $\tilde{\beta}_T$ is the BP-MA estimate of β_T (see above), then:

$$\tilde{y}_{T+1} = x_{T+1}\tilde{\beta}_T$$

The Monte Carlo replications are used to estimate the mean squared error of the predictions:

$$MSE_j^y = E \left[(y_{T+1} - \tilde{y}_{T+1})^2 \right]$$

where $j = \text{TVC-MA, TVC-MS, TVC-II, TVC-}\pi, \text{OLS, BP, BP-MA}$ depending on which of the above methods has been used to forecast y_{T+1} .

To improve the accuracy of our Monte Carlo estimates of MSE_j^y , we use the fact that:

$$E \left[(y_{T+1} - \tilde{y}_{T+1})^2 \right] = E \left[v_{T+1}^2 \right] + E \left[\left(\beta_{T+1} - \tilde{\beta}_{T+1} \right)^\top x_{T+1}^\top x_{T+1} \left(\beta_{T+1} - \tilde{\beta}_{T+1} \right) \right]$$

Since $E \left[v_{T+1}^2 \right]$ is known, we use the Monte Carlo simulations to estimate only the second summand on the right hand side of the above equation.

Table 2 reports the Monte Carlo estimates of MSE_j^y . The variation in MSE_j^y across models and design parameters broadly reflects the variation in MSE_j^β we have discussed above. To avoid repetitions, we point out the only significant difference, which concerns the highly persistent design ($\rho = 0.99$), while the TVC and BP estimates give rise to an MSE_j^β that is around two orders of magnitude higher than MSE_{OLS}^β , the part of their MSE_j^y attributable to estimation error ($MSE_j^y - 1$) compares much more favorably to its OLS counterpart, especially in the designs where x_t includes three lags of y_t and u_t . This might be considered evidence of the identification problem mentioned above.

4.2. Performance When the DGP Is a Regression with a Discrete Structural Break

In this subsection we present the results of a set of Monte Carlo simulations aimed at understanding how our TVC model performs when regression coefficients experience a single discrete structural break. As in the previous subsection, we analyze both losses in forecasting performance and losses in estimation precision.

The Monte Carlo design is the same one employed in the previous subsection, except for the fact that the data generating process is now subject to a discrete structural break at an unknown date:

- *Data generating process:* y_t is generated according to:

$$\begin{aligned} y_t &= \rho y_{t-1} + u_{t-1} + v_t \text{ if } t < \tau \\ y_t &= \rho y_{t-1} + (1 + b) u_{t-1} + v_t \text{ if } t \geq \tau \end{aligned}$$

where $y_0 = 0$, $u_t \sim T(0, 1, 5)$ i.i.d., $v_t \sim N(0, 1)$ i.i.d., and u_t and v_t are serially and cross-sectionally independent; τ is the stochastic breakpoint date, extracted from a discrete uniform distribution on the set of sample dates (from 1 to T); $b \sim N(0, 1)$ is the stochastic break in regression coefficients.

Table 2. Prediction accuracy when coefficients are stable—Monte Carlo evidence—MSE of one-step-ahead predictions.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.0384	1.0438	1.0358	1.0319	1.0301	1.0942	1.0332
T = 200	1.0208	1.0214	1.0181	1.0161	1.0155	1.2914	1.0165
T = 500	1.0101	1.0084	1.0073	1.0064	1.0060	1.0121	1.0073
$\rho = 0.50$							
T = 100	1.0393	1.0443	1.0366	1.0325	1.0311	1.0978	1.0351
T = 200	1.0217	1.0229	1.0193	1.0170	1.0159	1.0443	1.0168
T = 500	1.0101	1.0086	1.0075	1.0065	1.0060	1.0144	1.0061
$\rho = 0.80$							
T = 100	1.0453	1.0526	1.0423	1.0371	1.0348	1.1301	1.0420
T = 200	1.0218	1.0230	1.0191	1.0162	1.0155	1.0757	1.0166
T = 500	1.0103	1.0089	1.0078	1.0065	1.0062	1.0229	1.0063
$\rho = 0.99$							
T = 100	1.2256	1.2710	1.2238	1.2206	1.0489	1.2199	1.0484
T = 200	1.1349	1.1548	1.1317	1.1306	1.0224	1.0496	1.0225
T = 500	1.0419	1.0462	1.0392	1.0377	1.0078	1.0198	1.0078
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.0841	1.0914	1.0813	1.0762	1.0746	1.0808	1.0759
T = 200	1.0433	1.0454	1.0412	1.0386	1.0375	1.0379	1.0379
T = 500	1.0183	1.0171	1.0158	1.0146	1.0142	1.0143	1.0143
$\rho = 0.50$							
T = 100	1.0845	1.0923	1.0817	1.0769	1.0752	1.0928	1.0772
T = 200	1.0435	1.0454	1.0412	1.0380	1.0372	1.0375	1.0375
T = 500	1.0179	1.0167	1.0154	1.0144	1.0141	1.0141	1.0141
$\rho = 0.80$							
T = 100	1.0889	1.0967	1.0864	1.0816	1.0802	1.0855	1.0815
T = 200	1.0418	1.0435	1.0397	1.0373	1.0363	1.0369	1.0367
T = 500	1.0177	1.0168	1.0155	1.0143	1.0139	1.0140	1.0140
$\rho = 0.99$							
T = 100	1.1381	1.1705	1.1298	1.1353	1.0943	1.1128	1.0886
T = 200	1.0647	1.0727	1.0609	1.0617	1.0439	1.0504	1.0425
T = 500	1.0196	1.0192	1.0178	1.0171	1.0163	1.0160	1.0160

The estimation precision and the forecasting performance are evaluated by comparing the estimates of the coefficient vector at time T and the predictions of y_{T+1} with their true values.

