Bayesian Analysis of Coefficient Instability in Dynamic Regressions

Emanuela Ciapanna and Marco Taboga*

Bank of Italy, Research Department

Abstract

This paper proposes a Bayesian regression model with time-varying coefficients (TVC) that allows to jointly estimate the degree of instability and the time-path of regression coefficients. Thanks to its computational tractability, the model proves suitable to perform the first (to our knowledge) Monte Carlo study of the finite-sample properties of a TVC model. Under several specifications of the data generating process, the estimation precision and the forecasting accuracy of the proposed TVC model compare favourably to those of other methods commonly employed to deal with parameter instability. Furthermore, the TVC model leads to small losses of efficiency under the null of stability and it is robust to mis-specification, providing a satisfactory performance also when regression coefficients experience discrete structural breaks.

JEL codes: C11, C32, C50.

Keywords: time-varying regression, coefficient instability.

*Any views expressed in this article are the authors' and do not necessarily represent those of the Bank of Italy. The bulk of this research project was done while Marco Taboga was visiting the Centre for Econometric Analysis at Cass Business School. We wish to thank Giovanni Urga and seminar participants at Cass Business School, as well as participants at the Scottish Economic Society 2011 Conference for helpful comments. Corresponding author: Marco Taboga, Via Nazionale 91, 00184 - Roma, Italy. e-mail: marco.taboga@bancaditalia.it .

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).

There are two main approaches to address instability in regression coefficients:

1. formulating and estimating regression models under the hypothesis of constant coefficients, testing for the presence of structural breaks (e.g.: Chow - 1960, Brown, Durbin and Evans - 1975, Nyblom - 1989) and identifying the break-points (e.g.: Andrews, Lee, and Ploberger - 1996, Bai and Perron - 1998);

2. formulating regression models with time-varying coefficients (TVC) and estimating the path of their variation (e.g.: Doan, Litterman and Sims - 1984, Stock and Watson - 1996, Cogley and Sargent - 2001).

Approach (1) allows to search for time spans over which the hypothesis of constant coefficients is not rejected by the data. However, it can happen that regression coefficients change so frequently that the hypothesis of constant coefficients does not fit any time span (or only time spans that are too short to be of any interest to the econometrician). In these cases, approach (2) can be utilized, as it is suitable to deal also with frequently changing coefficients. On the other side of the coin, approach (2) often relies on dynamic specifications that are (at least in theory) not suitable to detect infrequent and abrupt changes in regression coefficients.

In the absence of strong priors about the ways in which relationships between variables change, the two approaches can arguably be considered complementary and it seems reasonable to use them in conjunction. However, approach (1) is apparently much more frequently utilized than approach (2) in empirical work (e.g.: Kapetanios - 2008).

One possible reason why TVC models are less popular is that tests for structural breaks are often quite easy to implement, while specifying and estimating TVC models is usually a difficult task that relies on complex and computationally intensive numerical techniques and requires careful specification of the dynamics of the coefficients. Even if the development of Markov chain Monte Carlo (MCMC) methods has somewhat facilitated the estimation of TVC models (e.g.: Carter and Kohn - 1994 and Chib and Greenberg - 1995), the technical skills and the computing time required by these techniques are still far superior to those required to estimate regressions with constant coefficients[1].

In this paper, we propose a Bayesian TVC model that aims to fill this gap. The model has low computational requirements and allows to compute analytically the posterior probability that the regression is stable, the estimates 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 plain-vanilla OLS regressions with constant coefficients.

Another possible reason why TVC models are less popular than OLS-based alternatives is that the properties of the former are thus far largely unknown, while the latter have been extensively studied both theoretically (e.g.: Bai and Perron - 1998) and by means of Monte Carlo simulations (e.g.: Hansen - 2000 and Bai and Perron - 2006). Thanks to the computational tractability of our TVC model, we are able to perform the first (to our knowledge) Monte Carlo study of the finite sample properties of a TVC model.

The main goal of our Monte Carlo study 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 my data has indeed been generated by a regression with constant coefficients; how much do I lose, in terms of estimation precision and forecasting accuracy, when I 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 they are comparable to the losses from using frequentist breakpoint detection procedures, such as Bai and Perron's (1998 and 2003) sequential procedure and its model-averaging variant (Pesaran and Timmermann - 2007). Under most simulation scenarios, the mean squared estimation error increases by about 5 per cent when one of the proposed TVC estimators is used in place of OLS to estimate the coefficients of a stable regression.

Another concern is robustness to mis-specification. *Suppose my data has been*

[1] In recent years, several empirical papers have successfully applied MCMC methods to the estimation of regression models with time varying parameters (e.g. Sargent, Williams and Zha - 2006, Canova and Ciccarelli - 2009, Canova and Gambetti - 2009, Koop, Leon-Gonzalez and Strachan - 2009, Cogley, Primiceri and Sargent - 2010). See also the survey by Koop and Korobilis (2009).

generated by a regression with few discrete structural breaks; how much do I loose from using the TVC model instead of standard frequentist procedures for breakpoint detection? Our Monte Carlo evidence indicates that also in this case the estimation precision and the forecasting accuracy of the TVC model are comparable to those of standard frequentist procedures.

Finally, a third concern is efficiency under the null of instability. *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 find that it generally does and that in some cases this gain in efficiency can be quite large (TVC can reduce the mean squared estimation error by up to 60 per cent with respect to the best performing OLS-based method).

All in all, the TVC model seems to be a valid complement to frequentist procedures for breakpoint detection, as the performances of the two approaches are, in general, comparable, but the TVC model fares better in the presence of frequently changing coefficients. There is, however, an important exception to this general result: when the regression includes a lag of the dependent variable and the autoregressive coefficient is near unity. In this case, the performance of the TVC model degrades steeply, and so, but to a lesser extent, does the performance of frequentist methods for breakpoint detection. We argue that this phenomenon is due to an identification problem (already pointed out in similar contexts by Hatanaka and Yamada - 1999 and Perron and Zhu - 2005) which can be alleviated by adding more regressors or increasing the sample size.

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 consider a regression commonly employed to estimate how stock returns are related to market-wide risk factors. We find that the coefficients of this regression are unstable with high probability for a vast majority of the stocks included in the S&P 500 index. We also find 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 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 analyze 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 find that such a decision rule performs well across different scenarios, leading to the smallest losses under the null of stability and still being able to produce satisfactory results when coefficients are indeed unstable.

Some features of our model are borrowed from existing TVC models (in particular Doan, Littermann and Sims - 1984, Stock and Watson - 1996, Cogley and Sargent - 2001), whereas other features 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 re-scalings of the regressors[2]: this property is essential to obtain a completely automatic specification of priors. Another original feature 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 coefficient instability, while retaining a considerable parsimony. Our modelling choices have two main practical consequences: 1) the priors are specified in a completely automatic way so that regressors and regressands are the only input required from the final user[3]; 2) the computational burden of the model is minimized, because analytical estimators are available both for the regression coefficients and for their degree of instability. To our knowledge, none of the existing models has these two characteristics, that allow to use the model in large-scale applications such as Monte Carlo simulations.

The paper is organized as follows: Section 1 presents 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

