A Monte Carlo Study of Time Varying Coefficient (TVC) Estimation*

Stephen G. Hall†
University of Leicester
Bank of Greece

Heather D. Gibson
Bank of Greece

G.S. Tavlas
Bank of Greece
University of Leicester

Mike G. Tsionas
Lancaster University
Athens University of Economics and Business

Abstract

A number of recent papers have proposed a time-varying-coefficient (TVC) procedure that, in theory, yields consistent parameter estimates in the presence of measurement errors, omitted variables, incorrect functional forms, and simultaneity. The key element of the procedure is the selection of a set of driver variables. With an ideal driver set the procedure is both consistent and efficient. However, in practice it is not possible to know if a perfect driver set exists. We construct a number of Monte Carlo experiments to examine the performance of the methodology under (i) clearly-defined conditions and (ii) a range of model misspecifications. We also propose a new Bayesian search technique for the set of driver variables underlying the TVC methodology. Experiments are performed to allow for incorrectly specified functional form, omitted variables, measurement errors, unknown nonlinearity and endogeneity. In all cases except the last, the technique works well in reasonably small samples.

Keywords: Time-varying coefficients, specification errors, Monte Carlo

JEL classifications: C130, C190, C220

* We are grateful to Fredj Jawadi, and two referees for helpful comments on an earlier draft.
† Corresponding author: Professor Stephen G. Hall, Head of Economics, School of Business, University of Leicester, University Road, Leicester, LE1 7RH, UK; tel.: +44 (0)116 2522827; email: s.g.hall@le.ac.uk.

1. Introduction

A series of papers have proposed the use of time-varying coefficient (TVC) models to uncover the bias-free estimates of a set of model coefficients in the presence of omitted variables, measurement error and an unknown true functional form.[1] There have also been a reasonably-large number of successful applications of the technique.[2] However, it is difficult to establish the usefulness of a technique strictly through applications since we can never be certain of the accuracy of the results. This paper attempts to bridge the gap between the asymptotic theoretical results of the theoretical papers and the apparently good performance of the applied papers by constructing a set of Monte Carlo experiments to examine (1) how well the technique performs under clearly-defined conditions and (2) the limits on the technique's ability to perform successfully under a broad range of model misspecifications.

The technique is motivated by an important theorem that was first proved by Swamy and Mehta (1975) and has recently been confirmed by Granger (2008) who quoted a proof that he attributed to Hal White. This theorem states that any nonlinear function may be exactly represented by a linear relationship with time-varying parameters. The importance of this theorem is that it allows us to capture an unknown true functional form in this framework. The parameters of this time-varying-coefficient model are, of course, not consistent estimates of the true functional form since they will be contaminated by the usual biases due to omitted variables, measurement error and simultaneity. The technique being investigated here allows us, in principal, to decompose the TVCs into two components; we associate the first component with the true nonlinear structure, which we interpret as the derivative of the dependent variable with respect to each of the independent variables in the unknown, nonlinear, true function; we associate the second component with the biases emanating from misspecification, and which we then

[1] Swamy, Chang, Mehta and Tavlas (2003), Swamy, Tavlas, Hall and Hondroyiannis (2010), Swamy, Hall and Tavlas (2012), Swamy, Hall and Tavlas, (2014), Swamy, Mehta, Tavlas and Hall (2014).

[2] Empirical applications include Hall, Hondroyiannis, Swamy, and Tavlas (2009), Hall, Hondroyiannis, Swamy and Tavlas (2010), Tavlas, Swamy, Hall and Kenjegaliev (2013), Hondroyiannis, Kenjegaliev, Hall, Swamy and Tavlas (2013), Kenjegaliev, Hall, Tavlas and Swamy (2013), Hall, Swamy, and Tavlas (2017).

remove from the TVC to give us our consistent estimates. Potentially, this technique offers an interesting way forward in dealing with model misspecification. It has generally been applied in a time series setting but it can equally well be interpreted as a cross section[3] or panel estimation technique.

The remainder of this paper is structured as follows. Section 2 outlines the basic (TVC) theoretical framework. Section 3 discusses some computational issues associated with estimating the model. Section 4 reports on a series of Monte Carlo experiments. Section 5 concludes. An Appendix provides details on the computational methods used in the Monte Carlo simulations.

2. The Theoretical Framework

We follow Swamy, Tavlas, Hall and Hondroyiannis (2010) who set the groundwork for uncovering causal economic laws. We assume

$$y_t^* = f(\mathbf{x}_t^*, \mathbf{e}_t^*) \ , \tag{1}$$

where $\mathbf{x}_t^*, \mathbf{e}_t^*$ are the true determinants of y_t. Alternatively we can represent this relationship by

$$y_t^* = \alpha_{0t} + \boldsymbol{\alpha}_{1t}' \mathbf{x}_t^* + \boldsymbol{\alpha}_{2t}' \mathbf{e}_t^* \tag{2}$$

We have the auxiliary equations:

$$\mathbf{e}_t^* = \boldsymbol{\Psi}_t \mathbf{x}_t^* + \mathbf{v}_t^* \tag{3}$$

Substituting (3) into (2) gives:

$$y_t^* = \left(\alpha_{0t} + \boldsymbol{\alpha}_{2t}' \mathbf{v}_t^*\right) + \left(\boldsymbol{\alpha}_{1t}' + \boldsymbol{\alpha}_{2t}' \boldsymbol{\Psi}_t\right) \mathbf{x}_t^* \ . \tag{4}$$

To deal with errors in variables we assume:

$$y_t = y_t^* + v_t, \tag{5}$$

[3] Time varying coefficients are meaningless in a cross section setting; in such a setting the coefficients vary across the individual units in the cross section. We simply re-interpret the t-subscript as i-subscripts.

