Article

An Alternative Estimation Method for Time-Varying Parameter Models

Mikio Ito [1], **Akihiko Noda** [2,3] **and Tatsuma Wada** [4,*]

1. Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan; ito@econ.keio.ac.jp
2. School of Commerce, Meiji University, 1-1 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8301, Japan; anoda@meiji.ac.jp
3. Keio Economic Observatory, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
4. Faculty of Policy Management, Keio University, 5322 Endo, Fujisawa 252-0882, Kanagawa , Japan
* Correspondence: twada@sfc.keio.ac.jp

Abstract: A multivariate, non-Bayesian, regression-based, or feasible generalized least squares (GLS)-based approach is proposed to estimate time-varying VAR parameter models. Although it has been known that the Kalman-smoothed estimate can be alternatively estimated using GLS for univariate models, we assess the accuracy of the feasible GLS estimator compared with commonly used Bayesian estimators. Unlike the maximum likelihood estimator often used together with the Kalman filter, it is shown that the possibility of the pile-up problem occurring is negligible. In addition, this approach enables us to deal with stochastic volatility models, models with a time-dependent variance–covariance matrix, and models with non-Gaussian errors that allow us to deal with abrupt changes or structural breaks in time-varying parameters.

Keywords: Kalman filter; non-Bayesian time-varying model; generalized least squares; vector autoregressive model

1. Introduction

Most macroeconomists recognize that time-varying parameter models are sufficiently flexible to capture the complex nature of a macroeconomic system, thereby fitting the data better than models with constant parameters. The instability of the parameters in econometric models has often been incorporated in Markov-switching models (e.g., Hamilton 1989) and structural change models (e.g., Perron 1989). However, time-varying models allow the parameters to change gradually over time, which is the main difference between time-varying models and Markov-switching or structural break models.

In the literature on the application of time-varying vector autoregressive (TV-VAR) models in macroeconomics, Bernanke and Mihov (1998) consider that the autoregressive parameters may be time-varying. However, after confirming the stability of the parameters using the parameter consistency test of Hansen (1992), they employ the time-invariant (i.e., usual) VAR model. Indeed, Cogley and Sargent (2005) find that Hansen's (1992) test has low power and is unreliable and instead propose a TV-VAR model with stochastic volatility in the error term. Primiceri (2005) sheds light on a technical aspect of the time-varying model, namely, the Bayesian estimation technique used for the time-varying parameters. In general, difficulties in dealing with time-varying parameter models arise when free parameters and unobserved variables need to be estimated. Primiceri (2005) thus presents a clear estimation procedure based on the Bayesian Markov Chain Monte Carlo (MCMC) method.

Several studies, including Primiceri (2005), claim that the Bayesian method is preferred to the maximum likelihood (ML) method because the former (i) is less likely to suffer from the so-called pile-up problem (Sargan and Bhargava 1983); (ii) is less likely to have computation problems such as a degenerated likelihood function or multiple local minima; and (iii) helps find statistical inferences such as standard errors. However, both the Bayesian

and the ML methods require Kalman filtering to estimate an unobservable state vector that includes the time-varying parameters.[1]

Duncan and Horn (1972), Sant (1977), Maddala and Kim (1998), more recently, Chan and Jeliazkov (2009) and Durbin and Koopman (2012) have attempted to understand Kalman filtering through the lens of conventional regression models.[2] To the best of our knowledge, Duncan and Horn (1972) are the first to show that the generalized least squares (GLS) estimator for basic state-space models equivalently uncovers the unobserved state vector estimated through Kalman filtering. Sant (1977) proves the equivalence between the GLS estimator and Kalman-smoothed estimator. From a practical perspective, the series of papers by Ito et al. (2014, 2016, 2021) apply the TV-VAR, time-varying autoregressive (TV-AR) and time-varying vector error correction (TV-VEC) models to stock prices and exchange rates using the regression method as opposed to the Kalman filter.

In this study, in the spirit of Duncan and Horn (1972) and Sant (1977), we propose a multivariate, regression-based approach or GLS-based approach that utilizes ordinary least squares (OLS) or GLS in lieu of the Kalman smoother. More precisely, our GLS-based approach includes OLS as a variety of GLS. It is employed by Ito et al. (2014, 2016, 2021) to evaluate market efficiency in stock markets and foreign exchange markets.[3] However, the main purpose of employing the Kalman filter (or smoother) is to avoid using a system of large matrices required by GLS—at least until computers became capable of handling large matrices.

The equivalence between GLS and the Kalman smoother leads us to the following question: if GLS yields Kalman-smoothed estimates, to what extent can the GLS-based approach recover time-varying parameters? This question is practical and important because the finite sample properties of the GLS estimator are, generally, unknown.[4] Another question pertains to the seriousness of the pile-up problem, which occurs when the ML estimation of the variance of the state equation error is zero, even though its true value is non-zero (but small). While our proposed method is not identical to ML because we do not maximize the likelihood function with respect to the variances of the errors, whether our GLS-based approach suffers from the pile-up problem to the same degree as ML is not immediately obvious.

We also consider the possibility of non-independent and identically distributed (i.i.d.) or non-Gaussian errors in the model. The former are frequently used in this field because it is reasonable to assume that the variance of the errors may be time-varying. The latter are important in empirical studies because they allow us to model abrupt changes or structural breaks in time-varying parameters similar to the strategy employed by Perron and Wada (2009) among others.

To summarize, the contributions of this study are as follows: GLS estimates the true time-varying parameters fairly well even when the errors are not i.i.d. or not Gaussian, provided an appropriate way to implement FGLS is carefully chosen based primarily on the relative size of the variances of the errors or signal-to-noise ratio (SNR). The pile-up problem that is often cumbersome to ML is shown to be negligible. In addition, our GLS outperforms the commonly used Bayesian estimation method in recovering the time-varying parameter values.

The rest of this paper is organized as follows. Section 2 presents our model together with its likelihood function. We explain the GLS-based approach for the class of TV-AR models and steps used to implement FGLS in Section 3. Section 4 evaluates the GLS-based approach in a variety of conditions such as a small SNR, non-i.i.d. errors, and non-Gaussian errors.

An application to macroeconomic data, including a comparison with the Bayesian MCMC method, is demonstrated in Section 5. Section 6 concludes.

2. Model

Our model allows the class of TV-AR models and permits two different matrix forms. The first matrix form is that of Durbin and Koopman (2012), which they use as a device to

find the Kalman-smoothed estimate of an unobserved state vector. The second matrix form is an extended version of Maddala and Kim (1998), which we employ in this study. As it will become clear, this form allows us to use GLS to estimate the time-varying parameters. We can then formally demonstrate that the Kalman-smoothed estimate of the first matrix form is equivalent to the GLS estimates of the second matrix form, showing that GLS estimates are an alternative estimation method to the Kalman smoother.

2.1. Basic State-Space Model of the Class of TV-AR Models

Our model is given by:

$$y_t = Z_t \beta_t + \varepsilon_t \tag{1}$$
$$\beta_t = \beta_{t-1} + \eta_t, \tag{2}$$

where y_t is a $k \times 1$ vector of observable variables; Z_t is a $k \times m$ matrix of observable variables; β_t is an $m \times 1$ vector of time-varying parameters; and ε_t and η_t are $k \times 1$ and $m \times 1$ vectors of normally distributed error terms with the variance–covariance matrices of H_t and Q_t, respectively:[5]

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} H_t & 0 \\ 0 & Q_t \end{bmatrix} \right).$$

The variance–covariance matrices H_t and Q_t are allowed to be time-dependent, as in the stochastic volatility model. For the initial value of β_t, we assume

$$\beta_0 \sim N(b_0, P_0).$$

If we assume b_0 and P_0 are known, it is reasonable to use the diffuse prior for P_0 because β_t follows a non-stationary process. In this case, the diagonal elements of P_0 should be large numbers (e.g., Harvey 1989; Koopman 1997). Alternatively, we can simply ignore P_0 as zero when we assume β_0 is known and not stochastic.

Equations (1) and (2) can be used for a variety of TV-AR models. For example, when $k = 1$, $Z_t = y_{t-1}$ yields a TV-AR(1) model. Similarly, the TV-VAR(1) model $y_t = A_t y_{t-1} + \varepsilon_t$ with $A_t = A_{t-1} + \eta_t$ is expressed by setting $Z_t = (y'_{t-1} \otimes I_k)$ and $\beta_t = vec(A_t)$. It is also possible to include intercepts that vary over time. For a TV-AR(1) model, for example, one can set $Z_t = (1, y_{t-1})$, meaning that the first element of β_t is the time-varying intercept.

Below, we present two specifications of our model, (1) and (2). The first specification allows us to derive the Kalman-smoothed estimate as explained by Durbin and Koopman (2012). The second specification is in the spirit of Duncan and Horn (1972), leading us to the GLS-based approach. As we will see, both specifications yield the same smoothed estimate.