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

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

```
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.

empirical application; Section 6 concludes. Proofs and other technical details are relegated to the Appendix.

1 The Bayesian model

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

$$y_t = x_t \beta_t + v_t \tag{1}$$

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:

$$\beta_t = \beta_{t-1} + w_t \tag{2}$$

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

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:

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

If a $k \times 1$ random vector z has a multivariate Student's t distribution with mean m, scale matrix S and n degrees of freedom, we write:

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

and its density is parametrized as follows:

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

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

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

and its density is parametrized as follows:

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

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

$$X = \begin{bmatrix} x_1^{\top} & \ldots & x_T^{\top} \end{bmatrix}^{\top}$$

1.2 Structure of prior information and updating

In this subsection we state the main assumptions on the structure of prior information and we 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[4], i.e.:

- β_1 has a multivariate normal distribution conditional on V, with known mean $\widehat{\beta}_{1,0}$ and covariance equal to $V \cdot F_{\beta,1,0}$ where $F_{\beta,1,0}$ is a known matrix;

[4]This prior is frequently utilized in Bayesian regressions with constant coefficients (e.g. Hamilton - 1994)

- the reciprocal of V has a Gamma distribution, with known parameters $\widehat{V}_0$ and n_0.

The second set of assumptions regards W, which is assumed to be proportional to the prior variance of β_1:

$$W = \lambda V F_{\beta,1,0} \tag{3}$$

where λ is a coefficient of proportionality[5].

When $\lambda = 0$, the covariance matrix of w_t is zero and regression coefficients are stable. On the contrary, when $\lambda > 0$, w_t has non-zero covariance matrix and the regression coefficients are unstable (i.e. they change through time). The higher λ is, the greater the variance of w_t is and the more unstable regression coefficients are.

The constant of proportionality λ is parametrized as:

$$\lambda = \lambda(\theta) \tag{4}$$

where $\lambda(\theta)$ is a strictly increasing function and θ is a random variable with finite support R_θ:

$$R_\theta = \{\theta_1, \ldots, \theta_q\} \subset [0, 1]$$

The prior probabilities of the q possible values of θ are denoted by $p_{0,1}$, ..., $p_{0,q}$. The discussion of how θ_1, ..., θ_q and $p_{0,1}$, ..., $p_{0,q}$ are chosen is postponed to the next section.

It is also assumed that $\theta_1 = 0$ and $\lambda(\theta_1) = 0$. Therefore, regression coefficients are stable when $\theta = \theta_1 = 0$ and unstable when $\theta \neq \theta_1$ (the closer θ is to 1, the more unstable regression coefficients are).

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

[5] The assumption that $W^* \propto F_{\beta,1,0}$ is made also by Doan, Littermann and Sims (1984) in their seminal paper on TVC models. 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.

Assumption 1 *The priors on the unknown parameters are:*

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

and the initial information set is:

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

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

Proposition 2 *Let priors and initial information be as in Assumption 1. Let $p_{t,i} = p\,(\theta = \theta_i \,|D_t\,)$. Then:*

$$\begin{aligned} p\,(\beta_t \,|D_{t-1}\,) &= \sum_{i=1}^{q} p\,(\beta_t \,|\theta = \theta_i, D_{t-1}\,)\, p_{t-1,i} \qquad (5) \\ p\,(y_t \,|D_{t-1}, x_t\,) &= \sum_{i=1}^{q} p\,(y_t \,|\theta = \theta_i, D_{t-1}, x_t\,)\, p_{t-1,i} \\ p\,(1/V \,|D_{t-1}\,) &= \sum_{i=1}^{q} p\,(1/V \,|\theta = \theta_i, D_{t-1}\,)\, p_{t-1,i} \\ p\,(\beta_t \,|D_t\,) &= \sum_{i=1}^{q} p\,(\beta_t \,|\theta = \theta_i, D_t\,)\, p_{t,i} \end{aligned}$$

where

$$\begin{aligned} &(\beta_t \,|\theta = \theta_i, D_{t-1}\,) \sim T\left(\widehat{\beta}_{t,t-1,i}, \widehat{V}_{t-1,i}F_{\beta,t,t-1,i}, n_{t-1,i}\right) \qquad (6) \\ &(y_t \,|\theta = \theta_i, D_{t-1}, x_t\,) \sim T\left(\widehat{y}_{t,t-1,i}, \widehat{V}_{t-1,i}F_{y,t,t-1,i}, n_{t-1,i}\right) \\ &(1/V \,|\theta = \theta_i, D_{t-1}\,) \sim G\left(\widehat{V}_{t-1,i}, n_{t-1,i}\right) \\ &(\beta_t \,|\theta = \theta_i, D_t\,) \sim T\left(\widehat{\beta}_{t,t,i}, \widehat{V}_{t,i}F_{\beta,t,t,i}, n_{t,i}\right) \end{aligned}$$

and the parameters of the above distributions are obtained recursively as:

$$
\begin{array}{ll}
\widehat{\beta}_{t,t-1,i} = \widehat{\beta}_{t-1,t-1,i} & F_{\beta,t,t-1,i} = F_{\beta,t-1,t-1,i} + \lambda\left(\theta_i\right) F_{\beta,1,0} \\
\widehat{y}_{t,t-1,i} = x_t \widehat{\beta}_{t,t-1,i} & F_{y,t,t-1,i} = 1 + x_t F_{\beta,t,t-1,i} x_t^\top \\
e_{t,i} = y_t - \widehat{y}_{t,t-1,i} & P_{t,i} = F_{\beta,t,t-1,i} x_t^\top / F_{y,t,t-1,i} \\
\widehat{\beta}_{t,t,i} = \widehat{\beta}_{t,t-1,i} + P_{t,i} e_{t,i} & F_{\beta,t,t,i} = F_{\beta,t,t-1,i} - P_{t,i} P_{t,i}^\top F_{y,t,t-1,i} \\
n_{t,i} = n_{t-1,i} + 1 & \widehat{V}_{t,i} = \frac{1}{n_{t,i}}\left(n_{t-1,i}\widehat{V}_{t-1,i} + \frac{e_{t,i}^2}{F_{y,t,t-1,i}}\right)
\end{array}
\tag{7}
$$

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

$$
p_{t,i} = \frac{p_{t-1,i}\, p\left(y_t \left|\theta = \theta_i, D_{t-1}, x_t\right.\right)}{\sum_{j=1}^{q} p_{t-1,j}\, p\left(y_t \left|\theta = \theta_j, D_{t-1}, x_t\right.\right)}
\tag{8}
$$

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

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.

In the above proposition, the priors on the regression coefficients β_t in a generic time period t are updated using only information received up to that same time t. However, after observing the whole sample (up to time T), one might want to revise her priors on the regression coefficients β_t 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 3 *Let priors and initial information be as in Assumption 1. Then, for* $0 \leq \tau \leq T-1$*:*

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

where

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

The mixing probabilities $p_{T,i}$ and the parameters $\widehat{V}_{T,i}$, $n_{T,i}$ are obtained from the recursions in Proposition 2 while the parameters $\widehat{\beta}_{T-\tau,T,i}$ and $F_{\beta,T-\tau,T,i}$ are obtained from the following backward recursions:

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

starting from $\tau = 1$ and taking as 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 2.

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

Lemma 4 *For $1 \le t \le T$ and $s \in \{t-1, t, T\}$, the following equalities hold*[6]*:*

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

where $\mathrm{E}\left[\beta_t \,|\theta = \theta_i, D_s\right]$, $\mathrm{E}\left[y_t \,|x_t, D_s\right]$, $\mathrm{Var}\left[\beta_t \,|D_s\right]$ and $\mathrm{Var}\left[y_t \,|x_t, D_s\right]$ can be calculated analytically for each θ_i as in Propositions 2 and 3.

Thus, parameter estimates ($\mathrm{E}\left[\beta_t \,|D_s\right]$) and predictions ($\mathrm{E}\left[y_t \,|D_s\right]$) in any time period can be computed analytically and their variances are known in closed form.

[6] These are trivial consequences of the linearity of the integral (for the expected values) and of the law of total variance (for the variances).

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 $k^* = \arg\min \left\{k : \sum_{i=1}^{k} p_{T,i} \geq z\right\}$; 2) given k^*, extract β_T from the Student's t distribution $(\beta_T \,|\theta = \theta_{k^*}, D_T)$, which is given by Proposition 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 plain-vanilla OLS regressions with constant coefficients.

The above goals are pursued by extending Zellner's (1986) g-prior to TVC models and by parametrizing $\lambda(\theta)$ in such a way that the support of θ is invariant (it needs not 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 $t = 1$:

Assumption 5 *The prior mean is zero, corresponding to a prior belief of no predictability:*

$$\widehat{\beta}_{1,0} = 0 \tag{9}$$

while the prior covariance matrix is proportional to $\left(X^\top X\right)^{-1}$*:*

$$F_{\beta,1,0} = g\left(X^\top X\right)^{-1} \tag{10}$$

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, Ley and Steel - 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, Kohn and Wood - 1999):

Assumption 6 *The coefficient of proportionality is $g = T$.*

Thus, the amount of prior information (in the Fisher sense) about the coefficients is equal to the amount of average information contained in one observation from the sample.