Panel A of Table 3 reports the Monte Carlo estimates of MSE_j^β for the case in which x_t includes only the first lags of y_t and u_t . As before, we first discuss the cases in which $\rho \neq 0.99$. The OLS estimates, which have the smallest MSEs in the stable case (see previous subsections) are now those with the highest MSEs. Both the frequentist methods (BP and BP-MS) and the TVC methods (all four kinds) achieve a significant reduction in the MSEs with respect to OLS. Although TVC-MA and TVC-MS perform slightly better than TVC-II and TVC- π , there is not a clear ranking between the former two and the two frequentist methods: Their MSEs are on average comparable, but TVC-MA and TVC-MS tend to perform better when the sample size is small ($T = 100$), while BP and BP-MA tend to perform better when the sample size is

large ($T = 200, 500$). This might be explained by the fact that BP and BP-MA require the estimation of a considerable number of parameters when one or more breakdates are found and these parameters are inevitably estimated with low precision when the sample size is small. In the case in which $\rho = 0.99$, results are again substantially different: The MSEs of the TVC estimates (all four kinds) and of the BP estimates become much larger than the MSEs of the OLS estimates (and the BP estimates fare better than the TVC estimates), while the MSEs of the BP-MA estimates remain below those of the OLS estimates. The remarks about potential identification problems made in the previous subsections also apply to these results.

Panel B of Table 3 reports the Monte Carlo estimates of MSE_j^β for the case in which x_t includes three lags of y_t and u_t . The patterns are roughly the same found in Panel A (see previous paragraph), with the relative performance of the TVC methods and the frequentist methods depending on the sample size T . The only difference worth mentioning is that when $\rho = 0.99$, the increase in the MSEs is milder and the TVC-MA estimates are more precise than the BP estimates.

Table 3. Estimation precision when coefficients experience one discrete break—Monte Carlo evidence—MSE of coefficient estimates.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.1772	0.1767	0.1807	0.1943	0.3769	0.1737	0.2252
T = 200	0.1148	0.1145	0.1183	0.1250	0.3434	0.1088	0.1863
T = 500	0.0689	0.0692	0.0711	0.0741	0.3453	0.0452	0.1738
$\rho = 0.50$							
T = 100	0.1790	0.1792	0.1831	0.1951	0.3655	0.2140	0.2248
T = 200	0.1237	0.1228	0.1265	0.1346	0.3545	0.1053	0.1930
T = 500	0.0692	0.0694	0.0713	0.0744	0.3452	0.0533	0.1733
$\rho = 0.80$							
T = 100	0.1887	0.1906	0.1924	0.2038	0.3705	0.3016	0.2226
T = 200	0.1261	0.1255	0.1290	0.1360	0.3463	0.1309	0.1894
T = 500	0.0707	0.0708	0.0729	0.0760	0.3449	0.0803	0.1733
$\rho = 0.99$							
T = 100	5.1868	5.7741	5.1508	5.1618	0.4283	2.5268	0.4261
T = 200	3.3069	3.5054	3.2908	3.2956	0.3626	1.4681	0.2585
T = 500	0.7054	0.7362	0.7032	0.7053	0.3542	0.6431	0.1946
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.3299	0.3368	0.3322	0.3424	0.4389	0.3628	0.3320
T = 200	0.2127	0.2144	0.2145	0.2202	0.3761	0.2186	0.2421
T = 500	0.1357	0.1361	0.1372	0.1400	0.3540	0.1053	0.1974
$\rho = 0.50$							
T = 100	0.3272	0.3347	0.3291	0.3377	0.4368	0.3948	0.3292
T = 200	0.2201	0.2218	0.2215	0.2270	0.3826	0.2194	0.2460
T = 500	0.1401	0.1405	0.1413	0.1435	0.3507	0.1078	0.1956
$\rho = 0.80$							
T = 100	0.3529	0.3627	0.3547	0.3628	0.4585	0.4161	0.3476
T = 200	0.2420	0.2441	0.2436	0.2499	0.3913	0.2501	0.2591
T = 500	0.1475	0.1478	0.1485	0.1513	0.3638	0.1155	0.2040

Table 3. Cont.

Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0.99$							
T = 100	1.3642	1.6216	1.3155	1.3468	0.5319	1.3814	0.5141
T = 200	0.7399	0.8044	0.7279	0.7389	0.4296	0.7678	0.3266
T = 500	0.2532	0.2569	0.2541	0.2565	0.3651	0.3168	0.2147

As far as out-of-sample forecasting performance is concerned (Table 4, Panels A and B), the patterns in the MSE_j^y broadly reflect the patterns in the MSE_j^β . Again, there is an exception to this: When $\rho = 0.99$, high values of MSE_j^β do not translate into high values of MSE_j^y and as a consequence, despite the aforementioned identification problem, the BP and the four TVC forecasts are much more accurate than the OLS forecasts (and in some cases are also more accurate than the BP-MA forecasts).

Table 4. Prediction accuracy when coefficients experience one discrete break—Monte Carlo evidence—MSE of one-step-ahead predictions.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.2975	1.2989	1.3030	1.3243	1.6029	1.3322	1.3784
T = 200	1.1810	1.1794	1.1867	1.1987	1.5315	1.1777	1.2937
T = 500	1.1101	1.1105	1.1138	1.1208	1.6174	1.0930	1.3093
$\rho = 0.50$							
T = 100	1.2645	1.2643	1.2753	1.2991	1.5421	1.2941	1.3343
T = 200	1.2031	1.2039	1.2095	1.2192	1.6492	1.1700	1.3604
T = 500	1.1099	1.1106	1.1141	1.1213	1.6162	1.0908	1.3064
$\rho = 0.80$							
T = 100	1.2814	1.2815	1.2882	1.3202	1.5781	1.3859	1.3552
T = 200	1.1693	1.1684	1.1742	1.1876	1.5674	1.1481	1.2955
T = 500	1.1082	1.1085	1.1126	1.1187	1.6158	1.0867	1.3069
$\rho = 0.99$							
T = 100	1.4014	1.4357	1.4051	1.4245	1.6442	1.3933	1.3764
T = 200	1.2821	1.2951	1.2841	1.2929	1.5419	1.2005	1.3027
T = 500	1.1807	1.1857	1.1846	1.1887	1.6844	1.1552	1.3481
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.4007	1.4022	1.4042	1.4272	1.6117	1.5408	1.4496
T = 200	1.2689	1.2692	1.2738	1.2855	1.5737	1.3308	1.3612
T = 500	1.1770	1.1773	1.1803	1.1877	1.5753	1.1150	1.3172
$\rho = 0.50$							
T = 100	1.4022	1.4059	1.4070	1.4255	1.6490	1.6550	1.4823
T = 200	1.2832	1.2835	1.2872	1.2969	1.6013	1.3515	1.3825
T = 500	1.1576	1.1582	1.1600	1.1634	1.5436	1.1125	1.2884

Table 4. Cont.

Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0.80$							
T = 100	1.4206	1.4245	1.4242	1.4426	1.6930	1.5160	1.4954
T = 200	1.2864	1.2831	1.2891	1.3045	1.6034	1.2747	1.3675
T = 500	1.1733	1.1727	1.1758	1.1820	1.6348	1.1153	1.3417
$\rho = 0.99$							
T = 100	1.4718	1.5054	1.4852	1.5134	1.6644	1.4521	1.4802
T = 200	1.3123	1.3224	1.3183	1.3365	1.6158	1.3670	1.3750
T = 500	1.1916	1.1934	1.1967	1.2029	1.5966	1.1138	1.3327

4.3. Performance When the DGP Is a Regression with Frequently Changing Coefficients

In this subsection we present the results of a set of Monte Carlo simulations aimed at understanding how our TVC model performs when regression coefficients experience frequent changes.

We analyze both losses in forecasting performance and losses in estimation precision, using the same Monte Carlo design employed in the previous two subsections. The only difference is that the data is now generated by a regression whose coefficients change at every time period:

- Data generating process: y_t is generated according to:

$$y_t = \rho y_{t-1} + b_t u_{t-1} + v_t$$

$$b_t = b_{t-1} + w_t$$

where $y_0 = 0, b_0 = 1, u_t \sim T(0, 1, 5)$ i.i.d., $v_t \sim N(0, 1)$ i.i.d., $w_t \sim N(0, V)$ i.i.d., and u_t, v_t and w_t are serially and cross-sectionally independent. To ease comparisons with the previous subsection, V is chosen in such a way that $b_T \sim N(1, 1)$, irrespective of the sample size T :

$$V = \frac{1}{T}$$

Note that, although one coefficient of the regression is frequently changing (b_t), the other coefficient (ρ) is stable, therefore the true DGP does not exactly fit any of the possible DGPs contemplated by the TVC model. We prefer to adopt this specification over a specification in which the TVC model is correctly specified, because the results obtained with the latter specification are trivial (the TVC estimates are the best possible estimates). Furthermore, controlling ρ (keeping it fixed) allows a better understanding of its effects on model performance.

Panel A of Table 5 reports the Monte Carlo estimates of MSE_j^β for the case in which x_t includes only the first lags of y_t and u_t . We first summarize the results obtained when $\rho \neq 0.99$. The lowest MSEs are achieved by the TVC-MA estimates. The TVC-MS estimates are the second best (in some cases MSE_{TVC-MS}^β is almost identical to MSE_{TVC-MA}^β). TVC-II and TVC- π also have a performance comparable to that of TVC-MA (the increase in the MSEs is on average less than 5 per cent). The BP estimates are significantly less precise than the TVC estimates (their MSEs are roughly between 30 per cent and 70 per cent higher than MSE_{TVC-MA}^β). Finally, BP and BP-MA have a comparable performance when $T = 100$, but BP-MA is much less precise when the sample size increases ($T = 200, 500$).

When $\rho = 0.99$, we again observe a sharp increase in the MSEs of the TVC estimates (all four kinds) and of the BP estimates: Their MSEs become several times those of the OLS estimates. BP-MA achieves a significant reduction in MSE over OLS with larger sample sizes ($T = 200, 500$). Thus, with frequently

changing coefficients, BP-MA seems to be the only method capable of simultaneously dealing with coefficient instability and a highly persistent lagged dependent variable.

Panel B of Table 5 reports the Monte Carlo estimates of MSE_j^β for the case in which x_t includes three lags of y_t and u_t . Similar to what we found in the previous subsections, the only noticeable difference with respect to the one-lag case is that when $\rho = 0.99$ the increase in the MSEs is milder.

Table 5. Estimation precision when coefficients change every period—Monte Carlo evidence—MSE of coefficient estimates.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.1768	0.1778	0.1803	0.1939	0.3653	0.2333	0.2331
T = 200	0.1207	0.1210	0.1224	0.1280	0.3535	0.1914	0.2048
T = 500	0.0718	0.0718	0.0722	0.0735	0.3409	0.0925	0.1766
$\rho = 0.50$							
T = 100	0.1856	0.1865	0.1891	0.2031	0.3775	0.2689	0.2413
T = 200	0.1216	0.1219	0.1234	0.1287	0.3531	0.1635	0.2055
T = 500	0.0719	0.0720	0.0724	0.0735	0.3490	0.0982	0.1770
$\rho = 0.80$							
T = 100	0.2003	0.2042	0.2029	0.2146	0.3738	0.3252	0.2403
T = 200	0.1268	0.1275	0.1285	0.1333	0.3353	0.1982	0.1971
T = 500	0.0733	0.0733	0.0737	0.0747	0.3413	0.1108	0.1759
$\rho = 0.99$							
T = 100	5.8872	6.5930	5.8655	5.8674	0.4481	3.8089	0.4898
T = 200	3.2478	3.4124	3.2435	3.2428	0.3766	1.6324	0.2718
T = 500	0.8609	0.8916	0.8608	0.8609	0.3395	0.5805	0.1835
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	0.3279	0.3378	0.3304	0.3407	0.4386	0.3903	0.3544
T = 200	0.2187	0.2221	0.2202	0.2247	0.3810	0.2514	0.2723
T = 500	0.1349	0.1358	0.1354	0.1364	0.3526	0.1559	0.2151
$\rho = 0.50$							
T = 100	0.3369	0.3483	0.3391	0.3478	0.4395	0.3954	0.3553
T = 200	0.2279	0.2313	0.2292	0.2339	0.3834	0.2607	0.2753
T = 500	0.1389	0.1397	0.1393	0.1404	0.3553	0.1608	0.2176
$\rho = 0.80$							
T = 100	0.3608	0.3735	0.3632	0.3728	0.4500	0.4528	0.3665
T = 200	0.2397	0.2442	0.2410	0.2456	0.3901	0.2756	0.2806
T = 500	0.1489	0.1498	0.1493	0.1504	0.3655	0.1686	0.2241
$\rho = 0.99$							
T = 100	1.3679	1.6564	1.3119	1.3457	0.5345	1.2136	0.5107
T = 200	0.7702	0.8373	0.7648	0.7690	0.4231	0.5511	0.3365
T = 500	0.2467	0.2511	0.2472	0.2480	0.3633	0.2323	0.2291

As far as out-of-sample forecasting performance is concerned (Table 6, Panels A and B), the patterns in the MSE_j^y broadly reflect the patterns in the MSE_j^β . Again, the case of $\rho = 0.99$ constitutes an exception.