$$\mathbf{x}_t = \mathbf{x}_t^* + \mathbf{w}_t. \tag{6}$$

Substituting into (4), we obtain:

$$\begin{aligned} y_t &= \left(\alpha_{0t} + \boldsymbol{\alpha}_{2t}'\mathbf{v}_t^* + v_t\right) + \left(\boldsymbol{\alpha}_{1t}' + \boldsymbol{\alpha}_{2t}'\boldsymbol{\Psi}_t\right)\left(I - D_{wt}D_{xt}^{-1}\right)\mathbf{x}_t = \\ &\beta_{0t} + \boldsymbol{\beta}_{xt}'\mathbf{x}_t \equiv \mathbf{x}_{et}'\boldsymbol{\beta}_t, \end{aligned} \tag{7}$$

where $\beta_{0t} = \alpha_{0t} + \boldsymbol{\alpha}_{2t}'\mathbf{v}_t^* + v_t$, $\boldsymbol{\beta}_{xt}' = \left(\boldsymbol{\alpha}_{1t}' + \boldsymbol{\alpha}_{2t}'\boldsymbol{\Psi}_t\right)\left(I - D_{wt}D_{xt}^{-1}\right)$, $\mathbf{x}_{et} = \left[1, \mathbf{x}_t'\right]'$, $\boldsymbol{\beta}_t = \left[\beta_{0t}, \boldsymbol{\beta}_{xt}'\right]'$ and D_{wt}, D_{xt} are diagonal containing $\mathbf{w}_t$ and $\mathbf{x}_t$, respectively, along the diagonal. Finally, we assume there exists a vector $\mathbf{z}_t$ of drivers such that

$$\boldsymbol{\beta}_t = \boldsymbol{\Pi}\mathbf{z}_{et} + \boldsymbol{\varepsilon}_t, \tag{8}$$

where $\mathbf{z}_{et} = \left[1, \mathbf{z}_t'\right]'$. Under the assumption:

$$\boldsymbol{\varepsilon}_t \sim \mathrm{IN}\left(\mathbf{0}, \boldsymbol{\Sigma}\right). \tag{9}$$

It is straightforward to obtain the following

$$\mathrm{E}\left(y_t \mid \mathbf{x}_t, \mathbf{z}_t\right) = \left(\mathbf{z}_{et}' \otimes \mathbf{x}_{et}'\right)\mathrm{vec}\left(\boldsymbol{\Pi}\right) \tag{10}$$

In matrix notation we have

$$\mathrm{E}\left(\mathbf{y} \mid \mathbf{X}_z\right) = \mathbf{X}_z\mathrm{vec}\left(\boldsymbol{\Pi}\right), \mathrm{cov}\left(\mathbf{y} \mid \mathbf{X}_z\right) = \sigma_a^2\boldsymbol{\Omega} \tag{11}$$

where

$$X_z = \left(\mathbf{z}_{e1} \otimes \mathbf{x}_{e1}, \ldots, \mathbf{z}_{eT} \otimes \mathbf{x}_{eT}\right)', \ \sigma_a^2\boldsymbol{\Omega} = \mathbf{D}_x\left(I_T \otimes \sigma_a^2\boldsymbol{\Sigma}\right)\mathbf{D}_x',$$

and $\mathbf{D}_x = \mathrm{diag}\left[\mathbf{x}_{e1}', \ldots, \mathbf{x}_{eT}'\right]$. A restrictive version of (7) is

$$y_t = \mathbf{x}_{et}'\boldsymbol{\beta}_t + u_t, \ u_t \sim \mathrm{IN}\left(0, \sigma_u^2\right), \tag{12}$$

where β_{0t} is redefined as $\alpha_{0t} + v_t$ and β_t is independent of $\boldsymbol{\alpha}_{2t}'\mathbf{v}_t^* = u_t$

and (8) as

$$\boldsymbol{\beta}_t = \boldsymbol{\Pi}_r \mathbf{z}_{et} + \boldsymbol{\varepsilon}_{rt} \tag{13}$$

where the first row of Π_r post-multiplied by $\mathbf{z}_{et}$ does not contain the mean of $\boldsymbol{\alpha}'_{2t}\mathbf{v}^*_t$, and the first element of $\boldsymbol{\varepsilon}_{rt}$ is independent of $\boldsymbol{\alpha}'_{2t}\mathbf{v}^*_t$. The error vectors ε_t and $\boldsymbol{\varepsilon}_{rt}$ are introduced in (8) and (13). Substituting we have

$$y_t = \mathbf{x}'_{et}\boldsymbol{\Pi}_r \mathbf{z}_{et} + u_t + \mathbf{x}'_{et}\boldsymbol{\varepsilon}_{rt} = \left(z'_{et} \otimes x'_{et}\right)\mathrm{vec}\left(\boldsymbol{\Pi}_r\right) + u_t + \mathbf{x}'_{et}\boldsymbol{\varepsilon}_{rt}. \tag{14}$$

Defining $z_{et} \otimes x_{et} = \mathbf{X}_t$, and $\boldsymbol{\pi}_r = \mathrm{vec}\left(\boldsymbol{\Pi}_r\right)$, we have:

$$y_t = \mathbf{X}'_t\boldsymbol{\pi}_r + u_t + \mathbf{x}'_{et}\boldsymbol{\varepsilon}_{rt}. \tag{15}$$

By assumption, $\mathrm{E}\left(u_t + \mathbf{x}'_{et}\boldsymbol{\varepsilon}_{rt}\right) = 0$ and $\mathrm{var}\left(u_t + \mathbf{x}'_{et}\boldsymbol{\varepsilon}_{rt}\right) = \sigma_u^2 + \mathbf{x}'_{et}\sigma_r^2\boldsymbol{\Sigma}_r\mathbf{x}_{et}$.