Remark 7 *Given that $W^* = \lambda(\theta) F_{\beta,1,0}$ (equations 3 and 4), Zellner's prior (10) implies that also the covariance matrix of w_t is proportional to $\left(X^\top X\right)^{-1}$:*

$$W^* \propto \left(X^\top X\right)^{-1}$$

This proportionality condition has been imposed in a TVC model also by Stock and Watson[7] (1996), who borrow it from Nyblom (1989). A similar hypothesis is adopted also by Cogley and Sargent[8] (2001).

Remark 8 *Given (10) and (9), all the coefficients β_t have zero prior mean and covariance proportional to $\left(X^\top X\right)^{-1}$ conditional on D_0:*

$$\begin{aligned} \mathrm{E}\left[\beta_t \left| D_0, V, \theta\right.\right] &= 0 \ , \ t = 1 \ldots, T \ , \ \forall \theta \\ \mathrm{Var}\left[\beta_t \left| D_0, V, \theta\right.\right] &= \left[T + (t-1)\lambda(\theta)\right] V \left(X^\top X\right)^{-1} \ , \ t = 1, \ldots, T \ , \ \forall \theta \end{aligned}$$

[7] 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 $(X^\intercal X)^{-1}$. 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 to derive analytically a posterior distribution for λ.

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

This property will be used later, together with other properties of the priors, to prove that posterior model probabilities are scale invariant in the covariates.

2.2 The variance parameters $\widehat{V}_0$ 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:

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

With this choice, the updating equations in Proposition 2 would have to be replaced with a different set of updating equations until reaching the first non-zero observation of y_t (see e.g. 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, which yields almost identical results in reasonably sized samples:

Assumption 9 *The first observation in the sample (denote it by y_0) is used to form the prior on V :*

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

After using it to form the prior, we discard the first observation and start updating the equations (5)-(8) from the following 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

We have assumed that $W^* = \lambda\left(\theta\right) F_{\beta,1,0}$ where θ is a random variable having finite support $R_\theta = \{\theta_1, \ldots, \theta_q\} \subset [0, 1]$, $\theta_1 = 0$ and $\lambda\left(\theta\right)$ is strictly increasing in θ and such that $\lambda\left(\theta_1\right) = 0$. We now propose a specification of the function $\lambda\left(\theta\right)$ 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.

First, note that:

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

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

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

Assumption 10 *θ is the fraction of* $\mathrm{Var}\left[x_t w_t + v_t \,|x_t, D_0, \lambda\right]$ *generated on average by innovations to the regression coefficients:*

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

Given this assumption on θ, it is immediate to prove that

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

where

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

$\lambda\left(\theta\right)$ is strictly increasing in θ and such that $\lambda\left(0\right) = 0$, 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).

As far as the support of θ is concerned, we make the following assumption:

Assumption 11 *The support of θ is a geometrically spaced grid, consisting of q points:*

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

where $0 \leq \theta_{\max} < 1$ *and* $0 < c < 1$.

Notice that $\theta_{\max}$ cannot be chosen to be exactly equal to 1 (because $\lambda\left(1\right) = \infty$), but it can be set equal to any number arbitrarily close to 1.

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, Mammen and Spokoiny - 1997 and Spokoiny - 2001): in our model, it allows 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 grid[9].

If the geometric grid is considered as an approximation of a finer set of points (possibly a continuum), the geometric spacing ensures that the maximum relative round-off error is constant on all subintervals $[\theta_i, \theta_{i+1}]$ such that $1 < i < q$. The maximum relative round-off error is approximately $\frac{1-c}{2}$ on these subintervals and it can be controlled by an appropriate choice of c. On the contrary, the maximum relative round-off error cannot be controlled (it always equals 1) on the subinterval $[\theta_0, \theta_1]$, because the latter contains the point $\theta_0 = 0$. Only the absolute round-off error (equal to $\frac{\theta_{\max}}{2} c^{q-2}$) can be controlled on $[\theta_0, \theta_1]$, by an appropriate choice of q. Therefore, setting the two parameters c and q can be assimilated to setting the absolute and relative error tolerance in a numerical approximation problem.

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

Assumption 12 *The prior mixing probabilities are assumed to be:*

$$p\left(\theta_i \left| D_0 \right.\right) = q^{-1} \ , i = 1, \ldots, q$$

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.

2.4 Scale invariance

A crucial property of the automatic specification of priors proposed in the previous sections is that it guarantees scale invariance. The scale invariance property is satisfied if multiplying the regressors by an invertible matrix R, the posterior distribution of the coefficients is re-scaled accordingly[10] (it is multiplied by R^{-1}). Virtually all the TVC

[9] 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.

[10] For example, when a regressor is multiplied by 100, the posterior mean of the coefficient multiplying that regressor is divided by 100.

models we have found in the literature do not satisfy the scale invariance property, in the sense that they do not contemplate a mechanism to guarantee scale invariance by automatically re-scaling priors when the scale of regressors is changed. Although scale invariance might seem a trivial property, it is indispensable to achieve one of the main goals of this paper: having a completely automatic model that requires only regressors and regressands as inputs from the econometrician. Furthermore, it guarantees replicability of results: two researchers using the same data, but on different scales, will obtain the same results.

Scale invariance is formally defined as follows:

Definition 13 *Given the initial information set D_0, the information sets*

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

and a full-rank $k \times k$ matrix R, an initial information set D_0^* is said to be R-scale invariant with respect to D_0 if and only if:

$$(\beta_t \,| D_s^*) = \left(R^{-1}\beta_t \,| D_s\right) \qquad \forall s, t \leq T$$

where

$$D_t^* = D_{t-1}^* \cup \{y_t, x_t R\} \qquad t = 1, \dots, T$$

Note that the initial information set D_0, which contains the priors, is automatically specified as a function of y_0 and $\left(X^\top X\right)^{-1}$. We can write:

$$D_0 = D\left(y_0, \left(X^\top X\right)^{-1}\right) \tag{11}$$

The following proposition, proved in the Appendix, shows in what sense our TVC model is scale-invariant:

Proposition 14 *For any full-rank $k \times k$ matrix R, the initial information set D_0^* defined by:*

$$D_0^* = D\left(y_0, \left(R^\top X^\top X R\right)^{-1}\right)$$

is R-scale invariant with respect to the initial information set D_0, as defined in (11).

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 measure of stability, denoted by Π, is based on credible intervals (e.g.: Robert - 2007):

Definition 15 (Π-stability) *Let H_θ be a higher posterior probability set defined as follows:*[11]

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

The stability measure Π is defined by:

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

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

[11] H_θ contains all points of R_θ having higher posterior probability than $\theta = 0$ (recall that $\theta = 0$ means that regression coefficients are stable).

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 of stability, denoted by π, is constructed as a posterior odds ratio and it is based on the probability of the posterior mode of θ.

Definition 16 (π-stability) *Let p^* be the probability of (one of) the mode(s) of the posterior distribution of θ:*

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

The stability measure π is defined by:

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

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 γ and decide to reject the hypothesis of stability if the measure of stability is below the threshold ($\Pi < \gamma$ or $\pi < \gamma$). In case Π is used, the procedure can be assimilated to a frequentist test of hypothesis, where $1 - \gamma$ 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 stability[12]. In case π is used, one can resort to Jeffrey'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.