Despite their high MSE_j^β , the BP and the four TVC forecasts are more accurate than the OLS forecasts (and the TVC-MA and TVC- Π forecasts are also more accurate than the BP-MA forecasts).

Table 6. Prediction accuracy when coefficients change every period—Monte Carlo evidence—MSE of one-step-ahead predictions.

Panel A—One Lag in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.2637	1.2660	1.2696	1.2887	1.5793	1.3656	1.3679
T = 200	1.1825	1.1833	1.1858	1.1942	1.5627	1.2622	1.3261
T = 500	1.1094	1.1096	1.1101	1.1119	1.5546	1.1515	1.2784
$\rho = 0.50$							
T = 100	1.2760	1.2751	1.2830	1.3093	1.5948	1.4155	1.3831
T = 200	1.1867	1.1868	1.1894	1.2000	1.5554	1.2433	1.3298
T = 500	1.1161	1.1159	1.1170	1.1192	1.5943	1.1564	1.3054
$\rho = 0.80$							
T = 100	1.3009	1.3026	1.3070	1.3339	1.6125	1.4257	1.4093
T = 200	1.1894	1.1903	1.1926	1.2047	1.5493	1.2723	1.3286
T = 500	1.1116	1.1115	1.1123	1.1139	1.5399	1.1672	1.2843
$\rho = 0.99$							
T = 100	1.4028	1.4363	1.4067	1.4220	1.6146	1.4992	1.3924
T = 200	1.2848	1.2972	1.2867	1.2917	1.5963	1.2675	1.3379
T = 500	1.1564	1.1596	1.1574	1.1588	1.5581	1.1752	1.2882
Panel B—Three Lags in the Estimated Equation							
	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
$\rho = 0$							
T = 100	1.4312	1.4333	1.4359	1.4636	1.6667	1.5028	1.5340
T = 200	1.2653	1.2653	1.2681	1.2800	1.5577	1.3186	1.3912
T = 500	1.1713	1.1715	1.1723	1.1737	1.5745	1.2156	1.3453
$\rho = 0.50$							
T = 100	1.4214	1.4247	1.4252	1.4476	1.6373	1.4982	1.5109
T = 200	1.2954	1.2961	1.2993	1.3114	1.6280	1.3697	1.4369
T = 500	1.1687	1.1689	1.1695	1.1711	1.5631	1.2145	1.3402
$\rho = 0.80$							
T = 100	1.4303	1.4342	1.4353	1.4614	1.6616	1.5534	1.5281
T = 200	1.2803	1.2816	1.2829	1.2941	1.5949	1.3301	1.4126
T = 500	1.1652	1.1648	1.1659	1.1681	1.5584	1.2240	1.3402
$\rho = 0.99$							
T = 100	1.5138	1.5417	1.5246	1.5664	1.6848	1.5574	1.5380
T = 200	1.3405	1.3533	1.3469	1.3624	1.6228	1.3507	1.4279
T = 500	1.1838	1.1851	1.1853	1.1889	1.5842	1.2445	1.3475

5. Empirical Application: Estimating Common Stocks' Exposures to Risk Factors

In this section we briefly illustrate an empirical application of our TVC model. We use the model to estimate the exposures of S&P 500 constituents to market-wide risk factors. We track the weekly returns of the S&P 500 constituents for 10 years (from January 2000 to December 2009). An uninterrupted time series of returns is available for 432 of the 500 constituents (as of December 2009). The list of constituents and their returns are downloaded from Datastream. The risk factors we consider are the [Fama and French's \(1993,](#)

1995, 1996) risk factors (excess return on the market portfolio, return on the Small Minus Big portfolio, return on the High Minus Low portfolio), downloaded from Kenneth French’s website.

The exposures to the risk factors are the coefficients β_t in the regression

$$y_t = x_t\beta_t + v_t$$

where y_t is the excess return on a stock at time t ,

$$x_t = \left[1 \quad r_{M,t} - r_{f,t} \quad SMB_t \quad HML_t \right]$$

$r_{M,t}$ is the return on the market portfolio at time t , $r_{f,t}$ is the risk-free rate of return, and SMB_t and HML_t are the returns at time t on the SMB and HML portfolios respectively.

The procedures illustrated in the previous section are employed to understand whether the risk exposures β_t are time-varying and whether the TVC model provides good estimates of these risk exposures.

For a vast majority of the stocks included in our sample, we find evidence that β_t is indeed time-varying. $\theta = 0$ is the posterior mode of the mixing parameter for only 11 stocks out of 432. Furthermore, $P < 0.1$ and $\pi < 0.1$ for 92 per cent and 81 per cent of the stocks respectively. On average, P is 0.046 and π is 0.010. Furthermore, the frequentist method provides evidence that most stocks experience instability in their risk exposures. According to the BP sequential estimates, more than 78 per cent of stocks experience at least one break in β_t .

To evaluate the forecasting performance, we use the out-of-sample forecasts of y_t obtained after the first 400th week. The methods used to make predictions are those described in the previous section ($j = \text{TVC-MA, TVC-MS, TVC-II, TVC-}\pi, \text{OLS, BP, BP-MA}$). For each stock i and for each prediction method j , the mean squared error is computed as:

$$\overline{MSE}_{i,j} = \frac{1}{T - T_0} \sum_{t=T_0+1}^T (y_{t,i,j} - \tilde{y}_{t,i,j})^2$$

where T_0 is the number of periods elapsed before the first out-of-sample forecast is produced, $\tilde{y}_{t,i,j}$ denotes the prediction of the excess return of the i -th stock at time t , conditional on x_t , produced by method j , and $y_{t,i,j}$ is the corresponding realization.