2.2. Model Matrix Formulation of the State-Space Model

Following Durbin and Koopman (2012), we employ the matrix formulation of Equations (1) and (2). For $t = 1, \ldots, T$, we have a system of equations:

$$Y_T = Z\beta + \varepsilon \tag{3}$$
$$\beta = C(b_0^* + \eta) \tag{4}$$

$$\varepsilon \sim N(0, H), \quad \eta \sim N(0, Q),$$

with

$$Y_T = \begin{bmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_T \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{p+1} & & & 0 \\ & Z_{p+2} & & \\ & & \ddots & \\ 0 & & & Z_T \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_{p+1} \\ \beta_{p+2} \\ \vdots \\ \beta_T \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

$$H = \begin{bmatrix} H_{p+1} & & & 0 \\ & H_{p+2} & & \\ & & \ddots & \\ 0 & & & H_T \end{bmatrix}, \quad C = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & I & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ I & I & I & I \end{bmatrix}, \quad b_0^* = \begin{bmatrix} b_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$P_0^* = \begin{bmatrix} P_0 & 0 & & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_{p+1} \\ \eta_{p+2} \\ \vdots \\ \eta_T \end{bmatrix}, \quad Q = \begin{bmatrix} Q_{p+1} & & & 0 \\ & Q_{p+2} & & \\ & & \ddots & \\ 0 & & & Q_T \end{bmatrix}.$$

Unlike a more general state-space model in which Equation (2) has a transition matrix that includes unknown parameters to be estimated, the matrix formulation of the time-varying parameter model is largely simplified. For example, matrix C is often called the random walk generating matrix (e.g., Tanaka 2017), which is non-singular, and there are no free parameters to be estimated in the matrix. In addition, if H_t and Q_t are time invariant (i.e., if no GARCH effects or stochastic volatility exists in the model), matrices H and Q are simplified substantially.

For simplicity, we assume b_0 is known and non-stochastic; hence, $P_0 = 0$.[6]

2.3. Likelihood Function

Since we assume that ε and η are normally distributed, our matrix formulation of (1) and (2) allows us to write the log-likelihood function for Y_T given the covariance matrices of the errors (H and Q), and initial value vector (b_0^*) as

$$\log p(Y_T|H, Q, b_0^*) = -\frac{(T-p)k}{2}\log 2\pi - \frac{1}{2}\log|\Omega| - \frac{1}{2}(Y_T - ZCb_0^*)'\Omega^{-1}(Y_T - ZCb_0^*), \quad (5)$$

where

$$\Omega = H + ZCQC'Z'.$$

Interestingly, provided that H, Q, and b_0^* are known, the likelihood function does not involve our main parameter vector of interest, β.

3. Estimation of the TV-AR Models
3.1. Regression Lemma and Kalman Smoothing

Before showing the equivalence of our estimator and Kalman smoother, let us clarify the outcomes of the Kalman smoother when the model is described by Equations (1) and (2). According to Durbin and Koopman (2012), the Kalman-smoothed state of β is given by the expectation of β conditional on the information on all the observations of y_t:

$$\widetilde{\beta} = E[\beta|Y_T] = E[\beta] + Cov(\beta, Y_T)Var(Y_T)^{-1}(Y_T - E[Y_T]). \quad (6)$$

Note that we assume normal errors to derive Equation (6). The variance of β, given all the observations Y_T, is then

$$Var(\beta|Y_T) = Var(\beta) - Cov(\beta, Y_T)Var(Y_T)^{-1}Cov(\beta, Y_T)'. \quad (7)$$

The Kalman-smoothed estimate and its mean squared error (MSE) are given by (6) and (7), respectively. Durbin and Koopman (2012) call these equations the regression lemma,

which derives the mean and variance of the distribution of β conditional on Y_T, assuming the joint distribution of β and Y_T is a multivariate normal distribution. It follows that for (3) and (4), the Kalman-smoothed estimate is

$$\widetilde{\beta} = E[\beta|Y_T] = Cb_0^* + CQC'Z'\Omega^{-1}(Y_T - ZCb_0^*), \qquad (8)$$

and the conditional variance (or MSE) of the smoothed estimate is

$$Var(\beta|Y_T) = CQC' - CQC'Z'\Omega^{-1}ZCQC', \qquad (9)$$

where $\Omega = H + ZCQC'Z'$. Equations (8) and (9) are obtained given that $Cov(\beta, Y_T) = Var(\beta)Z' = CQC'Z'$ and $Var(Y_T) = ZCQC'Z' + H = \Omega$, and by substituting them into Equations (6) and (7). It is well known that (8) is a minimum-variance linear unbiased estimate of β, given Y_T, even though we do not assume the errors are normally distributed (e.g., Durbin and Koopman 2012).

3.2. Equivalence of the GLS-Based Estimator and Kalman Smoother

Equations (3) and (4) can be written in another matrix form to apply conventional regression analysis:

$$\left[\begin{array}{c} Y_T \\ -b_0^* \end{array} \right] = \left[\begin{array}{c} Z \\ -C^{-1} \end{array} \right] \beta + \left[\begin{array}{c} \varepsilon \\ \eta \end{array} \right]. \qquad (10)$$

This specification is similar to those of Duncan and Horn (1972) and Maddala and Kim (1998). The main difference between our specification and that of Duncan and Horn (1972) is that the former applies to a time-varying parameter model, while the latter is for a more general state-space model, which allows the transition equation to have a transition matrix F (i.e., when Equation (2) is $\beta_t = F\beta_{t-1} + \eta_t$). Since we do not need to estimate the transition matrix, our regressors in Equation (10) are all known.[7]

As mentioned by Duncan and Horn (1972), the confusion around the similarities and differences between Kalman filtering (including smoothing) and the conventional regression model stems from the fact that the former is the expectation of β, conditional on the information on Y_T, which is the linear projection of β onto the space spanned by Y_T (provided that the errors are normally distributed). On the contrary, the latter is a linear projection of the dependent variable onto the space spanned by the regressor, which is the projection of the left hand side of Equation (10) onto the space spanned by $\left[\begin{array}{cc} Z' & -C^{-1\prime} \end{array} \right]'$. However, Duncan and Horn (1972) show that GLS for (10) until the time-t observation yields the Kalman-filtered estimate.

As shown by Sant (1977), a natural extension of Duncan and Horn (1972) is that we obtain the Kalman-smoothed estimate of β when GLS is applied to all the observations, Y_T. We summarize the equivalence of the GLS estimator of model (10) and Kalman-smoothed estimator (8) and its MSE matrix (9) in Appendix A. We further reveal the equivalence when the time-varying model has time-invariant intercepts in Appendix B. With the likelihood function (5), GLS yields the ML estimator for the time-invariant intercepts and the Kalman-smoothed estimator for the rest (time-varying coefficients).

3.3. GLS in Practice

As shown in the previous subsections, under the condition that the variance–covariance matrices of the errors (H and Q) are known, the GLS estimator of β is identical to the Kalman-smoothed estimates. However, in practice, those variance–covariance matrices are generally unknown. The following two-step approach is often used to find the FGLS estimator. First, β can be estimated using OLS. Then, the OLS residuals are used to estimate H and Q, which are denoted as $\widehat{H}$ and $\widehat{Q}$, respectively. In the second step, FGLS is applied to our model assuming $\widehat{H}$ and $\widehat{Q}$ are the variance–covariance matrices of ε and η, respectively.

However, FGLS suffers from two problems. First, H and Q may involve too many unknown parameters. For example, when a TV-VAR(p) model has many variables (i.e.,

when k is large), H has $(T - p)$ of $k \times k$ matrices, and Q has the same number of $k(kp + 1) \times k(kp + 1)$ matrices. The second problem is possible heteroskedasticity. Suppose that ε is much greater in magnitude than η. More precisely, when the average trace of H is much larger than the average trace of Q, our GLS-based approach has heteroskedasticity in regression Equation (10), potentially causing an imprecise estimation of β. This concern is largely mitigated when the average traces of H and Q are similar.

To solve these two problems, we propose the following FGLS procedure.

- Step 1. We estimate model (10) by OLS and obtain the estimate of β by OLS, $\widehat{\beta}^O$. From the OLS residuals, $\widehat{\varepsilon}_t$ and $\widehat{\eta}_t$, we construct the first-step estimates of H_t and Q_t:

$$\widehat{H}_t = \frac{1}{T - p} \sum_{t=p+1}^{T} \widehat{\varepsilon}_t \widehat{\varepsilon}_t' \text{ and } \widehat{Q}_t = \frac{1}{T - p} \sum_{t=p+1}^{T} \widehat{\eta}_t \widehat{\eta}_t'.$$

Then, to construct the estimates of H and Q, denoted as $\widehat{H}^O$ and $\widehat{Q}^O$, respectively, we set $\widehat{H}_{p+1} = \widehat{H}_{p+2} = \cdots = \widehat{H}_T$ and $\widehat{Q}_{p+1} = \widehat{Q}_{p+2} = \cdots = \widehat{Q}_T$ to assume that the variances of ε and η are time-invariant. This assumption is undesirable because a number of studies of TV-VAR models have focused on stochastic volatility models, which require $\widehat{Q}_t \neq \widehat{Q}_{t+1}$, for example. However, thanks to this assumption, H and Q are always invertible, and those inverses are readily computed. The simulations in the next section will reveal how severely this assumption affects our estimation when stochastic volatility is present. With $\widehat{H}^O$ and $\widehat{Q}^O$, the log-likelihood is computed by (A6) or (5).

- Step 2 (1FGLS). Given $\widehat{H}^O$ and $\widehat{Q}^O$, we apply FGLS to obtain $\widehat{\beta}^{G1}$, which is the FGLS or 1FGLS estimate of β. We also compute the estimates of H and Q, denoted as $\widehat{H}^{G1}$ and $\widehat{Q}^{G1}$, respectively, in the same way as we computed $\widehat{H}^O$ and $\widehat{Q}^O$ in the first step. Then, the value of the log-likelihood function is computed.