3 Computational aspects

Under a normality assumption in both u_t and $\boldsymbol{\varepsilon}_{rt}$ the likelihood function is

$$L(\theta;Y) \propto \prod_{t=1}^{T}\left(\sigma_u^2 + \mathbf{x}'_{et}\boldsymbol{\Sigma}_r\mathbf{x}_{et}\right)^{-1/2} \exp\left\{-\tfrac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - \mathbf{X}'_t\boldsymbol{\pi}_r\right)^2}{\left(\sigma_u^2 + \mathbf{x}'_{et}\boldsymbol{\Sigma}_e\mathbf{x}_{et}\right)}\right\} \tag{16}$$

where the parameter vector is $\theta = \left[\boldsymbol{\pi}'_r, \sigma_r, \sigma_u, \mathrm{vech}(\boldsymbol{\Sigma}_r)'\right]'$. Coupled with a prior, $p\left(\theta\right)$, by Bayes' theorem we get the kernel posterior distribution:

$$p(\theta \mid Y) \propto L(\theta;Y)\,p(\theta),\ \theta \in \Theta \tag{17}$$

We assume a standard non-informative prior:

$$p(\Sigma_r) \propto |\,\Sigma_r\,|^{-(d+1)/2}$$

where d is the dimensionality of Σ_r. The prior of π will be detailed below.

Markov Chain Monte Carlo (MCMC) techniques can be used to obtain a sample $\{\theta^{(s)}, s = 1,...,S\}$ that converges in distribution to the posterior $p(\theta \mid Y)$. One efficient MCMC strategy is the following.

(i) Obtain $\boldsymbol{\pi}_r$ from its conditional distribution:

(ii) $\boldsymbol{\pi}_r \mid \sigma_r, \sigma_u, \boldsymbol{\Sigma}_r, \mathbf{Y} \sim \mathrm{N}\left(\hat{\boldsymbol{\pi}}_r, \mathbf{V}\right)$ (18)

where $\hat{\boldsymbol{\pi}}_r = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$, $\mathbf{V} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}$, $\boldsymbol{\Omega} = \mathrm{diag}\left(\sigma_u^2 + \mathbf{x}_{et}'\boldsymbol{\Sigma}_r\mathbf{x}_{et}, t = 1,...,T\right)$.

(iii) Reparametrize $\boldsymbol{\Sigma}_r$ using $\mathbf{C}$ where $\boldsymbol{\Sigma}_r = \mathbf{C}_r'\mathbf{C}_r$, $\sigma_u \propto \exp\left(c_0\right)$ and $\sigma_r \propto \exp(C_{00})$. Assuming that different non-zero elements of $\mathbf{C}_r$ are $c_1,...,c_p$ the new parameter vector is $\boldsymbol{\pi}_r$ and $\mathbf{c} = [c_0, c_{00}, c_1, ..., c_p]' \in \mathbb{R}^{p+2}$. Drawings from the conditional posterior distribution of $\mathbf{c} \mid \boldsymbol{\pi}_r, \mathbf{Y}$ can be realized using the Girolami and Calderhead (2012) Metropolis Adjusted Langevin Diffusion method described in the Appendix.

If we define $\hat{\boldsymbol{\beta}}_t = \left(\mathbf{x}_t\mathbf{x}_t' + \sigma^2\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\mathbf{x}_t y_t + \sigma^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}\mathbf{z}_t\right)$ and $\mathbf{V}_{\beta t} = \left(\mathbf{x}_t\mathbf{x}_t' + \sigma^2\boldsymbol{\Sigma}^{-1}\right)^{-1}$ we obtain:

$$\boldsymbol{\beta}_t \mid \sigma, \boldsymbol{\Sigma}, Y \sim \mathrm{N}\left(\hat{\boldsymbol{\beta}}_t, \mathbf{V}_{\beta t}\right). \quad (19)$$

In this form we can avoid a possibly inefficient Gibbs sampler which relies on drawing $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}$ from (13), $\{\boldsymbol{\beta}_t, t = 1,...,T\}$ from (19) and σ from (12).

Selecting the drivers

Suppose we have (12) and instead of (13) we have

$$\boldsymbol{\beta}_t = \boldsymbol{\Pi}_m\mathbf{z}_t^{(m)} + \boldsymbol{\varepsilon}_t^{(m)}, \forall m = 1,...,M, \quad (20)$$

where $\mathbf{z}_t^{(m)}$ is a potential set of drivers (subset) from a universe $Z = \left\{\mathbf{z}_{t1},...,\mathbf{z}_{tG}\right\}$. Equations (12) and (20) define different models indexed by m. As searching through all possible combinations of variables in Z is infeasible, we follow the Stochastic Search Variable Selection (SSVS) approach of George, Sun and Ni (2008).[4] The SSVS involves a specific prior of the form:

$$\boldsymbol{\pi} \mid \boldsymbol{\delta} \sim \mathrm{N}(\mathbf{0}, \mathbf{D}) \quad (21)$$

[4] See also Jochmann, Koop and Strachan (2010).

where $\boldsymbol{\delta}$ is a vector of unknown parameters and its elements can be $\delta_j \in \{0,1\}$.