[12] Note, however, that the parallelism can be misleading, as Bayesian p-values have frequentist validity only 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.

4 Monte Carlo evidence

4.1 Performance when the 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 plain vanilla 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 and 2003) sequential procedure, as implemented by Pesaran and Timmermann (2002 and 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 z_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

$$x_t = \left[\begin{array}{ccc} 1 & y_{t-1} & u_{t-1} \end{array}\right]$$

  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

$$x_t = \left[\begin{array}{ccccccc} 1 & y_{t-1} & y_{t-2} & y_{t-3} & u_{t-1} & u_{t-2} & u_{t-3} \end{array}\right]$$

- *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.

Each Monte Carlo experiment consists of 10,000 simulations.

The loss in estimation precision is evaluated comparing the estimate of the coefficient vector at time T (denote it by $\widetilde{\beta}_T$) with its true value. We consider seven different estimates:

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

$$\widetilde{\beta}_T = \mathrm{E}\left[\beta_T \left| D_T \right.\right] = \sum_{i=1}^{q} p_{T,i} \mathrm{E}\left[\beta_T \left| D_T, \theta = \theta_i \right.\right]$$

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

$$\widetilde{\beta}_T = \mathrm{E}\left[\beta_T \left| D_T, \theta = \theta_{j^*} \right.\right]$$

  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):

$$\widetilde{\beta}_T = \begin{cases} \mathrm{E}\left[\beta_T \left| D_T, \theta = 0 \right.\right] \text{ if } \Pi \geq 0.1 \\ \mathrm{E}\left[\beta_T \left| D_T \right.\right] \text{ if } \Pi < 0.1 \end{cases}$$

  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):

$$\widetilde{\beta}_T = \begin{cases} \mathrm{E}\left[\beta_T \left| D_T, \theta = 0\right.\right] \text{ if } \pi \geq 0.1 \\ \mathrm{E}\left[\beta_T \left| D_T\right.\right] \text{ if } \pi < 0.1 \end{cases}$$

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):

$$\widetilde{\beta}_T = \mathrm{E}\left[\beta_T \left| D_T, \theta = 0\right.\right]$$

- OLS estimates obtained from Bai and Perron's (1998 and 2003) sequential[13] procedure (denoted by BP), using the SIC criterion to choose the number of breakpoints (Pesaran and Timmermann - 2002 and 2007). If $\widetilde{\tau}$ is the last estimated breakpoint date in the sample, then $\widetilde{\beta}_T$ is the OLS estimate of β_T obtained using all the sample points from $\widetilde{\tau}$ to T;

- estimates obtained from Pesaran and Timmermann'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 $\widetilde{\tau}$ is the last estimated breakpoint date in the sample, then:

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

  where $\widetilde{\beta}_{T,\tau}$ 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 ($\tau + k + 1 \leq t \leq T$).

[13] We estimate the breakpoint dates sequentially rather than simultaneously to achieve a reasonable computational speed in our Monte Carlo simulations. Denote by ν_S the number of breakpoints estimated by the sequential procedure and by ν_σ the number estimated by the simultaneous procedure. Given that we are using the SIC criterion to choose the number of points, if $\nu_\sigma \leq 1$, then $\nu_S = \nu_\sigma$; otherwise, if $\nu_\sigma > 1$, then $\nu_S \leq \nu_\sigma$. Therefore, in our Montecarlo 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.

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

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

where $\|\|$ is the Euclidean norm and j =TVC-MA, TVC-MS, TVC-Π, TVC-π, 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 choose a value that is numerically close to 1 ($\theta_{\max} = 0.999$). Thus, the relative round-off error is bounded at 5 per cent and the model is able to detect degrees of instability as low as $\theta \simeq 3 \cdot 10^{-5}$ (for concreteness, this means that coefficient instability can be detected by the model also in cases in which less than 0.01 per cent of total innovation variance is generated by coefficient instability).

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. As anticipated in the introduction, 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-Π. MSE^{β}_{TVC-MA} and MSE^{β}_{TVC-MS} are roughly between 20 and 60 per cent higher than $MSE^{\beta}_{TVC-OLS}$, while MSE^{β}_{TVC-BP} is on average equal to several multiples of $MSE^{\beta}_{TVC-OLS}$. Qualitatively speaking, the loss in precision from using TVC-Π, 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^{β}_{TVC} (all four kinds of TVC) and MSE^{β}_{BP} become almost two orders of magnitude greater than MSE^{β}_{OLS}, while MSE^{β}_{BP-MA} remains comparable to MSE^{β}_{OLS} (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:

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

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 intercept[14]. 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 3 percent increase in MSE with respect to OLS). All the other estimates (TVC-MA, TVC-MS, TVC-Π and BP) are somewhat less efficient, but their MSEs seldom exceed those of the OLS estimates by more than 30%. 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

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

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 $\widetilde{y}_{T+1}$) is compared to the actual value y_{T+1}. As for coefficient estimates, we consider seven different predictions:

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

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

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

$$\widetilde{y}_{T+1} = \mathrm{E}\left[y_{T+1} \left| D_T, x_{T+1}, \theta = \theta_{j^*}\right.\right]$$

- 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):

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

- 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):

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

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

$$\widetilde{y}_{T+1} = \mathrm{E}\left[y_{T+1} \left| D_T, x_{T+1}, \theta = 0\right.\right]$$

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

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

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

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

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

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

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

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

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

Since $\mathrm{E}\left[v_{T+1}^2\right]$ 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 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.

The estimation precision and the forecasting performance are evaluated 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 of the MSE with respect to OLS. Although TVC-MA and TVC-MS perform slightly better than TVC-Π 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 break-dates 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 apply also 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 the 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.

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}; 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 also more accurate than the BP-MA forecasts).

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:

$$\begin{aligned} y_t &= \rho y_{t-1} + b_t u_{t-1} + v_t \\ b_t &= b_{t-1} + w_t \end{aligned}$$

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

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

Note that, although one coefficient of the regression is frequently changing (b_t), the other coefficient (ρ) is stable. As a consequence, the true DGP does not fit exactly 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 to better understand 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-Π 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 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, also with frequently changing coefficients, BP-MA seems to be the only method capable of dealing simultaneously 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. Similarly 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.

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 $\rho = 0.99$ 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).

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 and 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,

$$x_t = \left[\begin{array}{cccc} 1 & r_{M,t} - r_{f,t} & SMB_t & HML_t \end{array} \right]$$