- Step 3 (2FGLS). We repeat Step 2, computing $\widehat{\beta}^{G2}$, which is the (second-time) FGLS or 2FGLS of β. More precisely, we use the FGLS residuals in Step 2 to construct $\widehat{H}_t$ and $\widehat{Q}_t$ to obtain $\widehat{\beta}^{G2}$. Then, the value of the log-likelihood function is computed. If the likelihood ratio (from OLS to 1FGLS or from 1FGLS to 2FGLS) cannot be computed or is extraordinarily large, such as greater than 1e+10, we disregard the 1FGLS and 2FGLS estimators because both indicate that the variance–covariance matrix is not precisely estimated (degenerated). In such a case, we only record OLS. In addition, we define 2FGLS' as GLS using $\widehat{Y}_T = Z\widehat{\beta}_1^{G1}$ and $-\widehat{b}_0^* = -C^{-1}\widehat{\beta}_2^{G1}$ in place of $\widehat{\varepsilon}_t$ and $\widehat{\eta}_t$, respectively, to compute $\widehat{H}_t$ and $\widehat{Q}_t$, where $\widehat{\beta}_1^{G1}$ and $\widehat{\beta}_2^{G1}$ are the corresponding elements of 1FGLS, $\widehat{\beta}^{G1}$. The reason why we use $\widehat{\beta}^{G2\prime}$, which denotes 2FGLS', is that it is expected to ameliorate the effects arising from poorly estimated $\widehat{\beta}^{G1}$. That is perhaps due to misspecified H and Q. When those matrices are not correctly estimated, $\widehat{\beta}^{G1}$ may be far from its true value; hence, the residuals computed from $\widehat{\beta}^{G1}$ should not be used for further FGLS because the repeated use of the wrong variance–covariance matrices may make the estimator worse. In such a case, it may make sense to obtain $\widehat{\beta}^{G2\prime}$ as it does not repeat the same type of misspecification.

In summary, our procedure is based on the assumptions that the error terms have time-invariant variances and that the heteroskedasticity arising from the different sizes of H and Q can be correctly handled by the repeated use of FGLS. To validate our assumptions and procedure, we investigate the degree to which our procedure precisely estimates the true β using simulations.

4. Simulations

Among some influential empirical studies of TV-VAR, both Cogley and Sargent (2001, 2005) and Primiceri (2005) employ a three-variable TV-VAR(2) model. Hence, in our simulation study, we adopt the same specification and use simulations to assess how well the GLS-based approach recovers the true time-varying parameters relative to the Bayesian

approach. First, we compute the means and variances of the estimated time-varying parameters and compare them with those of the true time-varying parameters to evaluate the accuracy of both the GLS-based and the Bayesian[8] approaches. While comparing the first and second moments of the estimates to those of the true process may be inadequate to determine whether the GLS-based approach yields precise estimates, it is a useful way to grasp the overall accuracy of the estimates.[9]

Second, we consider the possibility of the pile-up problem. According to Primiceri (2005), the Bayesian approach is preferred when estimating time-varying parameter models because, among other reasons, it can potentially avoid the pile-up problem. However, the extent to which this problem affects our estimate is not immediately obvious because the literature (e.g., Shephard and Harvey 1990) provides theoretical explanations only for limited (simple) cases. On the contrary, our model can have a vector of time-varying terms (β_t), unlike prior studies that have analyzed scalar time-varying terms for simplicity. Therefore, it is reasonable to conduct a simulation study to reveal the extent to which our GLS-based approach suffers from the pile-up problem. Because the concern about the pile-up problem grows when the variance of the state equation error or the SNR is small, we study the performance of the GLS-based approach more comprehensively by altering the SNR in the data-generating process.

Third, we also evaluate the performance of the GLS-based approach when stochastic volatility and non-Gaussian errors are present. We investigate the effect of stochastic volatility on the GLS-based approach because macroeconomic research, including Cogley and Sargent (2005) and Primiceri (2005), has been allowing such shocks in the TV-VAR model. While the GLS-based approach does not require the assumption of i.i.d. errors to estimate β_t, we are interested in the extent to which the accuracy of the GLS-based approach is affected by the stochastic volatility of the errors. For the non-Gaussian errors, we focus on the possible structural breaks in the time-varying coefficients, β_t. By allowing a mixture of normal errors, as explained in the following subsection, we can model structural breaks or abrupt changes in β_t, as opposed to the gradual changes that the time-varying model generally assumes. Our simulation study is thus expected to shed light on the performance of the GLS-based approach when such errors are present.

Finally, as mentioned in Section 1, since we consider OLS to be a component of the GLS-based approach, we study its performance using simulations to clarify the relative performance FGLS and OLS, especially for small samples.

4.1. Data-Generating Process

We generate pseudo data by the system of Equations (3) and (4) with $T = \{100, 250\}$, $H = \{0.02^2 I, 0.2^2 I, 1^2 I\}$ and $Q = \{0.03^2 I\}$. By changing the variance of the error to the observation equation, we consider the role of the SNR, which we define as the average trace of the variance–covariance matrix of η_t relative to that of ε_t: In our simulation, we consider the SNRs for $0.03^2/0.02^2$, $0.03^2/0.2^2$ and $0.03^2/1^2$. The SNR is particularly important when we consider the possibility of the pile-up problem, which will be discussed in the next section. For the initial values, we set $b_0^* = 0$.

4.1.1. Non-Gaussian Errors

The original motivation to employ time-varying models for macroeconomic research was to allow β_t to change gradually. However, structural breaks or abrupt changes may exist in β_t, which means that β_t is almost constant over time until some point in the sample, for example, T_b; it then jumps to a different level afterward. One way to model such a break is to assume non-Gaussian errors for η_t. In particular, we assume mixtures of normal distributions (e.g., Perron and Wada 2009 for each element of error vector η_t:

$$\eta_{it} = \lambda_t \zeta_{1,t} + (1 - \lambda_t)\zeta_{2,t}$$

where

$$
\begin{aligned}
\lambda_t &\sim \quad i.i.d. Bernoulli(0.95) \\
\zeta_{1,t} &\sim \quad N\!\left(0, 0.03^2\right), \quad \zeta_{2,t} \sim N\!\left(0, 0.1^2\right).
\end{aligned}
$$

Intuitively, with a probability of 95%, η_t is $\zeta_{1,t}$, which is drawn from a normal distribution with a small variance. This small η_t keeps β_t nearly constant over time. However, a large η_t, which is $\zeta_{2,t}$, is drawn from a normal distribution with a (relatively) large variance. This η_t causes β_t to jump to a new level, with a 5% probability. Since we use the assumption of Gaussian errors to derive the equivalence between GLS and the Kalman-smoothed estimator, how non-Gaussian errors affect the accuracy of the GLS estimator when estimating β_t should be evaluated using simulations.

4.1.2. Stochastic Volatility and Autoregressive Stochastic Volatility

As Cogley and Sargent (2005) argue, in response to the criticism of Cogley and Sargent (2001), it is more flexible and realistic to assume that the variance of the shock ε_t is time-varying. Intuitively, not all shocks are generated from the same i.i.d. process. Since the GLS-based approach can handle the heteroskedasticity in H_t and Q_t, we can estimate the time-varying model with stochastic volatility, such as the one used by Primiceri (2005), at least theoretically. The error term ε_t may also follow the autoregressive stochastic volatility process described by Taylor (2007) and elsewhere.

However, in general, FGLS is merely a remedy to more precisely estimate the coefficients (in our case, β_t) when heteroskedasticity is present and FGLS is not primarily designed to estimate the process that the error term (or its variance) follows.

Nevertheless, we use the following data-generating process to assess the performance of the GLS-based approach.

$$
\begin{aligned}
\varepsilon_{it} &= \quad \sqrt{h_{i,t}}\,\xi_t \\
\log h_{i,t} &= \quad \rho \log h_{i,t-1} + e_t
\end{aligned}
$$

where $\rho = 1$ when stochastic volatility is considered and $\rho = 0.9$ when autoregressive stochastic volatility is considered; ε_{it} is the i-th element of ε_t. We assume $\log h_{i,0} = 0$, $e_t \sim N\!\left(0, 0.02^2\right)$, and $\xi_t \sim N(0, 1)$.

4.1.3. Eliminating Outliers

Since time-varying parameters mean that the generated series are generally non-stationary, some such series may be explosive and thus disregarded because they lack practical usefulness. In particular, if the standard deviation of the last 50 observations of a generated series is at least three times greater than that of the first 50 observations or if the inverse of $Z'Z$ does not exist, the series is discarded.