Also $\mathbf{D} = \text{diag}\left[d_1^2, ..., d_G^2 \right]$:

$$d_j^2 = \underline{\kappa}_{0j}^2, \text{ if } \delta_j = 0, \text{ and } d_j^2 = \underline{\kappa}_{1j}^2, \text{ if } \delta_j = 1. \quad (22)$$

The prior implies a mixture of two normals:

$$\pi_j | \delta_j \sim (1-\delta_j) N(0, \underline{\kappa}_{0j}^2) + \delta_j N(0, \underline{\kappa}_{1j}^2) \quad (23)$$

If $\underline{\kappa}_{0j}$ is "small" and $\underline{\kappa}_{1j}$ is "large", then, when $\delta_j = 0$ chances are that variable j will be excluded from the model while if $\delta_j = 1$ chances are that variable j will be included in the model. The prior for the indicator parameter $\boldsymbol{\delta}$ is:

$$P(\delta_j = 1) = \underline{q}_j, \; P(\delta_j = 0) = 1 - \underline{q}_j \quad (24)$$

and we set $\underline{q}_j = \frac{1}{2}$. For $\underline{\kappa}_{0j}$ and $\underline{\kappa}_{1j}$, George, Sun and Ni (2008) propose a semi-automatic procedure based on $\underline{\kappa}_{0j}^2 = c_0 \hat{v}\left(\pi_j\right)$ and $\underline{\kappa}_{1j}^2 = c_1 \hat{v}\left(\pi_j\right)$ for $c_0 = \frac{1}{10}$, $c_1 = 10$ and $\hat{v}\left(\pi_j\right)$ is any preliminary estimate of the variance of π_j.

For the elements of $\mathbf{c}$ we follow a similar approach. If c_j corresponds to a diagonal element it is always included in the model. If not, we use a mixture-of-normals SSVS approach as above.

4 Monte Carlo results

In all cases below $\gamma_0 = \gamma_1 = \gamma_2 = 0.1$. The number of Monte Carlo simulations is set to 10,000. All $\varepsilon_{tj} \sim iidN(0,1)$. In case IV, we set $\sigma_\varepsilon = 0.3$.

4.1 Model I: Incorrect Functional Form

The true model is $y_t = \gamma_0 + \gamma_1 x_t + \frac{1}{2}\gamma_2 x_t^2, t = 1, ..., T$ and we have omitted the nonlinear term. The driver is $z_t = \alpha x_t + \varepsilon_t$. We have $\varepsilon_t \sim \text{iid}N(0,1)$

and $x_t \sim \text{iidN}(1,1)$. In this case the correlation between z_t and x_t^2 is $\rho = \frac{3\alpha}{\sqrt{3\left(3\alpha^2+1\right)}}$.

If the correlation were equal to 1, then this would be a perfect driver as it exactly recreates the missing quadratic term. The estimation procedure would then be unbiased and efficient. If the correlation were zero, then z_t would contain no information about the missing nonlinearity. We are, therefore, interested in varying this correlation and seeing how low the correlation can fall before the estimator ceases to be useful.

In this case the true effect is $\gamma_1 + \gamma_2 x_t$, that is, the derivative of y with respect to x. There are no omitted variables or other misspecifications other than the nonlinearity so the set S_2 is empty and the estimate of the derivative is given by $\gamma_1 + \gamma_2 x_t = \beta_1 - \varepsilon_t$.

Table 1 gives the results of this set of Monte Carlo experiments for sample sizes of 50, 100, 200 and 1000. When the correlation between z and x is very high, then even for small samples the bias is very small and the standard deviation of the results is also small at around 1%. As the sample size grows, both the bias and the standard deviation fall, and the estimator is clearly consistent and efficient. As we look across the table, where the correlation between the driver and the true x variable falls the estimation procedure still does very well until the correlation falls to about 0.5; at that point the bias and the standard error begins to rise quite substantially. This happens even more clearly with the very large sample size of T=1000 where both the bias and standard error are very small until the correlation falls below 0.5.

4.2 Model II: Omitted Variables

The second model focuses on omitted variables. The true model is $y_t = \gamma_0 + \gamma_1 x_{t1} + \gamma_2 x_{t2}$. The x_{t1}, x_{t2} are correlated: $x_{t2} = \gamma x_{t1} + \xi_t$, $\xi_t, x_{t1} \sim \text{iidN}(0,1)$.

The squared correlation between the two variables is $\rho_{12}^2 = \frac{\gamma^2}{\gamma^2 + 1}$. We set $\gamma = 2$ so that this is 0.80.

We estimate the TVC model $y_t = \beta_{0t} + \beta_{1t} x_{t1}$ and again use a driver $z_t = \alpha x_{t2} + \varepsilon_t$ and we see how well the estimator performs as the correlation between z_t and x_{t2} falls. The correlation between z_t and x_{t2} is $\rho = \frac{\alpha}{\sqrt{\alpha^2 + 1}}$.

In this case, the true effect is γ_1 and the bias free estimate is $\beta_{1t} - \pi_1 z_t - e_t$.

The results of this experiment are given in Table 2. The results show a similar picture to case 1 above. Both the bias and the standard deviation clearly decrease as the sample size increases. Even for the smallest sample size both the bias and the standard deviation are quite small while the correlation between the driver and the misspecification is above 0.5. Again, as the correlation falls below 0.5 the bias and standard deviation rise quite quickly.

4.3 Model III: Measurement Error

The third model deals with measurement error, so we generate data from $y_t = \gamma_0 + \gamma_1 x_t$ then create $y_t^* = y_t + \varepsilon_{t1}$ and $x_t^* = x_t + \varepsilon_{t2}$ then we estimate the TVC model $y_t^* = \beta_{0t} + \beta_{1t} x_t^*$ and use two z's as drivers $z_{t1} = \alpha_1 \varepsilon_{t1} + \varepsilon_{t3}$ and $z_{t2} = \alpha_2 \varepsilon_{t2} + \varepsilon_{t4}$ and again see how things change as α gets bigger.

The results of this experiment are reported in Table 3. The results are entirely consistent with the results in the earlier two cases. The technique is clearly consistent, as the sample rises the bias falls considerably. Even for a small sample the bias is quite low for correlations between the driver and the measurement error which is 0.5 or above.