$r_{M,t}$ 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 only for 11 stocks out of 432. Furthermore, $\Pi < 0.1$ and $\pi < 0.1$ for 92% and 81% of the stocks respectively. On average, Π is 0.046 and π is 0.010. Also the frequentist method provides evidence that most stocks experience instability in their risk exposures: according to the BP sequential estimates, more than 78% 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 =TVC-MA, TVC-MS, TVC-Π, TVC-π, OLS, BP, BP-MA). For each stock i and for each prediction method j, the mean squared error is computed as:

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

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

To be able 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 $GAIN_{i,j}$ (each stock i represents a 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.

6 Conclusions

We have proposed a Bayesian regression model with time-varying coefficients (TVC). With respect to existing TVC models, we have introduced some technical innovations aimed at making TVC models less computationally expensive and completely automatic (by completely automatic we mean that regressors and regressands are the only input required from the econometrician, so that he/she does not need to engage in technically demanding specifications of priors and model parametrizations).

We have conducted several Monte Carlo experiments to understand the pros and cons that might be encountered when using the TVC model in applied econometric analyses. We have found that the cons are generally limited, in the sense that the TVC model has satisfactory estimation precision and forecasting performance also when regression coefficients are indeed stable or when coefficient instability is present but the TVC model is mis-specified. In the presence of coefficient instability, there are potential rewards from using the TVC model: in some cases, its estimation precision and forecasting accuracy are significantly better than those of competing models.

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

Before concluding, two remarks on the applicability of our TVC model are in order. First, we have confined attention to single equation regression models, but the results presented in the paper can be extended in a straightforward manner to multiple equation models (for example VARs), by imposing the usual normal / inverse Wishart priors on the initial parameters. Second, we have not discussed the use of the model for the analysis of cross-sectional data: however, it is possible to use TVC models like ours to analyze cross-sectional data in the presence of non-linearities that are not explicitly captured by the regressors (see West and Harrison - 1997); this is usually accomplished by replacing the time index t with the rank statistic of the regressor that is presumably responsible for the non-linearity.

References

[1] Andrews, D. W. K. (1993) "Tests for parameter instability and structural change with unknown change point", *Econometrica*, 61, 821-56.

[2] Andrews, D. W. K., I. Lee and W. Ploberger (1996) "Optimal changepoint tests for normal linear regression model", *Journal of Econometrics*, 70, 9-38.

[3] Andrews, D. W. K. and W. Ploberger (1994) "Optimal tests when a nuisance parameter is present only under the alternative", *Econometrica*, 62, 1383-1414.

[4] Bai, J. and P. Perron (1998) "Estimating and testing linear models with multiple structural breaks", *Econometrica*, 66, 47-78.

[5] Bai, J. and P. Perron (2006) "Multiple structural change models: a simulation analysis", *Econometric theory and practice: frontier of analysis and applied research (Essays in honor of Peter Phillips)* ed. by Corbae, D., S. Durlauf and B. E. Hansen, Cambridge University Press.

[6] Brown, R., L., J. Durbin and J. M. Evans (1975) "Techniques for testing the constancy of regression coefficients over time", *Journal of the Royal Statistical Society B*, 37, 149-192.

[7] Canova, F. and M. Ciccarelli (2009) "Estimating multi-country VAR models", *International Economic Review*, 929–959.

[8] Canova, F. and L. Gambetti (2009) "Structural changes in the US economy: is there a role for monetary policy?", *Journal of Economic Dynamics and Control*, 477–490.

[9] Carter, C. K. and R. Kohn (1994) “On Gibbs sampling for state space models”, *Biometrika*, 81, 541-553.

[10] Chib, S. and E. Greenberg (1995) “Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models”, *Journal of Econometrics*, 68, 339-360.

[11] Chow, G. C. (1960) "Tests of equality between sets of coefficients in two linear regressions", *Econometrica*, 28, 591–605.

[12] Cogley, T., G. E. Primiceri and T. J. Sargent (2010) "Inflation-gap persistence in the US", *American Economic Journal: Macroeconomics*, 43-69.

[13] Cogley, T. and T.J. Sargent (2001) "Evolving Post World War II U.S. Inflation Dynamics", *NBER Macroeconomics Annual*, 16, 331-373.

[14] Datta, G. S. (1996) "On priors providing frequentist validity of Bayesian inference for multiple parametric functions", *Biometrika*, 83, 287-298.

[15] Datta, G. S. and J. K. Ghosh (1995) "On priors providing frequentist validity for Bayesian inference", *Biometrika*, 82, 37-45.

[16] Diebold, F. X. and C. Chen (1996) "Testing structural stability with endogenous breakpoint: a size comparison of analytic and bootstrap procedures", *Journal of Econometrics*, 70, 221-241.

[17] Doan, T., R. B. Litterman and C. A. Sims (1984) "Forecasting and conditional projection using realistic prior distributions", *Econometric Reviews*, 3, 1-100.

[18] Fama, E. and K. French (1993) "Common risk factors in the returns on bonds and stocks", *Journal of Financial Economics*, 33, 3-56.

[19] Fama, E. and K. French (1995) "Size and book-to-market factors in earnings and returns", *Journal of Finance*, 50, 131-155.

[20] Fama, E. and K. French (1996) "Multifactor explanations of asset pricing anomalies", *Journal of Finance*, 51, 55-84.

[21] Fernandez, C., E. Ley and M. F. Steel (2001) "Benchmark priors for Bayesian model averaging", *Journal of Econometrics*, 100, 381-427.

[22] George, E. I. and R. E. McCulloch (1997) "Approaches for Bayesian variable selection", *Statistica Sinica*, 7, 339-374.

[23] Ghosh, J. K. and R. Mukerjee (1993) "Frequentist validity of highest posterior density regions in the multiparameter case", *Annals of the Institute of Statistical Mathematics*, 45, 293-302.

[24] Guerre, E. and P. Lavergne (2005) "Data-driven rate-optimal specification testing in regression models", *The Annals of Statistics*, 33, 840-870.

[25] Hamilton J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, USA.

[26] Hansen, B. E. (1997) "Approximate Asymptotic P Values for Structural-Change Tests", *Journal of Business & Economic Statistics*, 15, 60-67.

[27] Hansen, B. E. (2000) "Testing for structural change in conditional models", *Journal of Econometrics*, 97, 93-115.

[28] Hatanaka, M. and K. Yamada (1999) "A unit root test in the presence of structural changes in I(1) and I(0) models", *Cointegration, causality and forecasting: a Festschrift in honour of Clive W. J Granger* ed. by Engle R. F. and H. White, Oxford University Press.

[29] Horowitz, J. L. and V. G. Spokoiny (2001) "An adaptive, rate-optimal test of a parametric model against a nonparametric alternative", *Econometrica*, 69, 599-631.

[30] Jeffreys, H. (1961) *The theory of probability* (3e), Oxford University Press.

[31] Kapetanios, G. (2008) "Bootstrap-based tests for deterministic time-varying coefficients in regression models", *Computational Statistics & Data Analysis*, 15, 534-545.

[32] Kass, R. E. and L. Wasserman (1995) "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion", *Journal of the American Statistical Association*, 90, 928-934.

[33] Koop, G. (2003) *Bayesian Econometrics*, John Wiley and Sons, Chichester, England.

[34] Koop, G., R. Leon-Gonzalez, and R. Strachan (2009) "On the evolution of the monetary policy transmission mechanism", *Journal of Economic Dynamics and Control*, 997–1017.

[35] Koop, G. and D. Korobilis (2009) "Bayesian multivariate time series methods for empirical macroeconomics", *Foundations and trends in econometrics*, 267–358.

[36] Lepski, O. V., E. Mammen and V.G. Spokoiny (1997) "Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors", *Annals of Statistics*, 25, 929-947.

[37] Liang F., R. Paulo, G. Molina, M. A. Clyde and J. O. Berger (2008) "Mixtures of g priors for Bayesian variable selection", *Journal of the American Statistical Association*, 103, 410-423.

[38] Mukerjee, R. and D. K. Dey (1993) "Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics", *Biometrika*, 80, 499-505.

[39] Nyblom, J. (1989): "Testing for the constancy of parameters over time", *Journal of the American Statistical Association*, 84, 348-368.

[40] Perron, P. and X. Zhu (2005) "Structural breaks with deterministic and stochastic trends", *Journal of Econometrics*, 129, 65-119.

[41] Robert, C. P. (2007) *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Springer Verlag.

[42] Sargent, T., Williams, N. and T. Zha (2006) "Shocks and government beliefs: the rise and fall of American inflation", *American Economic Review*, 1193-1224.

[43] Shively, T. S., R. Kohn and S. 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.