4.2. *Mean and Variance of the Estimated β_t and Likelihood*

Since we simulate a TV-VAR(2) model with time-varying intercepts, β_t is a 21×1 vector. Let $\beta_{t,i,n}$ denote the true (data-generating process) $\beta_{t,i,n}$ (i.e., the i-th element of vector $\beta_{t,n}$) and let $\widehat{\beta}_{t,i,n}$ denote the estimate of $\beta_{t,i,n}$. In the Bayesian MCMC case, we use the posterior mean for $\widehat{\beta}_{t,i,n}$.[10] Since b_0^* is unknown in practice when estimating $\beta_{t,i,n}$, we estimate b_0^* as the coefficient vector from a full-sample time-invariant (i.e., usual) VAR(2) model before estimating $\beta_{t,i,n}$ by OLS, GLS, or Bayesian. The sample means and sample standard deviations of the estimate over the sample period are then computed:

$$\overline{\widehat{\beta}}_{i,n} \;=\; \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\beta}_{t,i,n} \tag{11}$$

$$sd\left(\widehat{\beta}_{i,n}\right) \;=\; \sqrt{\frac{1}{T-p-1}\sum_{t=p+1}^{T}\left(\widehat{\beta}_{t,i,n}-\overline{\widehat{\beta}}_{i,n}\right)^{2}}. \tag{12}$$

Similarly, we compute those of the true (data-generating) process:

$$\overline{\beta}_{i,n} \;=\; \frac{1}{T-p}\sum_{t=p+1}^{T}\beta_{t,i,n} \tag{13}$$

$$sd(\beta_{i,n}) \;=\; \sqrt{\frac{1}{T-p-1}\sum_{t=p+1}^{T}\left(\beta_{t,i,n}-\overline{\beta}_{i,n}\right)^{2}}. \tag{14}$$

From (11) and (12) and their data-generating process counterparts, (13) and (14), we have 21 means and standard deviations for each replication. After $N = 1000$ replications, we compute the averages of $\overline{\widehat{\beta}}_{i,n}$, $\overline{\beta}_{i,n}$, $sd\left(\widehat{\beta}_{i,n}\right)$, and $sd(\beta_{i,n})$ over the replications. We then have 21 means of time-varying parameters and 21 means of standard deviations (i.e., $i = 1, 2, \ldots, 21$).

$$\widehat{m}_{i} \;=\; \frac{1}{N}\sum_{n=1}^{N}\overline{\widehat{\beta}}_{i,n};\;\; m_{i} \;=\; \frac{1}{N}\sum_{n=1}^{N}\overline{\beta}_{i,n} \tag{15}$$

$$\widehat{s}_{i} \;=\; \frac{1}{N}\sum_{n=1}^{N}sd\left(\widehat{\beta}_{i,n}\right);\;\; s_{i} \;=\; \frac{1}{N}\sum_{n=1}^{N}sd(\beta_{i,n}) \tag{16}$$

Since both m and s are aggregate means, a small difference between m and $\widehat{m}$ or between s and $\widehat{s}$ is only an indication that the estimator is close to what it is expected to estimate. Hence, we further investigate the similarities of β and $\widehat{\beta}$. Comparing each element of β, we define the distance, "*dist*", as follows:

$$dist_{i} = \frac{1}{N(T-p)}\sum_{n=1}^{N}\sum_{t=p+1}^{T}\left|\beta_{t,i,n}-\widehat{\beta}_{t,i,n}\right|. \tag{17}$$

Similarly, we compare the standard deviations of each element of β as a ratio of the standard deviation of $\widehat{\beta}_{i,n}$ to the standard deviation of the true process, $\beta_{i,n}$:

$$rat_{i} = \frac{1}{N}\sum_{n=1}^{N}\frac{sd\left(\widehat{\beta}_{i,n}\right)}{sd(\beta_{i,n})}. \tag{18}$$

In this simulation study, we focus on both $dist_{i}$ and rat_{i}. Our criteria for a good estimator are whether the $dist_{i}$ of an estimate is close to zero and whether the rat_{i} of that estimate is close to one.

4.3. Simulation Results 1: The SNR, Sample Size and Estimation Precision

Tables 1 and 2 display the medians of $dist_{i}$ and rat_{i} as well as the medians of m_{i} and s_{i} for $T = 100$ and $T = 250$, respectively.

Table 1. Simulation Results $T = 100$.

H	Q		True	OLS	1FGLS	2FGLS	2FGLS'	Primiceri
0.02^2	0.03^2	median m	−0.006	−0.001	0.000	0.001	0.001	0.000
		median s	0.086	0.033	0.016	0.019	0.010	0.000
		median $dist$		**0.129**	0.141	0.164	0.193	0.273
$SNR =$	2.25	median rat		**0.416**	0.206	0.252	0.124	0.006
0.2^2	0.03^2	median m	−0.002	−0.003	−0.003	−0.003	0.001	0.002
		median s	0.087	0.135	0.084	0.048	0.031	0.000
		median $dist$		0.164	0.131	**0.120**	0.122	0.273
$SNR =$	0.0225	median rat		1.759	**1.078**	0.609	0.397	0.006
1	0.03^2	median m	−0.001	−0.009	−0.009	−0.006	−0.005	−0.003
		median s	0.087	0.287	0.291	0.297	0.104	0.000
		median $dist$		0.272	0.277	0.278	**0.135**	0.273
$SNR =$	0.0009	median rat		3.761	3.812	3.904	**1.339**	0.005

Notes: (1) The numbers in the column under "True" are computed from the data-generating process described in Section 4.2. (2) "m", "s", "$dist$", and "rat" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS', Primiceri).

Table 2. Simulation Results $T = 250$.

H	Q		True	OLS	1FGLS	2FGLS	2FGLS'	Primiceri
0.02^2	0.03^2	median m	−0.007	−0.003	−0.002	−0.001	0.002	0.002
		median s	0.156	0.110	0.076	0.059	0.022	0.024
		median $dist$		**0.103**	0.126	0.150	0.263	0.348
$SNR =$	2.25	median rat		**0.718**	0.494	0.392	0.159	0.173
0.2^2	0.03^2	median m	−0.004	−0.006	−0.006	−0.005	0.002	0.003
		median s	0.156	0.201	0.150	0.105	0.061	0.048
		median $dist$		0.153	0.128	**0.120**	0.144	0.341
$SNR =$	0.0225	median rat		1.379	**1.013**	0.692	0.408	0.337
1	0.03^2	median m	−0.002	−0.003	−0.004	−0.005	−0.002	0.003
		median s	0.156	0.321	0.320	0.317	0.135	0.030
		median $dist$		0.243	0.243	0.243	**0.114**	0.341
$SNR =$	0.0009	median rat		2.285	2.282	2.265	**0.922**	0.208

Notes: (1) The numbers in the column under "True" are computed from the data-generating process described in Section 4.1. (2) "m", "s", "$dist$", and "rat" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS', Primiceri).

We focus on the median values of the estimated parameters because we sometimes encounter outliers. Our view is that the non-stationary nature of the data-generating process together with the possibility of poorly estimated H and Q, especially when the SNR is low, creates those outliers in the estimated coefficients. OLS works relatively well when the SNR is relatively large because, as Tables 1 and 2 show, the median distance of the estimate from the true process (i.e., $dist_i$) is small, and the median sample variance of the estimated β_t is closer to that of the true process compared with other approaches (i.e., rat_i is closer to one).[11] On the contrary, 2FGLS' works relatively well when the SNR is small. General tendencies from Table 1 can be summarized as follows. First, OLS, 1FGLS and 2FGLS share largely the same characteristics. However, the volatility of β_t estimated by 2FGLS' is much smaller than those estimated by OLS, 1FGLS and 2FGLS. Second, OLS, 1FGLS, 2FGLS and 2FGLS' all tend to have larger rat_i as the SNR increases. More precisely, OLS, 1FGLS, 2FGLS and 2FGLS' overestimate (underestimate) the volatility of β_t when the SNR is very small (large). Third, for the median distance of the estimate from the true process (i.e., $dist_i$) for T = 250, the best case for OLS and 1FGLS is when the SNR is 2.25. This phenomenon is easy to understand because an SNR that is either too small or too large make the estimation of β_t difficult, since an SNR far from one means the degree of heterogeneity is quite serious. In such a situation, it is easy to imagine that OLS cannot recover β_t well and 1FGLS may thus be unsuitable for implementing FGLS.

What is the effect of increasing the sample size? A comparison of Tables 1 and 2 show that the degrees to which the volatility of β_t is over- or underestimated is largely mitigated for OLS, 1FGLS and 2FGLS when the sample size increases from 100 to 250. At the same time, the median distances of the estimate from the true process for OLS, 1FGLS and 2FGLS generally shorten as the sample size increases, showing that the accuracy of OLS and 1FGLS improves with the sample size. However, such effects of an increased sample size do not clearly hold for 2FGLS'. Furthermore, Primiceri's (2005) Bayesian estimation is inaccurate at recovering the true parameter values throughout the simulation. As Primiceri (2005) explains in detail, one caveat is that the prior for Q is crucial for the volatility of the posterior mean of β_t. Hence, it may not be appropriate to use the same set of priors as Primiceri (2005) for this simulation study because more suitable prior values may improve the results. However, we must note that the other estimates do not necessitate such a choice depending on the SNR.

Additionally, instead of computing the median value of each measure over i, we can look closely at $dist_i$ and rat_i for each i. There are some parameters (i) where Primiceri's (2005) method outperforms the others. Yet, our focus here is to see the overall performance of each method.

4.4. Simulation Results 2: The Effects of Non-i.i.d. and Non-Gaussian Errors

Table 3 presents the effects of non-Gaussian errors as well as stochastic volatility and stochastic autoregressive errors.

Table 3. Stochastic Volatility and Autoregressive Stochastic Volatility.