4.4 Detecting Irrelevant Drivers

Next, we examine whether the SSVS[5] procedure, which we have not applied so far, can correctly identify the drivers z_{t1}, z_{t2}. To this end, we construct ten other drivers, say $z_{t2},...,z_{t,12}$ from a multivariate normal distribution with zero means and equal correlations of 0.70. In Table 4 we report the equivalent of Table 3 plus the proportion of cases, say Π^*, in which SSVS has correctly excluded $z_{t2},...,z_{t,12}$ from the set of possible drivers[6].

There is again a remarkable cut off at the correlation level of 0.5. Above this level the true driver set is correctly identified in around 60% of cases and for the largest sample in over 90% of cases, even for small samples. Once the correlation falls below 0.5, however, the proportion of correct identifications falls dramatically. An obvious conclusion here is that when we have drivers that are effective enough so that we will get reasonably good parameter estimates, the SSVS algorithm is very effective at detecting them.

4.5 Model IV: A More Complex Nonlinearity

The true model is $y_t = \gamma_0 + \gamma_1 x_t + \exp\left(-\delta x_t^2\right) + \varepsilon_t, t = 1,...,T$ and we have omitted the nonlinear term. The drivers form a Fourier basis $\left\{\cos(jx_t), \sin(jx_t), j = 1,...,J\right\}$ after transforming all series to lie in $\left(-\pi, \pi\right)$. We have $\varepsilon_t \sim \text{iidN}(0,1)$ and $x_t \sim \text{iidN}(0,1)$ ordered from smallest to largest. The drivers, that is powers of x_t are selected through the SSVS procedure. We set the maximum value of J to 10.

[5] An alternative to using the SSVS procedure would be the LASSO prior. The procedures are similar in terms of timing and purpose. There is some evidence that both perform well (Pavlou, Ambler, Seaman, De Lorio and Omar(2016)) and in a similar manner but further work is needed in this area.

[6] There is an issue here as to whether we need to start from a superset of drivers which includes the true ones. Clearly, if we do this then this is an ideal situation and the Monte Carlo tells us how well the procedure performs. However, from a theoretical point of view what we need is that the superset includes variables be highly correlated with the true drivers. In a data rich environment this would not be a strong restriction. Bai and Ng (2010) prove that common factors which drive all the variables in a system are valid instrumental variables. By the same reasoning, we could construct a set of factors from a large set of variables which would work well as drivers.

We again estimate the TVC model $y_t = \beta_{0t} + \beta_{1t} x_{t1}$ and this time the derivative of y with respect to x is $\gamma_1 - 2\delta x_t \exp(-\delta x_t^2)$. Our estimate of this is again given by $\beta_{1t} - e_t$

The results for this exercise are given in Table 5. In this case for δ in the range 0.1 to 5 the bias remains very small, as does the standard deviation. There is also a noticeable reduction in both bias and standard deviation as the sample size increases.

4.6 Model V: An Endogeneity Experiment

In this experiment we have: $y_t = \gamma_1 + \gamma_2 x_{t1} + \gamma_3 x_{t2} + u_t$. The correlation between u_t and x_{tj} is 0.80 (j = 1,2) so that endogeneity is quite strong in this model. Our drivers are four variables $z_{t5},...,z_{t8}$ *orthogonal* to the error u_t and four drivers $z_{t1},...,z_{t4}$ which are *correlated* with the error u_t but they are orthogonal to each other as well as orthogonal to the other four drivers. The degree of correlation between the drivers and u_t is ρ. We are interested in Π^*, the proportion of cases where *all* the drivers $z_{t1},...,z_{t4}$ are included in the model *and* the drivers $z_{t5},...,z_{t8}$ are *all* excluded. Of course, we do not force the correct drivers in final estimation.

The results for this experiment are given in Table 6. Here the table has some rather different results than the earlier tables. The bias remains quite high, even for quite high correlations for the sample size of 50 or 100. It is only for much larger sample sizes that the bias becomes negligible. For larger sample sizes the bias remains small again for correlations above 0.5 and the SSVS selection procedure works reasonably well.

6. Conclusions

This paper has investigated the performance of the TVC estimation procedure in a Monte Carlo setting. The key element of TVC estimation is the identification and selection of a set of driver variables. With an ideal driver set, it is straightforward to show that the procedure is both consistent and efficient.

However, in practice it is not possible to know if we have a perfect driver set. Therefore, we need to know how the procedure performs when the driver set is less than perfect. In this paper, we dealt with this issue in a Monte Carlo setting.

We construct a number of Monte Carlo experiments to examine the performance of the methodology under (i) clearly-defined conditions and (ii) a range of model misspecifications. We also propose a new Bayesian search technique for the set of driver variables underlying the TVC methodology. Experiments are performed to allow for incorrectly specified functional form, omitted variables, measurement errors, unknown nonlinearity and endogeneity. Our broad conclusion is that, even for relatively small samples, the technique works well so long as the correlation between the driver set and the misspecification in the model is greater than about 0.5. Both the bias and the efficiency of the estimators also improve as the sample size grows, but again a correlation of over 0.5 seems to be required. The only caveat to this result is that if we are considering strong simultaneity bias; in that case the sample size needs to be quite large (over 500) before the technique works reasonably well. Finally, we find that the SSVS technique also seems to perform well in finding an appropriate driver set from a much larger set of possible drivers.