[44] Smith, M. and R. Kohn (1996) "Nonparametric regression using Bayesian variable selection", *Journal of Econometrics*, 75, 317-343.

[45] Spokoiny, V.G. (2001) "Data-driven testing the fit of linear models", *Mathematical Methods of Statistics*, 10, 465-497.

[46] Stock, J. H. and M. W. Watson (1996) "Evidence on structural instability in macroeconomic time series relations", *Journal of Business & Economic Statistics*, 14, 11-30.

[47] West, M. and J. Harrison (1997), *Bayesian forecasting and dynamic models*, Second Edition, Springer Verlag, New York.

[48] Zellner, A. (1986) "On assessing prior distributions and Bayesian regression analysis with g-prior distributions", *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, eds. P. K. Goel and A. Zellner, 233-243, North-Holland/Elsevier.

7 Appendix

7.1 Proofs of propositions 2 and 3

In this section we derive the formulae presented in Propositions 2 and 3. To facilitate the exposition, we start from simpler information structures and then we tackle the more complex information structure assumed in Propositions 2 and 3 and summarized in Assumption 1.

7.1.1 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 17 (Priors and initial information) *The priors on the unknown parameters are:*

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

and the initial information set is:

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

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 18 (Forward updating) *Let priors and initial information be as in Case 17. Then:*

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

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

$$\begin{array}{ll} \widehat{\beta}_{t,t-1} = \widehat{\beta}_{t-1,t-1} & F_{\beta,t,t-1} = F_{\beta,t-1,t-1} + W^* \\ \widehat{y}_{t,t-1} = x_t \widehat{\beta}_{t,t-1} & F_{y,t,t-1} = 1 + x_t F_{\beta,t,t-1} x_t^\top \\ e_t = y_t - \widehat{y}_{t,t-1} & P_t = F_{\beta,t,t-1} x_t^\top / F_{y,t,t-1} \\ \widehat{\beta}_{t,t} = \widehat{\beta}_{t,t-1} + P_t e_t & F_{\beta,t,t} = F_{\beta,t,t-1} - P_t P_t^\top F_{y,t,t-1} \end{array} \tag{12}$$

starting from the initial values $\widehat{\beta}_{1,0}$ *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 (12) are just the usual updating equations of the Kalman filter (e.g.: Hamilton - 1994). ■

The smoothing equations are provided by the following proposition:

Proposition 19 (Backward updating) *Let priors and initial information be as in Case 17. Then:*

$$\left(\beta_{T-\tau} \left| D_T^{**}\right.\right) \sim N\left(\widehat{\beta}_{T-\tau,T}, V F_{\beta,T-\tau,T}\right)$$

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

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

and the backward recursions start from the terminal values of the forward recursions (12).

Proof. These are the usual backward Kalman recursions (e.g.: Hamilton - 1994). ■

7.1.2 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 20 (Priors and initial information) *The priors on the unknown parameters are:*

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

and the initial information set is:

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

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 21 (Forward updating) *Let priors and initial information be as in Case 20. Then:*

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

where the parameters of the above distributions are calculated recursively as in (12) 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:

$$\begin{aligned} \left(\beta_1 \left| D_1^*, V \right.\right) &\sim N\left(\widehat{\beta}_{1,1}, VF_{\beta,1,1}\right) & (13) \\ \left(1/V \left| D_1^* \right.\right) &\sim G\left(\widehat{V}_1, n_1\right) & (14) \end{aligned}$$

Since $\beta_2 = \beta_1 + w_2$ and

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

then, by the additivity of normal distributions:

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

Therefore, at time $t = 2$, $p\left(\beta_2 \left| D_1^*, V \right.\right)$ and $p\left(1/V \left| D_1^* \right.\right)$ 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 22 (Backward updating) *Let priors and initial information be as in Case 20. Then:*

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

where $\widehat{V}_T$ and n_T are calculated as in Proposition 18 and the other parameters of the above distributions are calculated recursively (backwards) as in Proposition 19.

Proof. From Proposition 19, we know that:

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

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

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

■

7.1.3 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 2 and 3):

Case 23 *The priors on the unknown parameters are:*

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

and the initial information set is:

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

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 17 and 20 in the previous subsections had the only purpose of introducing the more complex Case 23. Given the above assumptions, the posterior distributions of the parameters of the regression can be calculated as follows:

Proposition 24 *Let priors and initial information be as in Case 23. 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:

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

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

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

are calculated for each θ_i *as in Propositions 18 and 21.*

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 18 and 21 (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:

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

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 2 in the main text is obtained by combining propositions 18, 21 and 24 above. Proposition 3 results from propositions 19, 22 and 24 above.

7.2 Proof of proposition 14 (scale invariance)

When $x_t R$ is the vector of regressors, the prior covariance is:

$$F_{\beta,1,0} = T\left(R^\top X^\top X R\right)^{-1} = T R^{-1}\left(X^\top X\right)^{-1}\left(R^{-1}\right)^\top$$

The constant ω is unaffected by the rotation:

$$\begin{aligned}
\omega &= \frac{1}{T}\sum_{t=1}^{T} x_t R F_{\beta,1,0} R^\top x_t^\top \\
&= \frac{1}{T}\sum_{t=1}^{T} x_t R \left(T R^{-1}\left(X^\top X\right)^{-1}\left(R^{-1}\right)^\top\right) R^\top x_t^\top \\
&= \frac{1}{T}\sum_{t=1}^{T} x_t \left(T\left(X^\top X\right)^{-1}\right) x_t^\top
\end{aligned}$$

Since R_θ does not depend on the data and

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

the fact that ω does not change implies that also R_λ (the support of λ) remains unchanged. The prior probabilities assigned to the elements of R_λ also do not depend on the data. So, the prior distribution of λ is unaffected by the rotation.

As far as the recursive equations in Proposition 2 are concerned, note that the initial conditions

$$\begin{aligned}\widehat{\beta}_{1,0,i} &= \widehat{\beta}_{1,0} = 0 \\ \widehat{V}_{0,i} &= \widehat{V}_0 = y_0^2 \\ n_{0,i} &= n_0 = 1\end{aligned}$$

are not affected by the rotation, while the initial condition

$$F_{\beta,1,0,i} = F_{\beta,1,0} = TR^{-1}\left(X^\top X\right)^{-1}\left(R^{-1}\right)^\top$$

changes (it is pre-multiplied by R^{-1} and post-multiplied by $(R^{-1})^\top$).

For $t > 0$, it can be easily checked that $\widehat{y}_{t,t-1,i}$, $F_{y,t,t-1,i}$, $e_{t,i}$, $n_{t,i}$, $\widehat{V}_{t,i}$ remain unchanged, while $F_{\beta,t,t-1,i}$ and $F_{\beta,t,t,i}$ are pre-multiplied by R^{-1} and post-multiplied by $(R^{-1})^\top$ and $\widehat{\beta}_{t,t-1,i}$, $\widehat{\beta}_{t,t,i}$ and $P_{t,i}$ are pre-multiplied by R^{-1}. Therefore:

$$\left(\beta_t \left| D_{t-1}^*, \lambda = \lambda\left(\theta_i\right)\right.\right) = \left(R^{-1}\beta_t \left| D_{t-1}, \lambda = \lambda\left(\theta_i\right)\right.\right) \qquad \forall t \le T,\ i = 1, \dots, q$$

and

$$\left(\beta_t \left| D_t^*, \lambda = \lambda\left(\theta_i\right)\right.\right) = \left(R^{-1}\beta_t \left| D_t, \lambda = \lambda\left(\theta_i\right)\right.\right) \qquad \forall t \le T,\ i = 1, \dots, q$$

The model probabilities $p_{t,1}, \dots, p_{t,q}$ depend only on $\widehat{y}_{t,t-1,i}$, $F_{y,t,t-1,i}$, $n_{t-1,i}$ and $\widehat{V}_{t-1,i}$, which remain unchanged, so they remain unchanged as well. As a consequence, also unconditionally:

$$\left(\beta_t \left| D_s^*\right.\right) = \left(R^{-1}\beta_t \left| D_s\right.\right) \qquad \forall t \le T$$

for $s = t-1$ or $s = t$. Using similar arguments on the backward recursions of proposition 3, it is possible to prove that the above equality holds for any $s \le T$.

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.0257 (0.0003)	0.0286 (0.0003)	0.0239 (0.0003)	0.0215 (0.0002)	0.0205 (0.0002)	0.0540 (0.0031)	0.0221 (0.0002)
T=200	0.0138 (0.0001)	0.0142 (0.0002)	0.0121 (0.0001)	0.0106 (0.0001)	0.0102 (0.0001)	0.0479 (0.0150)	0.0108 (0.0001)