T	Q		True	OLS	1FGLS	2FGLS	2FGLS'	Primiceri
100	0.03^2	median m	−0.002	−0.002	−0.003	−0.002	−0.005	−0.001
	RW	median s	0.086	0.289	0.295	0.298	0.103	0.000
		median $dist$		0.273	0.278	0.279	**0.135**	0.273
		median rat		3.837	3.916	3.986	**1.369**	0.005
100	0.03^2	median m	−0.002	−0.002	−0.003	−0.002	−0.005	−0.002
	AR	median s	0.086	0.289	0.295	0.298	0.103	0.000
		median $dist$		0.273	0.278	0.279	**0.135**	0.273
		median rat		3.837	3.916	3.986	**1.369**	0.005
250	0.03^2	median m	−0.002	−0.006	−0.005	−0.005	−0.004	0.001
	RW	median s	0.154	0.321	0.320	0.318	0.136	0.030
		median $dist$		0.244	0.244	0.244	**0.114**	0.343
		median rat		2.299	2.291	2.277	**0.928**	0.211
250	0.03^2	median m	−0.002	−0.008	−0.008	-0.009	−0.005	0.002
	AR	median s	0.154	0.322	0.321	0.318	0.137	0.030
		median $dist$		0.243	0.244	0.244	**0.114**	0.338
		median rat		2.310	2.303	2.292	**0.937**	0.213

Notes: (1) The numbers in the column under "True" are computed from the data-generating process: $\eta_{it} = \lambda_t \zeta_t^1 + (1 - \lambda_t)\zeta_t^2$ where $\lambda_t \sim i.i.d. Bernoulli(0.95)$, $\zeta_{1,t} \sim N(0, 0.03^2)$, $\zeta_{2,t} \sim N(0, 0.1^2)$ and $\varepsilon_{it} = \sqrt{h_{i,t}}\xi_t$, $\log h_{i,t} = \rho \log h_{i,t-1} + e_t$ where $\rho = 1$ when stochastic volatility is considered (labeled as RW), $\rho = 0.9$ when autoregressive stochastic volatility is considered (labeled as AR),ε_{it} is the i-th element of ε_t; $\log h_{i,0} = 0$, $e_t \sim N(0, 0.02^2)$, and $\xi_t \sim N(0, 1)$. (2) "m", "s", "$dist$", and "rat" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS', Primiceri).

The general tendencies that appear in the Gaussian error case (Tables 1 and 2) are maintained. While both OLS, 1FGLS and 2FGLS overestimate the volatility of β_t, the degree of overestimation is largely mitigated when the sample size increases. Moreover, 2FGLS' underestimates the volatility of β_t, and increasing the sample size helps 2FGLS' estimate β_t more accurately, and Primiceri's (2005) Bayesian approach cannot estimate β_t well. Remarkably, given the value of the autoregressive parameter ρ, there is also a negligible difference between the stochastic volatility and autoregressive stochastic volatility cases.

When only the non-Gaussian error is considered, as Tables 4 and 5 show, we obtain mostly the same results as those presented in Tables 1 and 2. Once again, except for the Bayesian estimator, the degree of overestimation (underestimation) depends on the SNR. Similar to the results in Tables 1 and 2, a larger sample size generally improves the estimation by OLS, 1FGLS and 2FGLS in that the degree of over- or underestimation is largely reduced when the sample size increases. In addition, for OLS, 1FGLS and 2FGLS, the median distance between the true and estimated β_t shortens with the sample size. However, this tendency does not apply to 2FGLS'.

What is the effect of scholastic volatility or autoregressive volatility in the observation equation error (ε_t) on our estimation? Table 6 shows that except for Primiceri's (2005) Bayesian approach, the results arising from such errors are similar to the small SNR cases in Tables 1 and 2.

Table 4. Mixtures of Normals $T = 100$.

H	Q		True	OLS	1FGLS	2FGLS	2FGLS'	Primiceri
0.02^2		median m	−0.002	0.002	0.004	0.004	0.004	0.003
		median s	0.102	0.042	0.025	0.021	0.012	0.001
		median $dist$		**0.137**	0.151	0.171	0.229	0.328
		median rat		**0.477**	0.266	0.229	0.139	0.006
0.2^2		median m	−0.001	−0.004	−0.003	−0.003	0.006	0.007
		median s	0.103	0.138	0.092	0.059	0.033	0.000
		median $dist$		0.164	0.136	**0.129**	0.135	0.310
		median rat		1.497	**0.987**	0.633	0.362	0.006
1		median m	−0.002	−0.007	−0.007	−0.007	−0.007	0.000
		median s	0.105	0.277	0.281	0.284	0.099	0.000
		median $dist$		0.259	0.264	0.265	**0.135**	0.319
		median rat		3.068	3.112	3.139	**1.066**	0.005

Notes: (1) The numbers in the column under "True" are computed from the data-generating process: $\eta_{it} = \lambda_t \zeta_t^1 + (1 - \lambda_t)\zeta_t^2$ where $\lambda_t \sim i.i.d. Bernoulli(0.95)$, $\zeta_{1,t} \sim N(0, 0.03^2)$, $\zeta_{2,t} \sim N(0, 0.1^2)$. (2) "$m$", "$s$", "$dist$", and "$rat$" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS', Primiceri).

This is because the observation error (ε_t) has a variance larger than one due to the stochastic volatility ($\sqrt{h_t}$) term. The results of the stochastic volatility and autoregressive stochastic volatility cases are therefore similar.

Table 5. Mixtures of Normals $T = 250$.

H	Q		True	OLS	1FGLS	2FGLS	2FGLS′	Primiceri
0.02^2		median m	-0.006	-0.001	-0.002	-0.003	0.004	0.002
		median s	0.181	0.132	0.101	0.083	0.030	0.045
		median $dist$		**0.109**	0.130	0.153	0.291	0.388
		median rat		**0.738**	0.554	0.460	0.183	0.268
0.2^2		median m	-0.004	-0.005	-0.004	-0.003	0.003	0.002
		median s	0.182	0.212	0.169	0.128	0.064	0.074
		median $dist$		0.150	0.131	**0.129**	0.177	0.387
		median rat		1.235	**0.969**	0.726	0.373	0.450
1		median m	-0.002	-0.006	-0.006	-0.006	-0.007	0.000
		median s	0.181	0.318	0.317	0.329	0.138	0.063
		median $dist$		0.228	0.229	0.239	**0.127**	0.385
		median rat		1.918	1.906	1.961	**0.795**	0.373

Notes: (1) The numbers in the column under "True" are computed from the data-generating process: $\eta_{it} = \lambda_t \zeta_t^1 + (1 - \lambda_t)\zeta_t^2$ where $\lambda_t \sim i.i.d.Bernoulli(0.95)$, $\zeta_{1,t} \sim N(0, 0.03^2)$, $\zeta_{2,t} \sim N(0, 0.1^2)$. (2) "$m$", "$s$", "$dist$", and "$rat$" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS′, Primiceri).

Table 6. Stochastic Volatility, Autoregressive Stochastic Volatility, and Mixtures of Normals.

T	RW/AR		True	OLS	1FGLS	2FGLS	2FGLS′	Primiceri
100	RW	median m	-0.002	-0.010	-0.010	-0.008	-0.008	-0.004
		median s	0.104	0.275	0.278	0.282	0.098	0.000
		median $dist$		0.258	0.261	0.262	**0.134**	0.317
		median rat		3.070	3.131	3.175	**1.078**	0.005
	AR	median m	-0.002	-0.010	-0.010	-0.008	-0.008	-0.004
		median s	0.104	0.275	0.278	0.282	0.098	0.000
		median $dist$		0.258	0.261	0.262	**0.134**	0.317
		median rat		3.070	3.130	3.177	**1.078**	0.005
250	RW	median m	-0.001	-0.005	-0.005	-0.006	-0.004	-0.001
		median s	0.180	0.317	0.315	0.314	0.135	0.060
		median $dist$		0.228	0.228	0.229	**0.127**	0.381
		median rat		1.924	1.913	1.905	**0.785**	0.361
	AR	median m	-0.001	-0.005	-0.006	-0.007	-0.004	-0.001
		median s	0.180	0.317	0.315	0.314	0.135	0.060
		median $dist$		0.228	0.228	0.229	**0.127**	0.382
		median rat		1.923	1.916	1.911	**0.785**	0.362

Notes: (1) The numbers in the column under "True" are computed from the data-generating process: $\varepsilon_{it} = \sqrt{h_{i,t}}\xi_t$, $\log h_{i,t} = \rho \log h_{i,t-1} + e_t$ where $\rho = 1$ when stochastic volatility is considered (labeled as RW), $\rho = 0.9$ when autoregressive stochastic volatility is considered (labeled as AR), ε_{it} is the i-th element of ε_t; $\log h_{i,0} = 0$, $e_t \sim N(0, 0.02^2)$, and $\xi_t \sim N(0, 1)$. (2) "m", "s", "$dist$", and "rat" stand for the mean, the standard deviation, the distance from the true values, and the ratio of the standard deviation of the estimates to that of the true values of β. (3) The bold numbers are the smallest (for median "$dist$") and the closest to one (for median "rat"), indicating the best method out of the four (OLS, 1FGLS, 2FGLS, 2FGLS′, Primiceri).

4.5. Discussion: The Pile-Up Problem

Our results suggest that the GLS-based approach does not suffer from the pile-up problem and that lower SNRs often lead to the overestimation of the volatility of β_t, especially when OLS, 1FGLS or 2FGLS is used (Tables 1 and 2). Moreover, the degree of overestimation of the sample variance of β_t becomes more severe when the sample size is small. This may be puzzling given the fact that OLS and ML are generally equivalent and that GLS and ML are equivalent if the errors are not i.i.d. (i.e., the errors heteroskedastic or autocorrelated). However, this statement is not true if FGLS fails to deal with non-i.i.d. errors appropriately. As OLS, 1FGLS and 2FGLS would then be unable to estimate an equivalent β_t to that under ML, this explains why the GLS-based approach does not suffer from the pile-up problem. Interestingly, our simulations reveal that the Bayesian estimator yields much smaller variations over time.