APPENDIX

This Appendix provides details on the computational methods used in the Monte Carlo simulations. Following Girolami and Calderhead (2011) we utilize Metropolis-adjusted Langevin and Hamiltonian Monte Carlo sampling methods defined on the Riemann manifold, since we are sampling from target densities with high dimensions that exhibit strong degrees of correlation. Consider the Langevin diffusion:

$$d\boldsymbol{\theta}(t) = \tfrac{1}{2}\nabla \log p\left(\boldsymbol{\theta}(t);\mathbf{Y}\right)dt + d\mathbf{B}(t),$$

where $\mathbf{B}$ denotes the D-dimensional Brownian motion. The first-order Euler discretization provides the following candidate generation mechanism:

$$\boldsymbol{\theta}^{*} = \boldsymbol{\theta}^{o} + \tfrac{1}{2}\varepsilon^{2}\nabla \log p\left(\boldsymbol{\theta}^{o};\mathbf{Y}\right) + \varepsilon\mathbf{z},$$

where $\mathbf{z} \sim N_{D}(\mathbf{0},\mathbf{I})$, and $\varepsilon > 0$ is the integration step size. Since the discretization induces an unavoidable error in approximation of the posterior, a ***Metropolis step*** is used, where the proposal density is

$$q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{o}\right) = N_{D}\left(\boldsymbol{\theta}^{o} + \tfrac{1}{2}\varepsilon^{2}\nabla \log p\left(\boldsymbol{\theta}^{o};\mathbf{Y}\right),\ \varepsilon^{2}\mathbf{I}\right),$$

with acceptance probability $a\left(\boldsymbol{\theta}^{o},\boldsymbol{\theta}^{*}\right) = \min\left\{1,\ \dfrac{p\left(\boldsymbol{\theta}^{*} \mid \mathbf{Y}\right)q\left(\boldsymbol{\theta}^{o} \mid \boldsymbol{\theta}^{*}\right)}{p\left(\boldsymbol{\theta}^{o} \mid \mathbf{Y}\right)q\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{o}\right)}\right\}$. Here $\mathbf{Y}$ denotes the available data. The Brownian motion of the Riemann manifold is given by:

$$d\tilde{\mathbf{B}}_{i}(t) = \left|\mathbf{G}\left(\boldsymbol{\theta}(t)\right)\right|^{-1/2}\sum_{j=1}^{d}\frac{\partial}{\partial\boldsymbol{\theta}_{j}}\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)_{ij}\left|\mathbf{G}\left(\boldsymbol{\theta}(t)\right)\right|^{1/2}\right]dt + \left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)}d\mathbf{B}(t)\right]_{i},$$

for $i = 1,...,D$.

The discrete form of the above stochastic differential equations is:

$$\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_i^o + \tfrac{1}{2}\varepsilon^2\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\nabla \log p\left(\boldsymbol{\theta}^o;\mathbf{Y}\right)\right]_i - \varepsilon^2\sum_{j=1}^{d}\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\frac{\partial \mathbf{G}\left(\boldsymbol{\theta}^o\right)}{\partial \boldsymbol{\theta}_j}\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\right]_{ij} +$$
$$\varepsilon^2\sum_{j=1}^{d}\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)_{ij} tr\left[\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\frac{\partial \mathbf{G}\left(\boldsymbol{\theta}^o\right)}{\partial \boldsymbol{\theta}_j}\right] + \varepsilon\left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)}\mathbf{z}\right]_i \quad \square$$
$$\boldsymbol{\mu}\left(\boldsymbol{\theta}^o,\varepsilon\right)_i + \varepsilon\left[\sqrt{\mathbf{G}^{-1}\left(\boldsymbol{\theta}(t)\right)}\mathbf{z}\right]_i .$$

The *proposal density* is $\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o \sim N_d\left(\boldsymbol{\mu}\left(\boldsymbol{\theta}^o,\varepsilon\right),\ \varepsilon^2\mathbf{G}^{-1}\left(\boldsymbol{\theta}^o\right)\right)$ and the acceptance probability has the standard *Metropolis form*:

$$a\left(\boldsymbol{\theta}^o,\boldsymbol{\theta}^*\right) = \min\left\{1,\ \frac{p\left(\boldsymbol{\theta}^* \mid \mathbf{Y}\right)q\left(\boldsymbol{\theta}^o \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^o \mid \mathbf{Y}\right)q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^o\right)}\right\}.$$

The gradient and the Hessian are computed using analytic derivatives. All computations are performed in Fortran 77 making extensive use of IMSL subroutines.

The Metropolis-Hastings procedure we use is a simple random walk whose candidate generating density is a multivariate Student-*t* distribution with 5 degrees of freedom[7] and covariance equal to a scaled version of the covariance obtained from the Langevin Diffusion MCMC. The scale parameter is adjusted so that approximately 25% of the draws are accepted.

[7] This guarantees the existence of moments up to order four.

References

Bai, J., & Ng, S. (2010). Instrumental Variable Estimation in a data Rich Environment. *Econometric Theory, 26*, 1577-1606.

George, E., Sun, D., & Ni S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics, 142,* 553-580.

Girolami, M., & Calderhead B. (2012). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society*: Series B (Statistical Methodology) *73*(2), 123–214.

Granger, C. W. J. (2008). Nonlinear Models: Where Do We Go Next – Time-varying Parameter Models? *Studies in Nonlinear Dynamics and Econometrics*, *12*(3), 1-9.

Hall, S. G., Hondroyiannis, G., Swamy, P. A. V. B., & Tavlas, G. S. (2009). Where has all the money gone? Wealth and the demand for money in South Africa. *Journal of African Economies* 18(1): 84-112.

Hall, S. G., Hondroyiannis, G., Swamy, P. A. V. B., & Tavlas, G. S. (2010). The Fisher Effect Puzzle: A Case of Non-Linear Relationship. *Open Economies Review.*
DOI: 10.1007/s11079-009-9157-1.