T=500	0.0066 (0.0001)	0.0056 (0.0001)	0.0049 (0.0001)	0.0043 (0.0000)	0.0040 (0.0000)	0.0070 (0.0006)	0.0043 (0.0002)
	ρ=0.50						
T=100	0.0271 (0.0003)	0.0303 (0.0004)	0.0249 (0.0003)	0.0219 (0.0003)	0.0207 (0.0002)	0.0730 (0.0050)	0.0227 (0.0003)
T=200	0.0136 (0.0001)	0.0141 (0.0002)	0.0119 (0.0001)	0.0104 (0.0001)	0.0099 (0.0001)	0.0329 (0.0029)	0.0106 (0.0002)
T=500	0.0064 (0.0001)	0.0055 (0.0001)	0.0048 (0.0001)	0.0041 (0.0000)	0.0038 (0.0000)	0.0132 (0.0034)	0.0039 (0.0000)
	ρ=0.80						
T=100	0.0308 (0.0004)	0.0352 (0.0006)	0.0278 (0.0004)	0.0234 (0.0003)	0.0219 (0.0003)	0.1537 (0.0241)	0.0272 (0.0020)
T=200	0.0141 (0.0002)	0.0145 (0.0002)	0.0121 (0.0002)	0.0103 (0.0001)	0.0097 (0.0001)	0.0853 (0.0163)	0.0105 (0.0002)
T=500	0.0062 (0.0001)	0.0053 (0.0001)	0.0046 (0.0001)	0.0039 (0.0000)	0.0037 (0.0000)	0.0179 (0.0035)	0.0038 (0.0000)
	ρ=0.99						
T=100	4.7933 (0.1551)	5.3498 (0.1669)	4.7233 (0.1550)	4.7344 (0.1553)	0.0666 (0.0013)	1.9222 (0.3944)	0.1145 (0.0115)
T=200	2.5390 (0.0970)	2.7297 (0.1010)	2.5072 (0.0970)	2.5057 (0.0970)	0.0244 (0.0005)	0.7189 (0.1093)	0.0367 (0.0032)
T=500	0.4448 (0.0249)	0.4792 (0.0259)	0.4301 (0.0249)	0.4287 (0.0249)	0.0062 (0.0001)	0.1338 (0.0195)	0.0072 (0.0002)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.0885 (0.0007)	0.0964 (0.0009)	0.0860 (0.0007)	0.0804 (0.0006)	0.0785 (0.0006)	0.0827 (0.0013)	0.0795 (0.0006)
T=200	0.0433 (0.0003)	0.0453 (0.0004)	0.0414 (0.0003)	0.0385 (0.0003)	0.0376 (0.0003)	0.0379 (0.0003)	0.0378 (0.0003)
T=500	0.0186 (0.0001)	0.0175 (0.0001)	0.0162 (0.0001)	0.0150 (0.0001)	0.0146 (0.0001)	0.0146 (0.0001)	0.0146 (0.0001)
	ρ=0.50						
T=100	0.0953 (0.0008)	0.1033 (0.0009)	0.0920 (0.0007)	0.0863 (0.0007)	0.0842 (0.0006)	0.0905 (0.0021)	0.0850 (0.0006)
T=200	0.0457 (0.0004)	0.0478 (0.0004)	0.0434 (0.0003)	0.0406 (0.0003)	0.0398 (0.0003)	0.0399 (0.0003)	0.0398 (0.0003)
T=500	0.0194 (0.0002)	0.0184 (0.0002)	0.0170 (0.0001)	0.0159 (0.0001)	0.0155 (0.0001)	0.0157 (0.0002)	0.0156 (0.0001)
	ρ=0.80						
T=100	0.1049 (0.0009)	0.1134 (0.0010)	0.1017 (0.0008)	0.0959 (0.0008)	0.0940 (0.0007)	0.1024 (0.0041)	0.0947 (0.0007)
T=200	0.0511 (0.0004)	0.0532 (0.0005)	0.0484 (0.0004)	0.0454 (0.0004)	0.0442 (0.0003)	0.0447 (0.0004)	0.0443 (0.0003)
T=500	0.0215 (0.0002)	0.0203 (0.0002)	0.0188 (0.0002)	0.0174 (0.0001)	0.0169 (0.0001)	0.0169 (0.0001)	0.0169 (0.0001)
	ρ=0.99						
T=100	0.7213 (0.0409)	0.9599 (0.0498)	0.6244 (0.0402)	0.6713 (0.0408)	0.1505 (0.0016)	0.1830 (0.0072)	0.1531 (0.0019)
T=200	0.2619 (0.0203)	0.3052 (0.0226)	0.2317 (0.0199)	0.2425 (0.0202)	0.0632 (0.0006)	0.0867 (0.0142)	0.0636 (0.0007)
T=500	0.0327 (0.0017)	0.0310 (0.0018)	0.0279 (0.0017)	0.0264 (0.0017)	0.0213 (0.0002)	0.0213 (0.0002)	0.0213 (0.0002)

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.0384 (0.0010)	1.0438 (0.0013)	1.0358 (0.0009)	1.0319 (0.0008)	1.0301 (0.0006)	1.0942 (0.0098)	1.0332 (0.0008)
T=200	1.0208 (0.0005)	1.0214 (0.0005)	1.0181 (0.0004)	1.0161 (0.0004)	1.0155 (0.0004)	1.2914 (0.2190)	1.0165 (0.0004)
T=500	1.0101 (0.0002)	1.0084 (0.0002)	1.0073 (0.0002)	1.0064 (0.0002)	1.0060 (0.0001)	1.0121 (0.0026)	1.0073 (0.0012)
	ρ=0.50						
T=100	1.0393 (0.0009)	1.0443 (0.0011)	1.0366 (0.0009)	1.0325 (0.0008)	1.0311 (0.0007)	1.0978 (0.0126)	1.0351 (0.0013)
T=200	1.0217 (0.0006)	1.0229 (0.0007)	1.0193 (0.0006)	1.0170 (0.0005)	1.0159 (0.0004)	1.0443 (0.0047)	1.0168 (0.0004)
T=500	1.0101 (0.0003)	1.0086 (0.0002)	1.0075 (0.0002)	1.0065 (0.0002)	1.0060 (0.0001)	1.0144 (0.0027)	1.0061 (0.0001)
	ρ=0.80						
T=100	1.0453 (0.0011)	1.0526 (0.0014)	1.0423 (0.0011)	1.0371 (0.0010)	1.0348 (0.0009)	1.1301 (0.0212)	1.0420 (0.0021)
T=200	1.0218 (0.0006)	1.0230 (0.0008)	1.0191 (0.0006)	1.0162 (0.0003)	1.0155 (0.0003)	1.0757 (0.0212)	1.0166 (0.0005)
T=500	1.0103 (0.0003)	1.0089 (0.0003)	1.0078 (0.0002)	1.0065 (0.0002)	1.0062 (0.0001)	1.0229 (0.0084)	1.0063 (0.0001)
	ρ=0.99						
T=100	1.2256 (0.0089)	1.2710 (0.0102)	1.2238 (0.0089)	1.2206 (0.0089)	1.0489 (0.0010)	1.2199 (0.0630)	1.0484 (0.0015)
T=200	1.1349 (0.0046)	1.1548 (0.0051)	1.1317 (0.0046)	1.1306 (0.0046)	1.0224 (0.0005)	1.0496 (0.0030)	1.0225 (0.0008)
T=500	1.0419 (0.0018)	1.0462 (0.0019)	1.0392 (0.0018)	1.0377 (0.0018)	1.0078 (0.0002)	1.0198 (0.0032)	1.0078 (0.0002)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.0841 (0.0016)	1.0914 (0.0018)	1.0813 (0.0015)	1.0762 (0.0014)	1.0746 (0.0014)	1.0808 (0.0032)	1.0759 (0.0014)
T=200	1.0433 (0.0010)	1.0454 (0.0011)	1.0412 (0.0009)	1.0386 (0.0008)	1.0375 (0.0008)	1.0379 (0.0008)	1.0379 (0.0008)
T=500	1.0183 (0.0003)	1.0171 (0.0003)	1.0158 (0.0003)	1.0146 (0.0003)	1.0142 (0.0003)	1.0143 (0.0003)	1.0143 (0.0003)
	ρ=0.50						
T=100	1.0845 (0.0016)	1.0923 (0.0019)	1.0817 (0.0016)	1.0769 (0.0014)	1.0752 (0.0014)	1.0928 (0.0119)	1.0772 (0.0015)
T=200	1.0435 (0.0008)	1.0454 (0.0009)	1.0412 (0.0008)	1.0380 (0.0007)	1.0372 (0.0007)	1.0375 (0.0007)	1.0375 (0.0007)
T=500	1.0179 (0.0003)	1.0167 (0.0003)	1.0154 (0.0003)	1.0144 (0.0003)	1.0141 (0.0003)	1.0141 (0.0003)	1.0141 (0.0003)
	ρ=0.80						
T=100	1.0889 (0.0017)	1.0967 (0.0019)	1.0864 (0.0016)	1.0816 (0.0015)	1.0802 (0.0014)	1.0855 (0.0024)	1.0815 (0.0015)
T=200	1.0418 (0.0008)	1.0435 (0.0009)	1.0397 (0.0008)	1.0373 (0.0007)	1.0363 (0.0007)	1.0369 (0.0007)	1.0367 (0.0007)
T=500	1.0177 (0.0003)	1.0168 (0.0003)	1.0155 (0.0003)	1.0143 (0.0003)	1.0139 (0.0002)	1.0140 (0.0002)	1.0140 (0.0002)
	ρ=0.99						
T=100	1.1381 (0.0065)	1.1705 (0.0080)	1.1298 (0.0062)	1.1353 (0.0064)	1.0943 (0.0017)	1.1128 (0.0145)	1.0886 (0.0016)
T=200	1.0647 (0.0031)	1.0727 (0.0038)	1.0609 (0.0030)	1.0617 (0.0030)	1.0439 (0.0008)	1.0504 (0.0075)	1.0425 (0.0008)
T=500	1.0196 (0.0004)	1.0192 (0.0005)	1.0178 (0.0004)	1.0171 (0.0004)	1.0163 (0.0003)	1.0160 (0.0003)	1.0160 (0.0003)

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.1772 (0.0051)	0.1767 (0.0050)	0.1807 (0.0051)	0.1943 (0.0054)	0.3769 (0.0078)	0.1737 (0.0060)	0.2252 (0.0055)
T=200	0.1148 (0.0033)	0.1145 (0.0033)	0.1183 (0.0034)	0.1250 (0.0036)	0.3434 (0.0068)	0.1088 (0.0068)	0.1863 (0.0043)
T=500	0.0689 (0.0026)	0.0692 (0.0026)	0.0711 (0.0026)	0.0741 (0.0027)	0.3453 (0.0072)	0.0452 (0.0049)	0.1738 (0.0043)
	ρ=0.50						
T=100	0.1790 (0.0046)	0.1792 (0.0046)	0.1831 (0.0047)	0.1951 (0.0049)	0.3655 (0.0073)	0.2140 (0.0134)	0.2248 (0.0052)
T=200	0.1237 (0.0040)	0.1228 (0.0040)	0.1265 (0.0041)	0.1346 (0.0043)	0.3545 (0.0076)	0.1053 (0.0047)	0.1930 (0.0049)
T=500	0.0692 (0.0026)	0.0694 (0.0026)	0.0713 (0.0026)	0.0744 (0.0027)	0.3452 (0.0072)	0.0533 (0.0079)	0.1733 (0.0043)
	ρ=0.80						
T=100	0.1887 (0.0044)	0.1906 (0.0044)	0.1924 (0.0045)	0.2038 (0.0047)	0.3705 (0.0071)	0.3016 (0.0206)	0.2226 (0.0050)
T=200	0.1261 (0.0037)	0.1255 (0.0037)	0.1290 (0.0038)	0.1360 (0.0040)	0.3463 (0.0070)	0.1309 (0.0065)	0.1894 (0.0046)
T=500	0.0707 (0.0025)	0.0708 (0.0026)	0.0729 (0.0026)	0.0760 (0.0027)	0.3449 (0.0072)	0.0803 (0.0192)	0.1733 (0.0043)
	ρ=0.99						
T=100	5.1868 (0.1677)	5.7741 (0.1828)	5.1508 (0.1677)	5.1618 (0.1677)	0.4283 (0.0075)	2.5268 (0.2692)	0.4261 (0.0130)
T=200	3.3069 (0.1269)	3.5054 (0.1341)	3.2908 (0.1268)	3.2956 (0.1269)	0.3626 (0.0070)	1.4681 (0.1731)	0.2585 (0.0063)
T=500	0.7054 (0.0257)	0.7362 (0.0268)	0.7032 (0.0257)	0.7053 (0.0257)	0.3542 (0.0073)	0.6431 (0.1556)	0.1946 (0.0074)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.3299 (0.0057)	0.3368 (0.0057)	0.3322 (0.0058)	0.3424 (0.0060)	0.4389 (0.0074)	0.3628 (0.0110)	0.3320 (0.0058)
T=200	0.2127 (0.0042)	0.2144 (0.0042)	0.2145 (0.0042)	0.2202 (0.0044)	0.3761 (0.0069)	0.2186 (0.0121)	0.2421 (0.0047)
T=500	0.1357 (0.0035)	0.1361 (0.0035)	0.1372 (0.0035)	0.1400 (0.0036)	0.3540 (0.0070)	0.1053 (0.0042)	0.1974 (0.0044)
	ρ=0.50						
T=100	0.3272 (0.0054)	0.3347 (0.0054)	0.3291 (0.0055)	0.3377 (0.0057)	0.4368 (0.0074)	0.3948 (0.0194)	0.3292 (0.0057)
T=200	0.2201 (0.0043)	0.2218 (0.0043)	0.2215 (0.0044)	0.2270 (0.0045)	0.3826 (0.0069)	0.2194 (0.0097)	0.2460 (0.0048)
T=500	0.1401 (0.0036)	0.1405 (0.0036)	0.1413 (0.0037)	0.1435 (0.0037)	0.3507 (0.0073)	0.1078 (0.0039)	0.1956 (0.0046)
	ρ=0.80						
T=100	0.3529 (0.0058)	0.3627 (0.0059)	0.3547 (0.0059)	0.3628 (0.0060)	0.4585 (0.0076)	0.4161 (0.0110)	0.3476 (0.0058)
T=200	0.2420 (0.0049)	0.2441 (0.0048)	0.2436 (0.0049)	0.2499 (0.0051)	0.3913 (0.0074)	0.2501 (0.0117)	0.2591 (0.0052)
T=500	0.1475 (0.0036)	0.1478 (0.0036)	0.1485 (0.0036)	0.1513 (0.0037)	0.3638 (0.0072)	0.1155 (0.0044)	0.2040 (0.0045)
	ρ=0.99						
T=100	1.3642 (0.0514)	1.6216 (0.0605)	1.3155 (0.0508)	1.3468 (0.0515)	0.5319 (0.0076)	1.3814 (0.1203)	0.5141 (0.0129)
T=200	0.7399 (0.0253)	0.8044 (0.0277)	0.7279 (0.0251)	0.7389 (0.0253)	0.4296 (0.0078)	0.7678 (0.1073)	0.3266 (0.0077)
T=500	0.2532 (0.0069)	0.2569 (0.0070)	0.2541 (0.0069)	0.2565 (0.0069)	0.3651 (0.0076)	0.3168 (0.0654)	0.2147 (0.0049)

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.2975 (0.0272)	1.2989 (0.0275)	1.3030 (0.0277)	1.3243 (0.0286)	1.6029 (0.0349)	1.3322 (0.0566)	1.3784 (0.0278)
T=200	1.1810 (0.0124)	1.1794 (0.0120)	1.1867 (0.0127)	1.1987 (0.0136)	1.5315 (0.0287)	1.1777 (0.0248)	1.2937 (0.0181)