Therefore, our simulations seem to suggest both that the use of 2FGLS' is recommended when the sample size is small, and that OLS (1FGLS and 2FGLS as well) can recover the time-varying parameters fairly well when the sample size is large.

5. Application to the TV-VAR(2) Model with Interest Rates, Inflation, and Unemployment

A number of studies that employ TV-VAR models, including Cogley and Sargent (2005) and Primiceri (2005), focus on recovering the structural parameters from the estimated reduced form. Although we do not aim to identify fundamental shocks or compute impulse responses, we present the estimated TV-VAR(2) parameter using OLS (Figure 1), FGLS (Figure 2 for 1FGLS, Figure 3 for 2FGLS', and Figure 4 for 2FGLS), and the posterior mean of the time-varying approach using Primiceri's (2005) method (Figure 5).[12]
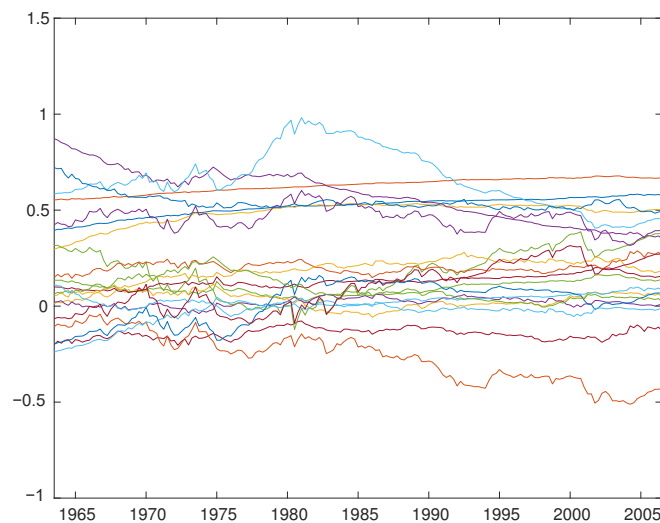

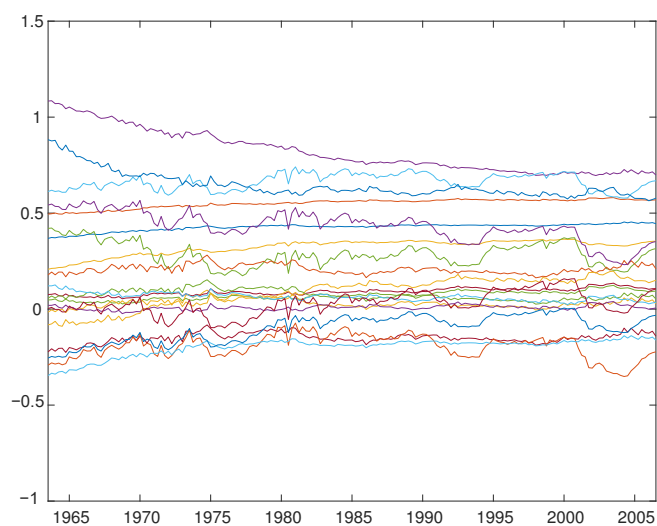**Figure 1.** The Estimated Time-Varying Parameters: OLS.

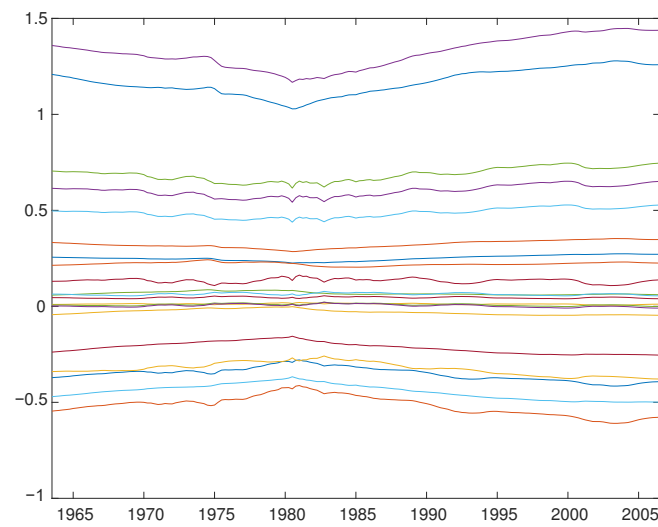**Figure 2.** The Estimated Time-Varying Parameters: 1FGLS.

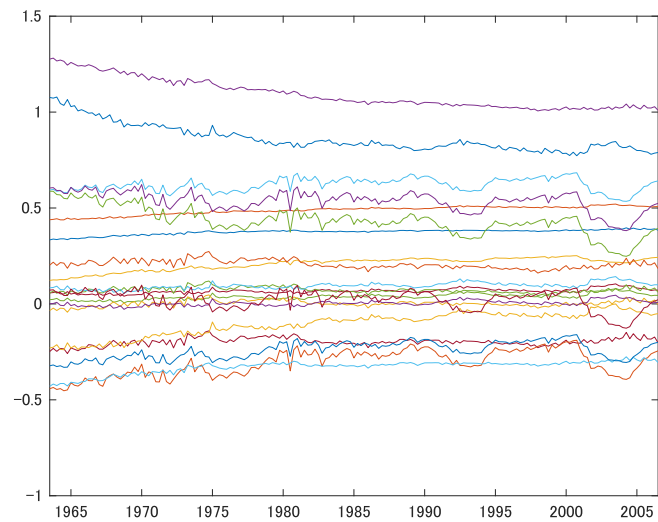**Figure 3.** The Estimated Time-Varying Parameters: 2GLS'.

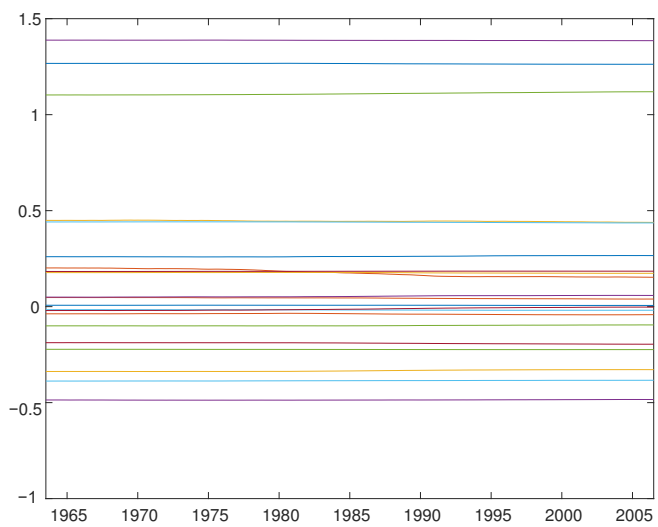**Figure 4.** The Estimated Time-Varying Parameters: 2GLS.

Figure 5. Replication of Primiceri (2005).

While Primiceri's (2005) Bayesian MCMC posterior means are virtually time-invariant and the estimates by 2FGLS' are slightly more volatile, the estimates provided by OLS, 1FGLS and 2FGLS have much larger volatility.

Interestingly, as detailed in the supplementary appendix (available upon request), the coefficients on the interest rate vary noticeably over time, exhibiting distinct patterns in the early 1980s (dip), late 1990s (up) and early 2000s (down). Similar to Primiceri's (2005) Bayesian posterior means, the intercepts (three time-varying coefficients) are largely stable over time. This finding is consistent with our simulation results. Overall, Primiceri's (2005) Bayesian estimator for the time-varying parameter (β_t) tends to have very small variation over time, resulting in virtually time-invariant parameters.[13]

6. Conclusions

The multivariate non-Bayesian regression-based or GLS-based approach for the time-varying parameter model is presented and assessed from a simulation aspect. Although this approach has already been theoretically justified and (at least partly) used by Ito et al. (2014, 2016, 2021), it is shown that using the GLS-based approach has at least four advantages. First, this approach can produce equivalent estimates without needing Kalman filtering or smoothing; it is also applicable to a wide range of time-varying parameter models (e.g., TV-AR, TV-VAR, and TV-VEC models) by adjusting the regression matrix accordingly. Second, the GLS-based approach works reasonably well in practice in that it can estimate the time-varying parameters even when non-i.i.d. errors or non-Gaussian errors exist in the model. In particular, we find that GLS outperforms Primiceri's (2005) Bayesian approach in recovering the true parameter values because it can take into account generally heteroskedastic error terms. The ability to deal with non-Gaussian errors is particularly important in empirical studies because it allows us to consider possible abrupt changes in time-varying parameters instead of gradual changes due to Gaussian errors. Remarkably, GLS can even outperform Primiceri's (2005) Bayesian approach that is often employed to estimate the TV-VAR models. However, in practice, the most appropriate method (OLS, 1FGLS, 2FGLS or 2FGLS') should be chosen depending on the sample size and SNR. More precisely, OLS is acceptable when the SNR is not far from one or the sample size is not small; otherwise, 2FGLS' is recommended. The reason why the sample size and SNR are important for choosing the preferred of the three methods is that 1FGLS, 2FGLS and 2FGLS' are not ideal GLS; hence, they cannot fully deal with the heterogeneity arising from our regression equation that includes both observation equation errors and state equation errors. However, because we do not maximize the unconditional likelihood function with respect to the variances of the errors and because 1FGLS, 2FGLS and 2FGLS' are not ideal GLS, the true variances are imprecisely estimated, and our GLS-based approach does not suffer from the pile-up problem that often occurs with ML.