Hall, S. G., Swamy, P. A. V. B., & Tavlas, G. S. (2017). "Time-Varying Coefficient Models: A Proposal For Selecting The Coefficient Driver Sets," Macroeconomic Dynamics, Cambridge University Press, vol. 21(05), pages 1158-1174, July

Hondroyiannis, G., Kenjegaliev, A., Hall, S. G., Swamy, P. A. V. B., & Tavlas, G. S. (2013). Is the relationship Between Prices and Exchange Rates Homogeneous *Journal of International Money and Finance* 37:411-436.
DOI: 10.1016/j.jimonfin.2013.06.014.

Jochmann, M., Koop, G., & Strachan, R. W. (2010). Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting* 26(2): 326-347.

Kenjegaliev, A., Hall, S. G., Tavlas, G. S., & Swamy, P. A. V. B. (2013). The forward rate premium puzzle: A case of misspecification? *Studies in Nonlinear Dynamics and Econometrics* 3: 265-280.
DOI: 10.1515/snde-2013-0009.

Pavlou, M., Ambler, G., Seaman, S., De Lorio, M., & Omar, Z. (2016). Review and Evaluation of Penalised Regression Methods for Risk prediction in Low-Dimensional data with few Events. *Statistics in Medicine* 37(7) 1159-1177.

Swamy, P. A. V. B., Chang, I-L., Mehta, J. S., & Tavlas, G. S. (2003). Correcting for Omitted-Variable and Measurement-Error Bias in Autoregressive Model Estimation with Panel Data,. *Computational Economics* 22: 225–253.

Swamy, P. A. V. B., Hall, S. G., & Tavlas, G. S. (2012). Generalized cointegration: a new concept with an application to health expenditure and health outcomes *Empirical Economics* 42: 603-18.
DOI: 10.1007/s00181-011-0483-y.

Swamy, P. A. V. B., Hall, S. G., & Tavlas, G. S. (2015). A Note on Generalizing the Concept of Cointegration. *Macroeconomic Dynamics,* 19(7): 1633-1646.
DOI: 10.1017/S1365100513000928.

Swamy, P. A. V. B., Hall, S. G., Tavlas, G. S., & Hondroyiannis, G. (2010). Estimation of Parameters in the Presence of Model Misspecification and Measurement Error. *Studies in Nonlinear Dynamics & Econometrics* 14: 1-35.

Swamy, P. A. V. B., & Mehta, J. S. (1975). Bayesian and Non-Bayesian Analysis of Switching Regressions and a Random Coefficient Regression Model. *Journal of the American Statistical Association*, 70, 593-602.

Tavlas, G. S., Swamy, P. A. V. B., Hall, S. G., & Kenjegaliev, A. (2013). Measuring Currency Pressures: The Cases of the Japanese Yen, the Chinese Yuan, and the U.K. Pound. *Journal of the Japanese and International Economies* 29: 1-20.
DOI: 10.1016/j.jjie.2013.04.001.

Table 1. Monte Carlo results for Model I

Corr ρ	0.95	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
T=50											
Bias	.017	.017	.025	0.028	.035	.048	.071	.098	.117	.125	.205
s.d.	.011	.011	.012	.019	.041	.057	.091	.120	.144	.189	.265
T=100											
Bias	.007	.007	.009	.017	.022	.035	.077	.114	.135	.181	.235
s.d.	.008	.008	.007	.014	.035	.052	.128	.140	.192	.272	.301
T=200											
bias	.004	.004	.007	.012	0.011	.070	.101	.177	.186	.244	.293
s.d.	.006	.006	.005	.007	.012	.044	.177	.186	.281	.316	.387
T=500											
bias	.003	.003	.005	.008	.011	.079	.136	.218	.222	.271	.332
s.d.	.004	.004	.006	.007	.032	.055	.190	.277	.334	.389	.415
T=1000											
bias	.001	.001	.003	.005	.007	.065	.225	.280	.345	.381	.414
s.d.	.003	.003	.005	.009	.041	.062	.217	.305	.376	.414	.520

Notes: Corr ρ is the degree of correlation between the true driver and the driver used. T is the sample size, Bias is the percent absolute bias; s.d. is the standard deviation of the bias.

Table 2. Monte Carlo results for Model II

Corr ρ	0.95	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
T=50											
bias	.021	.021	.029	.032	.038	.041	.058	.069	.082	.098	.125
s.d.	.013	.013	.015	.019	.022	.033	.066	.083	.102	.128	.155
T=100											
bias	.014	.014	.021	.028	.032	.055	.067	.091	.122	.144	.171
s.d.	.009	.009	.012	.017	.019	.028	.077	.107	.135	.176	.193
T=200											
bias	.009	.009	.017	.022	.027	.051	.083	.129	.142	.185	.215
s.d.	.007	.008	.010	.015	.017	.020	.154	.196	.226	.287	.303
T=500											
bias	.007	.008	.011	.018	.022	.049	.124	.155	.189	.212	.288
s.d.	.006	.007	.008	.013	.016	.022	.187	.234	.288	.317	.355
T=1,000											
bias	.005	.006	.008	.010	.017	.047	.171	.222	.287	.334	.345
s.d.	.004	.005	.007	.011	.014	.020	.199	.276	.302	.344	.381

Notes: Corr ρ is the degree of correlation between the true driver and the driver used. T is the sample size, Bias is the percent absolute bias; s.d. is the standard deviation of the bias.