T=500	1.1101 (0.0075)	1.1105 (0.0075)	1.1138 (0.0077)	1.1208 (0.0084)	1.6174 (0.0424)	1.0930 (0.0171)	1.3093 (0.0212)
	ρ=0.50						
T=100	1.2645 (0.0144)	1.2643 (0.0147)	1.2753 (0.0157)	1.2991 (0.0168)	1.5421 (0.0247)	1.2941 (0.0240)	1.3343 (0.0161)
T=200	1.2031 (0.0140)	1.2039 (0.0142)	1.2095 (0.0144)	1.2192 (0.0148)	1.6492 (0.0450)	1.1700 (0.0147)	1.3604 (0.0252)
T=500	1.1099 (0.0078)	1.1106 (0.0079)	1.1141 (0.0083)	1.1213 (0.0088)	1.6162 (0.0425)	1.0908 (0.0157)	1.3064 (0.0213)
	ρ=0.80						
T=100	1.2814 (0.0143)	1.2815 (0.0141)	1.2882 (0.0146)	1.3202 (0.0168)	1.5781 (0.0242)	1.3859 (0.0367)	1.3552 (0.0167)
T=200	1.1693 (0.0101)	1.1684 (0.0099)	1.1742 (0.0104)	1.1876 (0.0117)	1.5674 (0.0360)	1.1481 (0.0109)	1.2955 (0.0154)
T=500	1.1082 (0.0076)	1.1085 (0.0076)	1.1126 (0.0082)	1.1187 (0.0086)	1.6158 (0.0426)	1.0867 (0.0122)	1.3069 (0.0215)
	ρ=0.99						
T=100	1.4014 (0.0235)	1.4357 (0.0241)	1.4051 (0.0236)	1.4245 (0.0241)	1.6442 (0.0416)	1.3933 (0.0480)	1.3764 (0.0277)
T=200	1.2821 (0.0154)	1.2951 (0.0156)	1.2841 (0.0153)	1.2929 (0.0156)	1.5419 (0.0256)	1.2005 (0.0196)	1.3027 (0.0175)
T=500	1.1807 (0.0236)	1.1857 (0.0242)	1.1846 (0.0244)	1.1887 (0.0245)	1.6844 (0.0481)	1.1552 (0.0371)	1.3481 (0.0309)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.4007 (0.0171)	1.4022 (0.0161)	1.4042 (0.0172)	1.4272 (0.0215)	1.6117 (0.0263)	1.5408 (0.1302)	1.4496 (0.0217)
T=200	1.2689 (0.0155)	1.2692 (0.0158)	1.2738 (0.0157)	1.2855 (0.0162)	1.5737 (0.0243)	1.3308 (0.0842)	1.3612 (0.0172)
T=500	1.1770 (0.0131)	1.1773 (0.0131)	1.1803 (0.0132)	1.1877 (0.0143)	1.5753 (0.0283)	1.1150 (0.0088)	1.3172 (0.0168)
	ρ=0.50						
T=100	1.4022 (0.0184)	1.4059 (0.0183)	1.4070 (0.0186)	1.4255 (0.0196)	1.6490 (0.0282)	1.6550 (0.0968)	1.4823 (0.0231)
T=200	1.2832 (0.0139)	1.2835 (0.0140)	1.2872 (0.0141)	1.2969 (0.0144)	1.6013 (0.0246)	1.3515 (0.0764)	1.3825 (0.0185)
T=500	1.1576 (0.0088)	1.1582 (0.0089)	1.1600 (0.0089)	1.1634 (0.0090)	1.5436 (0.0270)	1.1125 (0.0076)	1.2884 (0.0145)
	ρ=0.80						
T=100	1.4206 (0.0222)	1.4245 (0.0224)	1.4242 (0.0223)	1.4426 (0.0227)	1.6930 (0.0325)	1.5160 (0.0584)	1.4954 (0.0231)
T=200	1.2864 (0.0269)	1.2831 (0.0254)	1.2891 (0.0269)	1.3045 (0.0303)	1.6034 (0.0384)	1.2747 (0.0302)	1.3675 (0.0237)
T=500	1.1733 (0.0107)	1.1727 (0.0105)	1.1758 (0.0108)	1.1820 (0.0114)	1.6348 (0.0435)	1.1153 (0.0075)	1.3417 (0.0228)
	ρ=0.99						
T=100	1.4718 (0.0199)	1.5054 (0.0206)	1.4852 (0.0204)	1.5134 (0.0212)	1.6644 (0.0332)	1.4521 (0.0274)	1.4802 (0.0234)
T=200	1.3123 (0.0138)	1.3224 (0.0143)	1.3183 (0.0140)	1.3365 (0.0145)	1.6158 (0.0407)	1.3670 (0.1117)	1.3750 (0.0193)
T=500	1.1916 (0.0130)	1.1934 (0.0131)	1.1967 (0.0132)	1.2029 (0.0136)	1.5966 (0.0295)	1.1138 (0.0070)	1.3327 (0.0168)

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.1768 (0.0020)	0.1778 (0.0020)	0.1803 (0.0021)	0.1939 (0.0023)	0.3653 (0.0051)	0.2333 (0.0061)	0.2331 (0.0029)
T=200	0.1207 (0.0014)	0.1210 (0.0014)	0.1224 (0.0014)	0.1280 (0.0015)	0.3535 (0.0049)	0.1914 (0.0391)	0.2048 (0.0026)
T=500	0.0718 (0.0008)	0.0718 (0.0008)	0.0722 (0.0008)	0.0735 (0.0008)	0.3409 (0.0048)	0.0925 (0.0015)	0.1766 (0.0022)
	ρ=0.50						
T=100	0.1856 (0.0022)	0.1865 (0.0021)	0.1891 (0.0022)	0.2031 (0.0025)	0.3775 (0.0054)	0.2689 (0.0097)	0.2413 (0.0032)
T=200	0.1216 (0.0014)	0.1219 (0.0014)	0.1234 (0.0014)	0.1287 (0.0015)	0.3531 (0.0050)	0.1635 (0.0039)	0.2055 (0.0027)
T=500	0.0719 (0.0008)	0.0720 (0.0008)	0.0724 (0.0008)	0.0735 (0.0008)	0.3490 (0.0050)	0.0982 (0.0020)	0.1770 (0.0023)
	ρ=0.80						
T=100	0.2003 (0.0023)	0.2042 (0.0024)	0.2029 (0.0024)	0.2146 (0.0025)	0.3738 (0.0051)	0.3252 (0.0102)	0.2403 (0.0030)
T=200	0.1268 (0.0014)	0.1275 (0.0014)	0.1285 (0.0014)	0.1333 (0.0015)	0.3353 (0.0048)	0.1982 (0.0069)	0.1971 (0.0026)
T=500	0.0733 (0.0008)	0.0733 (0.0008)	0.0737 (0.0008)	0.0747 (0.0008)	0.3413 (0.0048)	0.1108 (0.0034)	0.1759 (0.0023)
	ρ=0.99						
T=100	5.8872 (0.1915)	6.5930 (0.2162)	5.8655 (0.1915)	5.8674 (0.1915)	0.4481 (0.0057)	3.8089 (0.3557)	0.4898 (0.0186)
T=200	3.2478 (0.1078)	3.4124 (0.1123)	3.2435 (0.1078)	3.2428 (0.1078)	0.3766 (0.0050)	1.6324 (0.1588)	0.2718 (0.0038)
T=500	0.8609 (0.0329)	0.8916 (0.0340)	0.8608 (0.0329)	0.8609 (0.0329)	0.3395 (0.0047)	0.5805 (0.0686)	0.1835 (0.0023)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	0.3279 (0.0034)	0.3378 (0.0034)	0.3304 (0.0034)	0.3407 (0.0036)	0.4386 (0.0051)	0.3903 (0.0054)	0.3544 (0.0037)
T=200	0.2187 (0.0021)	0.2221 (0.0021)	0.2202 (0.0022)	0.2247 (0.0022)	0.3810 (0.0049)	0.2514 (0.0032)	0.2723 (0.0030)
T=500	0.1349 (0.0013)	0.1358 (0.0013)	0.1354 (0.0013)	0.1364 (0.0013)	0.3526 (0.0048)	0.1559 (0.0018)	0.2151 (0.0026)
	ρ=0.50						
T=100	0.3369 (0.0033)	0.3483 (0.0033)	0.3391 (0.0033)	0.3478 (0.0035)	0.4395 (0.0052)	0.3954 (0.0054)	0.3553 (0.0037)
T=200	0.2279 (0.0022)	0.2313 (0.0022)	0.2292 (0.0022)	0.2339 (0.0023)	0.3834 (0.0049)	0.2607 (0.0034)	0.2753 (0.0031)
T=500	0.1389 (0.0013)	0.1397 (0.0013)	0.1393 (0.0013)	0.1404 (0.0013)	0.3553 (0.0048)	0.1608 (0.0023)	0.2176 (0.0026)
	ρ=0.80						
T=100	0.3608 (0.0036)	0.3735 (0.0036)	0.3632 (0.0036)	0.3728 (0.0038)	0.4500 (0.0053)	0.4528 (0.0162)	0.3665 (0.0038)
T=200	0.2397 (0.0023)	0.2442 (0.0023)	0.2410 (0.0023)	0.2456 (0.0024)	0.3901 (0.0050)	0.2756 (0.0040)	0.2806 (0.0031)
T=500	0.1489 (0.0014)	0.1498 (0.0014)	0.1493 (0.0014)	0.1504 (0.0014)	0.3655 (0.0049)	0.1686 (0.0026)	0.2241 (0.0027)
	ρ=0.99						
T=100	1.3679 (0.0571)	1.6564 (0.0657)	1.3119 (0.0565)	1.3457 (0.0570)	0.5345 (0.0059)	1.2136 (0.1291)	0.5107 (0.0075)
T=200	0.7702 (0.0284)	0.8373 (0.0307)	0.7648 (0.0284)	0.7690 (0.0284)	0.4231 (0.0051)	0.5511 (0.0441)	0.3365 (0.0052)
T=500	0.2467 (0.0063)	0.2511 (0.0065)	0.2472 (0.0063)	0.2480 (0.0063)	0.3633 (0.0048)	0.2323 (0.0080)	0.2291 (0.0026)

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-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.2637 (0.0082)	1.2660 (0.0081)	1.2696 (0.0084)	1.2887 (0.0089)	1.5793 (0.0246)	1.3656 (0.0249)	1.3679 (0.0129)
T=200	1.1825 (0.0053)	1.1833 (0.0053)	1.1858 (0.0055)	1.1942 (0.0057)	1.5627 (0.0235)	1.2622 (0.0224)	1.3261 (0.0125)
T=500	1.1094 (0.0032)	1.1096 (0.0032)	1.1101 (0.0032)	1.1119 (0.0033)	1.5546 (0.0240)	1.1515 (0.0085)	1.2784 (0.0114)
	ρ=0.50						
T=100	1.2760 (0.0086)	1.2751 (0.0084)	1.2830 (0.0090)	1.3093 (0.0100)	1.5948 (0.0212)	1.4155 (0.0228)	1.3831 (0.0127)
T=200	1.1867 (0.0063)	1.1868 (0.0063)	1.1894 (0.0064)	1.2000 (0.0071)	1.5554 (0.0212)	1.2433 (0.0098)	1.3298 (0.0124)
T=500	1.1161 (0.0032)	1.1159 (0.0032)	1.1170 (0.0033)	1.1192 (0.0034)	1.5943 (0.0245)	1.1564 (0.0056)	1.3054 (0.0125)
	ρ=0.80						
T=100	1.3009 (0.0135)	1.3026 (0.0130)	1.3070 (0.0137)	1.3339 (0.0149)	1.6125 (0.0344)	1.4257 (0.0170)	1.4093 (0.0215)
T=200	1.1894 (0.0128)	1.1903 (0.0131)	1.1926 (0.0130)	1.2047 (0.0141)	1.5493 (0.0251)	1.2723 (0.0180)	1.3286 (0.0161)
T=500	1.1116 (0.0035)	1.1115 (0.0034)	1.1123 (0.0035)	1.1139 (0.0036)	1.5399 (0.0186)	1.1672 (0.0114)	1.2843 (0.0098)
	ρ=0.99						
T=100	1.4028 (0.0115)	1.4363 (0.0124)	1.4067 (0.0118)	1.4220 (0.0120)	1.6146 (0.0208)	1.4992 (0.0517)	1.3924 (0.0123)
T=200	1.2848 (0.0091)	1.2972 (0.0093)	1.2867 (0.0091)	1.2917 (0.0092)	1.5963 (0.0224)	1.2675 (0.0117)	1.3379 (0.0122)
T=500	1.1564 (0.0044)	1.1596 (0.0045)	1.1574 (0.0045)	1.1588 (0.0045)	1.5581 (0.0231)	1.1752 (0.0151)	1.2882 (0.0115)

Panel B – Three lags in the estimated equation

	TVC-MA	TVC-MS	TVC-Π	TVC-π	OLS	BP	BP-MA
	ρ=0						
T=100	1.4312 (0.0188)	1.4333 (0.0168)	1.4359 (0.0191)	1.4636 (0.0230)	1.6667 (0.0306)	1.5028 (0.0241)	1.5340 (0.0256)
T=200	1.2653 (0.0071)	1.2653 (0.0068)	1.2681 (0.0072)	1.2800 (0.0082)	1.5577 (0.0173)	1.3186 (0.0107)	1.3912 (0.0116)
T=500	1.1713 (0.0051)	1.1715 (0.0051)	1.1723 (0.0052)	1.1737 (0.0052)	1.5745 (0.0226)	1.2156 (0.0077)	1.3453 (0.0135)
	ρ=0.50						
T=100	1.4214 (0.0137)	1.4247 (0.0131)	1.4252 (0.0138)	1.4476 (0.0152)	1.6373 (0.0229)	1.4982 (0.0197)	1.5109 (0.0189)
T=200	1.2954 (0.0128)	1.2961 (0.0127)	1.2993 (0.0136)	1.3114 (0.0141)	1.6280 (0.0247)	1.3697 (0.0214)	1.4369 (0.0175)
T=500	1.1687 (0.0048)	1.1689 (0.0048)	1.1695 (0.0048)	1.1711 (0.0049)	1.5631 (0.0197)	1.2145 (0.0071)	1.3402 (0.0115)
	ρ=0.80						
T=100	1.4303 (0.0129)	1.4342 (0.0125)	1.4353 (0.0131)	1.4614 (0.0146)	1.6616 (0.0246)	1.5534 (0.0261)	1.5281 (0.0187)
T=200	1.2803 (0.0079)	1.2816 (0.0078)	1.2829 (0.0079)	1.2941 (0.0083)	1.5949 (0.0208)	1.3301 (0.0099)	1.4126 (0.0131)
T=500	1.1652 (0.0044)	1.1648 (0.0043)	1.1659 (0.0044)	1.1681 (0.0044)	1.5584 (0.0233)	1.2240 (0.0120)	1.3402 (0.0134)
	ρ=0.99						
T=100	1.5138 (0.0153)	1.5417 (0.0162)	1.5246 (0.0156)	1.5664 (0.0167)	1.6848 (0.0276)	1.5574 (0.0300)	1.5380 (0.0200)
T=200	1.3405 (0.0110)	1.3533 (0.0113)	1.3469 (0.0112)	1.3624 (0.0116)	1.6228 (0.0267)	1.3507 (0.0122)	1.4279 (0.0152)
T=500	1.1838 (0.0071)	1.1851 (0.0071)	1.1853 (0.0071)	1.1889 (0.0072)	1.5842 (0.0276)	1.2445 (0.0173)	1.3475 (0.0109)

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

	TVC-MA	TVC-MS	TVC-Π	TVC-π	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%