While our focus in this paper is the estimation method of relatively simple TV models, more flexible models, such as a TV-VAR with time-varying variances of the structural shocks, have a higher necessity for macroeconomic analysis. Extending our approach to such complex models is of great importance to both econometricians and macroeconomists.

Author Contributions: Conceptualization, M.I., A.N. and T.W.; methodology, M.I. and T.W.; software, M.I., A.N. and T.W.; validation, A.N. and T.W.; formal analysis, T.W.; investigation, A.N. and T.W.; resources, T.W.; data curation, T.W.; writing—original draft preparation, T.W.; writing—review and editing, A.N. and T.W.; visualization, A.N. and T.W.; supervision, T.W.; project administration, T.W.; funding acquisition, M.I., A.N. and T.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Abbreviations

The following abbreviations are used in this manuscript:

FGLS	Feasible generalized least squares
GLS	Generalized least squares
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MSE	Mean squared error
OLS	Ordinary least squares
SNR	Signal-to-noise ratio
TV-AR	time-varying autoregressive
TV-VAR	time-varying vector autoregressive
TV-VEC	time-varying vector error correction
VAR	Vector autoregressive

Appendix A. The Summary of GLS-Kalman Smoother Equivalence

Proposition A1. *(Sant (1977); and else) The GLS estimator of model (10) yields the Kalman-smoothed estimates (8) and its mean squared error matrix (9).*

Proof of Proposition A1.

Lemma A1. $\left(S - TU^{-1}V\right)^{-1} = S^{-1} + S^{-1}T\left(U - VS^{-1}T\right)^{-1}VS^{-1}$ *provided S^{-1} exists.*
□

The GLS estimate of β is

$$
\begin{aligned}
\widehat{\beta}_{GLS} &= \left[\begin{bmatrix} Z'H'^{-1/2} & C'^{-1}Q'^{-1/2} \end{bmatrix}\begin{bmatrix} H^{-1/2}Z \\ Q^{-1/2}C^{-1} \end{bmatrix}\right]^{-1} \\
&\quad \times \begin{bmatrix} Z'H'^{-1/2} & -C'^{-1}Q'^{-1/2} \end{bmatrix}\begin{bmatrix} H^{-1/2}Y_T \\ -Q^{-1/2}b_0^* \end{bmatrix} \\
&= \left(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\right)^{-1}\left(Z'H^{-1}Y_T + C^{-1'}Q^{-1}b_0^*\right) \\
&= \left[CQC' - CQC'Z'\Omega^{-1}ZCQC'\right]\left(Z'H^{-1}Y_T + C^{-1'}Q^{-1}b_0^*\right) \qquad \text{(A1)} \\
&= CQC'Z'\Omega^{-1}Y_T + \left[C - CQC'Z'\Omega^{-1}ZC\right]b_0^* \\
&= Cb_0^* + CQC'Z'\Omega^{-1}(Y_T - ZCb_0^*).
\end{aligned}
$$

Here, we used Lemma,

$$
\left(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\right)^{-1} = CQC' - CQC'Z'\left(H + Z^*CQC'Z'\right)^{-1}ZCQC'.
$$

From (A2), the conditional variance of $\widehat{\beta}_{GLS}$ is

$$Var\left(\widehat{\beta}_{GLS}|Y_T\right) = \left(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\right)^{-1}$$
$$= CQC' - CQC'Z'\Omega^{-1}ZCQC'.$$

Appendix B. Model with Time-Invariant Intercepts

While our model, (1) and (2), and its matrix formulation, (3) and (4), are flexible enough to admit time-varying coefficients, it is sometimes assumed that the class of TV-AR models has time-invariant intercepts. For the purpose of deriving the likelihood function, here, we modify our model to admit time-invariant intercepts. Suppose we have a $k \times k$ vector of time-invariant intercepts, v, in our model. Then, (1) and (2) become

$$y_t = v + Z_t\beta_t + \varepsilon_t \tag{A2}$$
$$\beta_t = \beta_{t-1} + \eta_t. \tag{A3}$$

In this case, it is convenient to use a matrix form to derive the likelihood function. With the vector of intercepts, our model in matrix form, (3) and (4), is then modified to

$$Y_T = \mathcal{I}v + Z\beta + \varepsilon \tag{A4}$$
$$\beta = C(b_0^* + \eta), \tag{A5}$$

where

$$\mathcal{I} = \begin{bmatrix} I_k & I_k & \cdots & I_k \end{bmatrix}',$$

and I_k is a $k \times k$ identity matrix. Similar to our assumption that time-varying intercepts, if they exist, are unknown, we assume that the vector of time-invariant intercepts, v, is the unknown parameter vector.

From our matrix formulation of (A4) and (A5), the log-likelihood function for Y_T given the covariance matrices of the errors (H and Q), intercept (v), and initial value vector (b_0^*) is

$$\log p(Y_T|H, Q, v, b_0^*) = -\frac{(T-p)k}{2}\log 2\pi - \frac{1}{2}\log|\Omega|$$
$$-\frac{1}{2}(Y_T - ZCb_0^* - \mathcal{I}v)'\Omega^{-1}(Y_T - ZCb_0^* - \mathcal{I}v). \tag{A6}$$

Appendix B.1. The GLS Estimator under the Presence of Time-Invariant Intercepts

As we discuss in the previous section, our model admits time-invariant intercepts. Therefore, it is straightforward to define the GLS estimator for such models. To do so, assuming that the time-invariant intercepts are unknown, let us define the vector of unknown parameters, $\beta^* = \begin{bmatrix} v' & \beta' \end{bmatrix}'$. Then, the matrix form for regression that is analogous to (10) is

$$\begin{bmatrix} Y_T \\ -b_0^* \end{bmatrix} = \begin{bmatrix} \mathcal{I} & Z \\ 0 & -C^{-1} \end{bmatrix}\beta^* + \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}. \tag{A7}$$

Here, one of the advantages of utilizing the regression approach (A7) over Kalman smoothing (A5) is that the unknown intercept vector v is estimated simultaneously with β. Then, it can be shown that the GLS estimate $\widehat{v}$ is indeed the maximum likelihood estimate.

Proposition A2. *The GLS estimate $\widehat{v}$ of model (A7) is the maximum likelihood estimate (MLE) of (A5), $\widehat{v}_{ML}$ conditional on $H, Q,$ and b_0^*.*

Proof. From the likelihood function, (A6), the normal equations pertaining to v are

$$\mathcal{I}'\Omega^{-1}(Y_T - ZCb_0^* - \mathcal{I}\widehat{v}_{ML}) = 0.$$

Therefore, the MLE for v is

$$\widehat{v}_{ML} = \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}(Y_T - ZCb_0^*). \tag{A8}$$

Now, the GLS estimates for β^* in model (A7) are

$$
\begin{aligned}
\widehat{\beta}^* &= \begin{bmatrix} \widehat{v} \\ \widehat{\beta} \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{I}'H^{-1}\mathcal{I} & \mathcal{I}'H^{-1}Z \\ Z'H^{-1}\mathcal{I} & Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1} \end{bmatrix}^{-1} \left[\begin{pmatrix} \mathcal{I}'H^{-1} \\ Z'H^{-1} \end{pmatrix} Y_T + \begin{pmatrix} O \\ C^{-1\prime}Q^{-1} \end{pmatrix} b_0^* \right].
\end{aligned} \tag{A9}
$$

Using the Lemma, we arrive at the following (see Appendix B.2 "Detailed Proof of Proposition A2" for details):

$$\widehat{v} = \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}(Y_T - ZCb_0^*).$$

This proves $\widehat{v}_{ML} = \widehat{v}$. □

Proposition A3. *The GLS estimate $\widehat{\beta}$ of model (A7) is the Kalman-smoothed estimate of model (A5).*

Proof. Thanks to the intercept, the Kalman-smoothed estimate is now

$$\widetilde{\beta} = Cb_0^* + CQC'Z'\Omega^{-1}(Y_T - \mathcal{I}v - ZCb_0^*). \tag{A10}$$

From (A9), it follows that

$$\widehat{\beta} = Cb_0^* + CQC'Z'\Omega^{-1}(Y_T - \mathcal{I}\widehat{v} - ZCb_0^*).$$

We prove the equivalence. □

It is clear that the GLS-based approach can compute the Kalman-smoothed β and estimate the unknown intercepts, v, simultaneously. The next question is how we can obtain the statistical inference about $\widehat{\beta}$. More precisely, at issue is whether the GLS-based approach yields the same MSE as the Kalman smoother. The answer to this question is negative for $\widehat{\beta}$.