Table 3. Monte Carlo results for Model III, $\rho_{\varepsilon 1,z1}$=0.50

Corr ρ	0.95	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
T=50											
bias	.022	.024	.031	.036	.041	.055	.062	.077	.085	.097	.105
s.d.	.011	.011	.015	.019	.022	.037	.055	.071	.080	.092	.101
T=100											
bias	.018	.019	.022	.029	.035	.050	.071	.082	.091	.108	.117
s.d.	.009	.009	.015	.021	.030	.047	.077	.093	.105	.120	.146
T=200											
bias	.009	.009	.015	.021	.031	.047	.087	.095	.119	.126	.141
s.d.	.007	.008	.012	.019	.027	.045	.090	.101	.138	.155	.188
T=500											
bias	.007	.008	.011	.016	.027	.040	.090	.122	.139	.155	.196
s.d.	.004	.005	.009	.017	.020	.039	.115	.137	.152	.188	.217
T=1,000											
bias	.005	.005	.007	.009	.016	.032	.117	.144	.162	.196	.213
s.d.	.003	.003	.006	.010	.016	.030	.141	.166	.195	.225	.255

Notes: Corr ρ is the degree of correlation between the true driver and the driver used. T is the sample size, Bias is the percent absolute bias; s.d. is the standard deviation of the bias.

Table 4. Monte Carlo results for Model III, $\rho_{\varepsilon 1,z1}$=0.50, SSVS

Corr ρ	0.95	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
T=50											
bias	.025	.026	.033	.038	.044	.059	.067	.079	.091	.099	.109
s.d.	.012	.012	.016	.020	.023	.039	.057	.075	.086	.095	.114
$\Pi*$	60.5%	60.0%	58.2%	57.3%	51.3%	33.3%	12.2%	8.3%	4.5%	0.0%	0.0%
T=100											
bias	.019	.020	.024	.031	.038	.053	.075	.087	.096	.112	.120
s.d.	.009	.011	.017	.023	.034	.049	.079	.098	.114	.126	.151
$\Pi*$	71.2%	71.0%	62.3%	64.8%	58.2%	55.4%	9.3%	7.5%	3.3%	0.0%	0.0%
T=200											
bias	.012	.012	.019	.027	.035	.049	.089	.099	.121	.127	.148
s.d.	.008	.008	.015	.021	.029	.047	.082	.103	.140	.159	.192
$\Pi*$	87.3%	87.0%	77.3%	71.2%	61.5%	59.2%	8.2%	3.7%	0.0%	0.0%	0.0%
T=500											
bias	.009	.009	.014	.019	.029	.043	.094	.128	.140	.162	.200
s.d.	.005	.006	.012	.019	.023	.040	.119	.141	.158	.193	.232
$\Pi*$	97.3%	96.5%	93.4%	85.5%	79.3%	62.7%	4.4%	1.0%	0.0%	0.0%	0.0%
T=1,000											
bias	.006	.006	.009	.015	.019	.035	.121	.147	.168	.201	.217
s.d.	.004	.004	.007	.015	.018	.034	.144	.169	.198	.230	.266
$\Pi*$	99.5%	98.3%	97.7%	91.2%	85.2%	77.7%	2.1%	0.0%	0.0%	0.0%	0.0%

Notes: Corr ρ is the degree of correlation between the true driver and the driver used. T is the sample size, Bias is the percent absolute bias; s.d. is the standard deviation of the bias. $\Pi*$ is the proportion of times the correct driver set is selected

Table 5. Monte Carlo results for Model IV, nonlinearity, SSVS / Fourier basis

δ	0.1	0.3	0.5	1.00	5.00
T=50					
bias	0.022	0.025	0.028	0.031	0.035
s.d.	0.014	0.015	0.016	0.019	0.023
T=100					
bias	0.017	0.022	0.025	0.029	0.032
s.d.	0.011	0.012	0.014	0.017	0.020
T=200					
bias	0.013	0.015	0.017	0.019	0.021
s.d.	0.008	0.009	0.011	0.012	0.017
T=500					
bias	0.009	0.010	0.012	0.014	0.018
s.d.	0.005	0.007	0.009	0.010	0.014
T=1,000					
bias	0.005	0.006	0.007	0.011	0.013
s.d.	0.003	0.004	0.006	0.009	0.011

Notes: δ is the degree of missing nonlinearity given as in Section 4.5 above. T is the sample size. Bias is the percent absolute bias; s.d. is the standard deviation of the bias.

Table 6. Monte Carlo results for Model V, Endogeneity, SSVS

corr ρ		0.1	0.3	0.5	0.7	0.9
T=50	bias	0.491	0.401	0.387	0.266	0.212
	s.d.	0.716	0.710	0.688	0.644	0.601
	Π^*	0.044	0.081	0.225	0.447	0.617
T=100	bias	0.386	0.316	0.303	0.201	0.138
	s.d.	0.355	0.350	0.281	0.277	0.252
	Π^*	0.051	0.128	0.315	0.517	0.645
T=200	bias	0.300	0.216	0.287	0.181	0.101
	s.d.	0.314	0.310	0.277	0.201	0.196
	Π^*	0.081	0.201	0.403	0.615	0.717
T=500	bias	0.295	0.290	0.101	0.087	0.071
	s.d.	0.201	0.200	0.096	0.061	0.055
	Π^*	0.091	0.261	0.462	0.687	0.775
T=1,000	bias	0.282	0.280	0.047	0.031	0.028
	s.d.	0.195	0.193	0.032	0.027	0.016
	Π^*	0.101	0.316	0.518	0.784	0.801
T=10,000	bias	0.190	0.190	0.039	0.001	0.001
	s.d.	0.182	0.182	0.030	0.011	0.008
	Π^*	0.115	0.320	0.615	0.813	0.917

Notes: Corr ρ is the degree of correlation between the true driver and the driver used. Π^* is the proportion of cases where *all* the drivers $z_{t1},...,z_{t4}$ are included in the model *and* the drivers $z_{t5},...,z_{t8}$ are *all* excluded. T is the sample size. Bias is the percent absolute bias; s.d. is the standard deviation of the bias.