In order to compare the performance of the various methods across stocks, we use the performance of OLS forecasts as a benchmark. Thus, the gain from using model j with stock i is defined as:

$$GAIN_{i,j} = 1 - \frac{\overline{MSE}_{i,j}}{\overline{MSE}_{i,OLS}}$$

i.e., $GAIN_{i,j}$ is the average reduction in MSE achieved by using model j instead of OLS. A positive value indicates an improvement in forecasting performance.

Table 7 reports some summary statistics of the sample distribution of $GAIN_{i,j}$ (to each stock i corresponds a different sample point). All the TVC methods achieve a reduction in MSE and, among the TVC methods, TVC-MA achieves the maximum average reduction (approximately 3 per cent). BP performs very poorly (it actually causes a strong increase in MSE), while the average reduction achieved by BP-MA is similar to that of TVC-MA (again, approximately 3 per cent). The four TVC models have similar sample distributions of gains, characterized by a pronounced skew to the right (several small gains and few very large gains); furthermore, all four have a more dispersed distribution than the BP-MA model.

Table 7. Prediction accuracy—Risk exposures—Reduction in the MSE of one-step-ahead predictions (benchmark = OLS).

	TVC-MA	TVC-MS	TVC-P	TVC-p	OLS	BP	BP-MA
Mean	2.94%	2.48%	2.70%	2.15%	0%	−121.72%	3.13%
Standard dev.	11.02%	11.68%	10.82%	10.63%	0%	−557.07%	6.00%
First quartile	−2.14%	−2.67%	−2.14%	−2.23%	0%	−46.94%	−0.14%
Median	1.85%	1.30%	1.17%	0.11%	0%	−5.76%	1.53%
Third quartile	7.83%	7.75%	7.47%	6.84%	0%	0.20%	6.37%

6. Conclusions

We have proposed a Bayesian regression model with time-varying coefficients (TVC) that has low computational requirements because it allows one to derive analytically the posterior distribution of coefficients, as well as the posterior probability that they are stable.

The model is completely automatic in the sense that regressors and regressands are the only input required from the econometrician, so that they do not need to engage in technically demanding specifications of priors and model parametrizations.

We conducted several Monte Carlo experiments to understand the finite-sample properties of the model. We found that the model had satisfactory estimation precision and forecasting performance even when regression coefficients were stable or when coefficient instability was present but the model was mis-specified. When coefficients were unstable, the estimation precision and the forecasting accuracy of the model were significantly better than those of other competing models.

A caveat emerged from our Monte Carlo experiments: When a highly persistent autoregressive component was included among the regressors, then the TVC model tended to have poor estimation precision; in this case, the performance of the TVC model could be improved by increasing the number of exogenous regressors or by increasing the sample size, otherwise, one could resort to model-averaging variants of frequentist breakpoint detection procedures.

To demonstrate a real-world application of our TVC model, we used it to estimate the exposures of S&P 500 stocks to market-wide risk factors. We found that a vast majority of stocks had time-varying exposures and that the TVC model helped to better forecast these exposures.

Author Contributions: All authors contributed equally to the paper.

Funding: This research received no external funding.

Acknowledgments: We thank Giovanni Urga and Fabrizio Venditti for helpful comments, as well as the audiences at the following seminars/conferences: The Bank of Italy, EIEF, Cass Business School, Scottish economic Society. The views expressed here are our own and do not necessarily reflect those of the Bank of Italy. Other usual disclaimers apply.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proofs of Propositions 1 and 2

In this section we derive the formulae presented in Propositions 1 and 2. To facilitate the exposition, we start from simpler information structures and then we tackle the more complex information structure assumed in Propositions 1 and 2 and summarized in Summary 1.

Appendix A.1.1. V and θ Known, β_1 Unknown

We start from the simple case in which V and θ are both known. The assumptions on the priors and the initial information are summarized as follows:

Case A1 (Priors and initial information). *The priors on the unknown parameters are:*

$$(\beta_1 | D_0^{**}) \sim N(\hat{\beta}_{1,0}, VF_{\beta,1,0})$$

and the initial information set is:

$$D_0^{**} = \{\hat{\beta}_{1,0}, F_{\beta,1,0}, V, \theta\}$$

Note that also $W^* = \lambda(\theta) F_{\beta,1,0}$ and $W = V\lambda(\theta) F_{\beta,1,0}$ are known, because θ and V are known. The information sets D_t^{**} satisfy the recursion $D_t^{**} = D_{t-1}^{**} \cup \{y_t, x_t\}$, starting from the set D_0^{**} . Given the above assumptions, as new information becomes available, the posterior distribution of the parameters of the regression can be calculated using the following results:

Proposition A1 (Forward updating). *Let priors and initial information be as in Case A1. Then:*

$$\begin{aligned} (\beta_t | D_{t-1}^{**}) &\sim N(\hat{\beta}_{t,t-1}, VF_{\beta,t,t-1}) \\ (y_t | D_{t-1}^{**}, x_t) &\sim N(\hat{y}_{t,t-1}, VF_{y,t,t-1}) \\ (\beta_t | D_t^{**}) &\sim N(\hat{\beta}_{t,t}, VF_{\beta,t,t}) \end{aligned}$$

where the means and variances of the above distributions are calculated recursively as follows:

$$\begin{aligned} \hat{\beta}_{t,t-1} &= \hat{\beta}_{t-1,t-1} & F_{\beta,t,t-1} &= F_{\beta,t-1,t-1} + W^* \\ \hat{y}_{t,t-1} &= x_t \hat{\beta}_{t,t-1} & F_{y,t,t-1} &= 1 + x_t F_{\beta,t,t-1} x_t^\top \\ e_t &= y_t - \hat{y}_{t,t-1} & P_t &= F_{\beta,t,t-1} x_t^\top / F_{y,t,t-1} \\ \hat{\beta}_{t,t} &= \hat{\beta}_{t,t-1} + P_t e_t & F_{\beta,t,t} &= F_{\beta,t,t-1} - P_t P_t^\top F_{y,t,t-1} \end{aligned}$$

starting from the initial values $\hat{\beta}_{1,0}$ and $F_{\beta,1,0}$.

Proof. Note that, given the above assumptions, the system:

$$\begin{cases} y_t = x_t \beta_t + v_t \\ \beta_t = \beta_{t-1} + w_t \end{cases}$$

is a Gaussian linear state-space system, where $y_t = x_t \beta_t + v_t$ is the observation equation and $\beta_t = \beta_{t-1} + w_t$ is the transition equation. Hence, the posterior distribution of the states can be updated using the Kalman filter. The recursive equations in Proposition A1 are just the usual updating equations of the Kalman filter (e.g., Hamilton 1994). □

The smoothing equations are provided by the following proposition:

Proposition A2 (Backward updating). *Let priors and initial information be as in Case A1. Then:*

$$(\beta_{T-\tau} | D_T^{**}) \sim N(\hat{\beta}_{T-\tau,T}, VF_{\beta,T-\tau,T})$$

where the means and the variances of the above distributions are calculated recursively (backwards) as follows:

$$\begin{aligned} Q_{T-\tau} &= F_{\beta,T-\tau,T-\tau} (F_{\beta,T-\tau+1,T-\tau})^{-1} \\ \widehat{\beta}_{T-\tau,T} &= \widehat{\beta}_{T-\tau,T-\tau} + Q_{T-\tau} (\widehat{\beta}_{T-\tau+1,T} - \widehat{\beta}_{T-\tau+1,T-\tau}) \\ F_{\beta,T-\tau,T} &= F_{T-\tau,T-\tau} + Q_{T-\tau} (F_{T-\tau+1,T} - F_{T-\tau+1,T-\tau}) Q_{T-\tau}^\top \end{aligned}$$

and the backward recursions start from the terminal values of the forward recursions calculated in Proposition 5.

Proof. These are the usual backward Kalman recursions (e.g., Hamilton 1994). \square

Appendix A.1.2. θ Known, β_1 and V Unknown

In this subsection we relax the assumption that V (the variance of v_t) is known and we impose a Gamma prior on the reciprocal of V . The assumptions on the priors and the initial information are summarized as follows:

Case A2 (Priors and initial information). *The priors on the unknown parameters are:*

$$\begin{aligned} (\beta_1 | D_0^*) &\sim N(\widehat{\beta}_{1,0}, VF_{\beta,1,0}) \\ (1/V | D_0^*) &\sim G(\widehat{V}_0, n_0) \end{aligned}$$

and the initial information set is:

$$D_0^* = \{\widehat{\beta}_{1,0}, F_{\beta,1,0}, \widehat{V}_0, n_0, \theta\}$$

Note that also $W^* = \lambda(\theta) F_{\beta,1,0}$ is known, because θ is known. The information sets D_t^* satisfy the recursion $D_t^* = D_{t-1}^* \cup \{y_t, x_t\}$, starting from the set D_0^* . Given the above assumptions, the posterior distributions of the parameters of the regression can be calculated as follows:

Proposition A3 (Forward updating). *Let priors and initial information be as in Case A2. Then:*

$$\begin{aligned} (\beta_t | D_{t-1}^*) &\sim T(\widehat{\beta}_{t,t-1}, \widehat{V}_{t-1} F_{\beta,t,t-1}, n_{t-1}) \\ (y_t | D_{t-1}^*, x_t) &\sim T(\widehat{y}_{t,t-1}, \widehat{V}_{t-1} F_{y,t,t-1}, n_{t-1}) \\ (\beta_t | D_t^*) &\sim T(\widehat{\beta}_{t,t}, \widehat{V}_t F_{\beta,t,t}, n_t) \\ (1/V | D_t^*) &\sim G(\widehat{V}_t, n_t) \end{aligned}$$

where the parameters of the above distributions are calculated recursively as in Proposition A1 and as follows:

$$\begin{aligned} n_t &= n_{t-1} + 1 \\ \widehat{V}_t &= \frac{1}{n_t} \left(n_{t-1} \widehat{V}_{t-1} + \frac{e_t^2}{F_{y,t,t-1}} \right) \end{aligned}$$

starting from the initial values $\widehat{\beta}_{1,0}, F_{\beta,1,0}, \widehat{V}_0$ and n_0 .

Proof. The proof is by induction. At time $t = 1$, $p(\beta_1 | D_0^*, V)$ and $p(1/V | D_0^*)$ are the conjugate normal/inverse gamma priors of a standard Bayesian regression model with constant coefficients (e.g., Hamilton 1994). Therefore, the usual results on the updating of these conjugate priors hold:

$$(\beta_1 | D_1^*, V) \sim N(\hat{\beta}_{1,1}, VF_{\beta,1,1}) \tag{A1}$$

$$(1/V | D_1^*) \sim G(\hat{V}_1, n_1) \tag{A2}$$

Since $\beta_2 = \beta_1 + w_2$ and

$$(w_2 | D_1^*, V) \sim N(0, VW^*)$$

then, by the additivity of normal distributions:

$$(\beta_2 | D_1^*, V) \sim N(\hat{\beta}_{1,1}, VF_{\beta,1,1} + VW^*) = N(\hat{\beta}_{2,1}, VF_{\beta,2,1})$$

Therefore, at time $t = 2$, $p(\beta_2 | D_1^*, V)$ and $p(1/V | D_1^*)$ are again the conjugate normal/inverse gamma priors of a standard Bayesian regression model with constant coefficients. Proceeding in the same way as for $t = 1$, one obtains the desired result for $t = 2$ and, inductively, for all the other periods. □

Posterior distributions of the coefficients that take into account all information received up to time T are calculated as follows:

Proposition A4 (Backward updating). *Let priors and initial information be as in Case A2. Then:*

$$(\beta_{T-\tau} | D_T^*) \sim T(\hat{\beta}_{T-\tau,T}, \hat{V}_T F_{\beta,T-\tau,T}, n_T)$$

where \hat{V}_T and n_T are calculated as in Proposition A3 and the other parameters of the above distributions are calculated recursively (backwards) as in Proposition A2.

Proof. From Proposition A2, we know that:

$$(\beta_{T-\tau} | D_T^*, V) = (\beta_{T-\tau} | D_T^{**}) \sim N(\hat{\beta}_{T-\tau,T}, VF_{\beta,T-\tau,T})$$

Furthermore, $(1/V | D_T^*) \sim G(\hat{V}_T, n_T)$. By the conjugacy of $(\beta_{T-\tau} | D_T^*, V)$ and $(1/V | D_T^*)$, it follows that:

$$(\beta_{T-\tau} | D_T^*) \sim T(\hat{\beta}_{T-\tau,T}, \hat{V}_T F_{\beta,T-\tau,T}, n_T)$$

□

Appendix A.1.3. θ , β_1 and V Unknown

In this subsection we relax the assumption that θ is known, using the same priors and initial information of the propositions in the main text of the article (Propositions 1 and 2):

Case A3. The priors on the unknown parameters are:

$$\begin{aligned} (\beta_1 | D_0, V, \theta) &\sim N(\widehat{\beta}_{1,0}, VF_{\beta,1,0}) \\ (1/V | D_0, \theta) &\sim G(\widehat{V}_0, n_0) \\ p(\theta_i | D_0) &= p_{0,i}, i = 1, \dots, q \end{aligned}$$

and the initial information set is:

$$D_0 = \{ \widehat{\beta}_{1,0}, F_{\beta,1,0}, \widehat{V}_0, n_0, p_{0,1}, \dots, p_{0,q} \}$$

The information sets D_t satisfy the recursion $D_t = D_{t-1} \cup \{y_t, x_t\}$, starting from the set D_0 . Note that the assumptions introduced in Cases A1 and A2 in the previous subsections had the sole purpose of introducing the more complex Case A3. Given the above assumptions, the posterior distributions of the parameters of the regression can be calculated as follows:

Proposition A5. Let priors and initial information be as in Case A3. Let $p_{t,i} = p(\theta = \theta_i | D_t)$. Then:

$$\begin{aligned} p(\beta_t | D_{t-1}) &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_{t-1}) p_{t-1,i} \\ p(y_t | D_{t-1}, x_t) &= \sum_{i=1}^q p(y_t | \theta = \theta_i, D_{t-1}, x_t) p_{t-1,i} \\ p(1/V | D_{t-1}) &= \sum_{i=1}^q p(1/V | \theta = \theta_i, D_{t-1}) p_{t-1,i} \\ p(\beta_t | D_t) &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_t) p_{t,i} \end{aligned}$$

The mixing probabilities are obtained recursively as:

$$p_{t,i} = \frac{p_{t-1,i} p(y_t | \theta = \theta_i, D_{t-1}, x_t)}{\sum_{j=1}^q p_{t-1,j} p(y_t | \theta = \theta_j, D_{t-1}, x_t)}$$

starting from the prior probabilities $p_{0,1}, \dots, p_{0,q}$. The conditional densities:

$$\begin{aligned} p(\beta_t | D_{t-1}, \theta = \theta_i) &= p(\beta_t | D_{t-1}^*) \\ p(y_t | D_{t-1}, x_t, \theta = \theta_i) &= p(y_t | D_{t-1}^*, x_t) \\ p(1/V | D_{t-1}, \theta = \theta_i) &= p(1/V | D_{t-1}^*) \\ p(\beta_t | D_t, \theta = \theta_i) &= p(\beta_t | D_t^*) \end{aligned}$$

are calculated for each θ_i as in Propositions A1 and A3.

Proof. Conditioning on $\theta = \theta_i$, the distributions of the parameters β_t and V and of the observations y_t are obtained from Propositions A1 and A3 (it suffices to note that $D_t \cup \theta = D_t^*$). Not conditioning on $\theta = \theta_i$,

the distributions of the parameters β_t and V and of the observations y_t are obtained marginalizing their joint distribution with θ . For example:

$$\begin{aligned} p(\beta_t | D_{t-1}) &= \sum_{i=1}^q p(\beta_t, \theta_i | D_{t-1}) \\ &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_{t-1}) p(\theta = \theta_i | D_{t-1}) \\ &= \sum_{i=1}^q p(\beta_t | \theta = \theta_i, D_{t-1}) p_{t-1,i} \end{aligned}$$

The mixing probabilities are obtained using Bayes' rule:

$$\begin{aligned} p_{t,i} &= p(\theta = \theta_i | D_t) \\ &= p(\theta = \theta_i | y_t, D_{t-1}, x_t) \\ &= \frac{p(y_t | \theta = \theta_i, D_{t-1}, x_t) p(\theta = \theta_i | D_{t-1}, x_t)}{p(y_t | D_{t-1}, x_t)} \\ &= \frac{p(y_t | \theta = \theta_i, D_{t-1}, x_t) p(\theta = \theta_i | D_{t-1})}{\sum_{j=1}^q p(y_t, \theta_j | D_{t-1}, x_t)} \\ &= \frac{p(y_t | \theta = \theta_i, D_{t-1}, x_t) p(\theta = \theta_i | D_{t-1})}{\sum_{j=1}^q p(y_t | \theta = \theta_j, D_{t-1}, x_t) p(\theta = \theta_j | D_{t-1}, x_t)} \\ &= \frac{p(y_t | \theta = \theta_i, D_{t-1}, x_t) p_{t-1,i}}{\sum_{j=1}^q p(y_t | \theta = \theta_j, D_{t-1}, x_t) p_{t-1,j}} \end{aligned}$$

□

Proposition 1 in the main text is obtained by combining Propositions A1, A3, and A5 above. Proposition 2 results from Propositions A2, A4, and A5 above.

References

- Andrews, Donald W. K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61: 821–56. [\[CrossRef\]](#)
- Andrews, Donald W. K., Inpyo Lee, and Werner Ploberger. 1996. Optimal changepoint tests for normal linear regression model. *Journal of Econometrics* 70: 9–38. [\[CrossRef\]](#)
- Bai, Jushan, and Pierre Perron. 1998. Estimating and testing linear models with multiple structural breaks. *Econometrica* 66: 47–78. [\[CrossRef\]](#)
- Bai, Jushan, and Pierre Perron. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18: 1–22. [\[CrossRef\]](#)
- Bai, Jushan, and Pierre Perron. 2006. Multiple structural change models: A simulation analysis. In *Econometric Theory and Practice: Frontier of Analysis and Applied Research (Essays in Honor of Peter Phillips)*. Edited by D. Corbae, S. Durlauf and B. E. Hansen. Cambridge: Cambridge University Press.
- Bitto, Angela, and Sylvia Frühwirth-Schnatter. 2019. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210: 75–97. [\[CrossRef\]](#)
- Brown, Robert L., James Durbin, and James M. Evans. 1975. Techniques for testing the constancy of regression coefficients over time. *Journal of the Royal Statistical Society B* 37: 149–92.
- Carter, Chris K., and Robert Kohn. 1994. On Gibbs sampling for state space models. *Biometrika* 81: 541–53. [\[CrossRef\]](#)

- Casarin, Roberto, German Molina, and Enrique ter Horst. 2019. A Bayesian time varying approach to risk neutral density estimation. *Journal of the Royal Statistical Society: Series A* 182: 165–95. [[CrossRef](#)]
- Chan, Joshua C. C. 2017. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* 35: 17–28.
- Chib, Siddhartha, and Edward Greenberg. 1995. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics* 68: 339–60. [[CrossRef](#)]
- Chow, Gregory C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28: 591–605. [[CrossRef](#)]
- Cogley, Timothy, Giorgio E. Primiceri, and Thomas J. Sargent. 2010. Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics* 2: 43–69. [[CrossRef](#)]
- Cogley, Timothy, and Thomas J. Sargent. 2001. Evolving Post World War II U.S. Inflation Dynamics. *NBER Macroeconomics Annual* 16: 331–73.
- Datta, Gauri Sankar. 1996. On priors providing frequentist validity of Bayesian inference for multiple parametric functions. *Biometrika* 83: 287–98. [[CrossRef](#)]
- Datta, Gauri Sankar, and Jayanta Kumar Ghosh. 1995. On priors providing frequentist validity for Bayesian inference. *Biometrika* 82: 37–45. [[CrossRef](#)]
- D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28: 82–101. [[CrossRef](#)]
- Doan, Thomas, Robert Litterman, and Christopher Sims. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3: 1–100. [[CrossRef](#)]
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on bonds and stocks. *Journal of Financial Economics* 33: 3–56. [[CrossRef](#)]
- Fama, Eugene F., and Kenneth R. French. 1995. Size and book-to-market factors in earnings and returns. *Journal of Finance* 50: 131–55. [[CrossRef](#)]
- Fama, Eugene F., and Kenneth R. French. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51: 55–84. [[CrossRef](#)]
- Fernandez, Carmen, Eduardo Ley, and Mark F. J. Steel. 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100: 381–427. [[CrossRef](#)]
- George, Edward I., and Robert E. McCulloch. 1997. Approaches for Bayesian variable selection. *Statistica Sinica* 7: 339–74.
- Ghosh, J. K., and Rahul Mukerjee. 1993. Frequentist validity of highest posterior density regions in the multiparameter case. *Annals of the Institute of Statistical Mathematics* 45: 293–302. [[CrossRef](#)]
- Guerre, Emmanuel, and Pascal Lavergne. 2005. Data-driven rate-optimal specification testing in regression models. *The Annals of Statistics* 33: 840–70. [[CrossRef](#)]
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, Bruce E. 2000. Testing for structural change in conditional models. *Journal of Econometrics* 97: 93–115. [[CrossRef](#)]
- Hatanaka, Michio, and Kazuo Yamada. 1999. A unit root test in the presence of structural changes in I(1) and I(0) models. In *Cointegration, Causality and Forecasting: a Festschrift in Honour of Clive W. J. Granger*. Edited by R. F. Engle and H. White. Oxford: Oxford University Press.
- Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14: 382–401.
- Horowitz, Joel L., and Vladimir G. Spokoiny. 2001. An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* 69: 599–631. [[CrossRef](#)]
- Jeffreys, Harold. 1961. *The Theory of Probability*, 3rd ed. Oxford: Oxford University Press.
- Kass, Robert E., and Larry Wasserman. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* 90: 928–34. [[CrossRef](#)]
- Kastner, Gregor. 2019. Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics* 210: 98–115. [[CrossRef](#)]

- Lepski, Oleg V., Enno Mammen, and Vladimir G. Spokoiny. 1997. Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Annals of Statistics* 25: 929–47.
- Liang, Feng, Rui Paulo, German Molina, Merlise A. Clyde, and Jim O. Berger. 2008. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association* 103: 410–23. [[CrossRef](#)]
- Mukerjee, Rahul, and Dipak K. Dey. 1993. Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* 80: 499–505. [[CrossRef](#)]
- Mumtaz, Haroon, and Konstantinos Theodoridis. 2018. The changing transmission of uncertainty shocks in the US. *Journal of Business & Economic Statistics* 36: 239–52.
- Nyblom, Jukka. 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84: 348–68. [[CrossRef](#)]
- Pacifico, Antonio. 2019. Structural Panel Bayesian VAR Model to Deal with Model Misspecification and Unobserved Heterogeneity Problems. *Econometrics* 7: 8. [[CrossRef](#)]
- Perron, Pierre, and Xiaokang Zhu. 2005. Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129: 65–119. [[CrossRef](#)]
- Pesaran, M. Hashem, and Allan Timmermann. 2002. Market timing and return prediction under model instability. *Journal of Empirical Finance* 9: 495–510. [[CrossRef](#)]
- Pesaran, M. Hashem, and Allan Timmermann. 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics* 137: 134–61. [[CrossRef](#)]
- Petrova, Katerina. 2019. A quasi-Bayesian local likelihood approach to time varying parameter VAR models. *Journal of Econometrics*. doi:10.1016/j.jeconom.2019.04.031. [[CrossRef](#)]
- Raftery, Adrian E., Miroslav Kárný, and Pavel Ettler. 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics* 52: 52–66. [[CrossRef](#)]
- Robert, Christian. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Berlin: Springer.
- Shively, Thomas S., Robert Kohn, and Sally Wood. 1999. Variable selection and function estimation in additive nonparametric regression using a data-based prior (with discussion). *Journal of the American Statistical Association* 94: 777–806. [[CrossRef](#)]
- Smith, Michael, and Robert Kohn. 1996. Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* 75: 317–43. [[CrossRef](#)]
- Spokoiny, Vladimir. 2001. Data-driven testing the fit of linear models. *Mathematical Methods of Statistics* 10: 465–97.
- Stock, James H., and Mark W. Watson. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* 14: 11–30.
- West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer.
- Zellner, Arnold. 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. Edited by P. K. Goel and A. Zellner. Amsterdam: North-Holland and Elsevier, pp. 233–43.