Proposition A4. *The mean squared error of the Kalman smoothed estimate is*

$$Var(\beta|Y_T) = CQC' - CQC'Z'\Omega^{-1}ZCQC', \tag{A11}$$

whereas the variance estimated from the GLS-based approach (A9) is

$$Var\left(\widehat{\beta}\right) = CQC' - CQC'Z'\Omega^{-1}ZCQC + CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZCQC'. \tag{A12}$$

Proof. See Appendix B.2 "Detailed Proof of Proposition A2" below. □

The difference between the Kalman-smoothed $Var(\beta|Y_T)$ and the GLS-based variance $Var\left(\widehat{\beta}\right)$ is $CQC'Z'\Omega^{-1}\mathcal{I}(\mathcal{I}'\Omega^{-1}\mathcal{I})^{-1}\mathcal{I}'\Omega^{-1}ZCQC'$, which pertains to the estimation of v. If we did not have to estimate v (as we assume for the Kalman-smoothed estimate), $Var(\beta|Y_T)$ and $Var\left(\widehat{\beta}\right)$ would be the same. In other words, if v is known, the MSE of the Kalman-smoothed estimate is the same as the variance of the GLS-based estimate. As a matter of fact, if $Var(\widehat{v}) = \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1} = 0$, the two estimates would be identical. This result reflects that the two approaches yield the same estimate and MSE, as in Proposition A1. Nevertheless, what is important here is that we can obtain (A11) by utilizing the estimated variance of $\widehat{\beta}^*$ of (A9). More specifically, we can estimate the MSE of the Kalman-smoothed estimate by

$$Var(\beta|Y_T) = Var\left(\widehat{\beta}\right) - Cov\left(\widehat{\beta}, \widehat{v}\right)Var(\widehat{v})^{-1}Cov\left(\widehat{\beta}, \widehat{v}\right)'.$$

Appendix B.2. Detailed Proof of Propositions A2

Lemma A2. *If G^{-1} and the inverse of $F = A - BG^{-1}E$ exist,*

$$\begin{pmatrix} A & B \\ E & G \end{pmatrix}^{-1} = \begin{pmatrix} F^{-1} & -F^{-1}BG^{-1} \\ -G^{-1}EF^{-1} & G^{-1} + G^{-1}EF^{-1}BG^{-1} \end{pmatrix}.$$

In our case,

$$A = \mathcal{I}'H^{-1}\mathcal{I}, \ B = \mathcal{I}'H^{-1}Z, G = Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}, E = Z'H^{-1}\mathcal{I};$$

and

$$\begin{aligned} F &= \mathcal{I}'H^{-1}\mathcal{I} - \mathcal{I}'H^{-1}Z\Big(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\Big)^{-1}Z'H^{-1}\mathcal{I} \\ &= \mathcal{I}'\Big[H^{-1} - H^{-1}Z\Big(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\Big)^{-1}Z'H^{-1}\Big]\mathcal{I}, \end{aligned}$$

whose inverse is

$$\begin{aligned} F^{-1} &= \Big\{\mathcal{I}'\Big[H^{-1} - H^{-1}Z\Big(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\Big)^{-1}Z'H^{-1}\Big]\mathcal{I}\Big\}^{-1} \\ &= \Big[\mathcal{I}'(H + ZCQC'Z')^{-1}\mathcal{I}\Big]^{-1} \\ &= \Big(\mathcal{I}'\Omega^{-1}\mathcal{I}\Big)^{-1}. \end{aligned}$$

Other useful equations are

$$\begin{aligned} G^{-1} &= \Big(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\Big)^{-1} \\ &= CQC' - CQC'Z'(H + ZCQC'Z')^{-1}ZCQC' \\ &= CQC' - CQC'Z'\Omega^{-1}ZCQC'; \end{aligned}$$

$$\begin{aligned} \Omega^{-1} &= (H + ZCQC'Z')^{-1} \\ &= H^{-1} - H^{-1}Z\Big(Z'H^{-1}Z + C'^{-1}Q^{-1}C^{-1}\Big)^{-1}Z'H^{-1} \\ &= H^{-1} - H^{-1}ZG^{-1}Z'H^{-1}. \end{aligned} \tag{A13}$$

Then, for (A9), we arrive at

$$\widehat{v} = \underbrace{\Big(F^{-1}\mathcal{I}'H^{-1} - F^{-1}BG^{-1}Z'H^{-1}\Big)}_{(\text{i})}Y_T - \underbrace{F^{-1}BG^{-1}C^{-1\prime}Q^{-1}b_0^*}_{(\text{ii})} \tag{A14}$$

$$\widehat{\beta} = \underbrace{\Big[-G^{-1}EF^{-1}\mathcal{I}'H^{-1} + \Big(G^{-1} + G^{-1}EF^{-1}BG^{-1}\Big)Z'H^{-1}\Big]}_{(\text{iii})}Y_T$$

$$+ \underbrace{\Big(G^{-1} + G^{-1}EF^{-1}BG^{-1}\Big)C^{-1\prime}Q^{-1}b_0^*}_{(\text{iv})}. \tag{A15}$$

(i) $\quad F^{-1}\mathcal{I}'H^{-1} - F^{-1}BG^{-1}Z'H^{-1} = \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}$

$$\times \left(I - ZCQC'Z'H^{-1} + ZCQC'Z'\Omega^{-1}ZCQC'Z'H^{-1}\right)$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}$$

$$\left[I - (\Omega - H)H^{-1} + (\Omega - H)\Omega^{-1}(\Omega - H)H^{-1}\right]$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}$$

$$\left(I - \Omega H^{-1} + I + \Omega H^{-1} - I - I + H\Omega^{-1}\right)$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}.$$

(ii) $\quad F^{-1}BG^{-1}C^{-1\prime}Q^{-1} = \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}Z\left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)C^{-1\prime}Q^{-1}$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}Z\left(C - CQC'Z'\Omega^{-1}ZC\right)$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}\left(ZC - ZCQC'Z'\Omega^{-1}ZC\right)$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}\left(I - ZCQC'Z'\Omega^{-1}\right)ZC$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}\left(I - (\Omega - H)\Omega^{-1}\right)ZC$$

$$= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC.$$

Therefore,

$$\begin{aligned}\widehat{v} &= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}Y_T - \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZCb_0^* \\ &= \left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}(Y_T - ZCb_0^*).\end{aligned}$$

(iii) $\quad -G^{-1}EF^{-1}\mathcal{I}'H^{-1} + \left(G^{-1} + G^{-1}EF^{-1}BG^{-1}\right)Z'H^{-1}$

$$= -G^{-1}EF^{-1}\left(\mathcal{I}'H^{-1} - BG^{-1}Z'H^{-1}\right) + G^{-1}Z'H^{-1}$$

$$= -G^{-1}EF^{-1}\mathcal{I}'\left(H^{-1} - H^{-1}ZG^{-1}Z'H^{-1}\right) + G^{-1}Z'H^{-1}$$

$$= -G^{-1}EF^{-1}\mathcal{I}'\Omega^{-1} + G^{-1}Z'H^{-1} \quad \text{(from (A13))}$$

$$= -G^{-1}Z'H^{-1}\mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1} + G^{-1}Z'H^{-1}$$

$$= -G^{-1}Z'H^{-1}\left(I - \mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1}\right)$$

$$= -\left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)Z'H^{-1}\left(I - \mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1}\right)$$

$$= -CQC'Z'H^{-1}\left(I - \mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1}\right) + CQC'Z'\Omega^{-1}(\Omega - H)H^{-1}\left(I - \mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1}\right)$$

$$= CQC'Z'\Omega^{-1}\left(I - \mathcal{I}F^{-1}\mathcal{I}'\Omega^{-1}\right)$$

$$= CQC'Z'\Omega^{-1} - CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}$$

$$= CQC'Z'\Omega^{-1}\left[I - \mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}\right]$$

(iv) $\left(G^{-1} + G^{-1}EF^{-1}BG^{-1}\right)C^{-1\prime}Q^{-1}$

$= G^{-1}C^{-1\prime}Q^{-1} + G^{-1}EF^{-1}BG^{-1}C^{-1\prime}Q^{-1}$

$= \left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)C^{-1\prime}Q^{-1} + G^{-1}E\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC$ (from (ii))

$= C - CQC'Z'\Omega^{-1}ZC$

$\quad + \left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)Z'H^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC$

$= C - CQC'Z'\Omega^{-1}ZC$

$\quad + CQC'\left(Z'H^{-1} - Z'\Omega^{-1}ZCQC'Z'H^{-1}\right)\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC$

$= C - CQC'Z'\Omega^{-1}ZC$

$\quad + CQC'\left[Z'H^{-1} - Z'\Omega^{-1}(\Omega - H)H^{-1}\right]\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC$

$= C - CQC'Z'\Omega^{-1}ZC + CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZC$

$= \left[I - CQC'Z'\Omega^{-1}Z + CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}Z\right]C.$

Therefore,

$$\begin{aligned}
\widehat{\beta} &= CQC'Z'\Omega^{-1}\left[I - \mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}\right]Y_T \\
&\quad + \left[I - CQC'Z'\Omega^{-1}Z + CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}Z\right]Cb_0^* \quad \text{(A16)}\\
&= Cb_0^* + CQC'Z'\Omega^{-1}(Y_T - \mathcal{I}\widehat{v} - ZCb_0^*).
\end{aligned}$$

The mean squared error matrix

$$Var(\beta|Y_T) = CQC' - CQC'Z'\Omega^{-1}ZCQC'.$$

From (A9) and Lemma, we can show

$$\begin{aligned}
Var\left(\widehat{\beta}\right) &= G^{-1} + G^{-1}EF^{-1}BG^{-1} \\
&= \left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right) + G^{-1}Z'H^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}ZG^{-1} \\
&= CQC' - CQC'Z'\Omega^{-1}ZCQC \\
&\quad + \left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)Z'H^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}Z \\
&\quad \times \left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right) \\
&= CQC' - CQC'Z'\Omega^{-1}ZCQC \\
&\quad + \left(CQC'Z'H^{-1} - CQC'Z'\Omega^{-1}ZCQC'Z'H^{-1}\right)\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}' \\
&\quad \times \left(H^{-1}ZCQC' - H^{-1}ZCQC'Z'\Omega^{-1}ZCQC'\right) \\
&= CQC' - CQC'Z'\Omega^{-1}ZCQC \\
&\quad + \left(CQC'Z'H^{-1} - CQC'Z'\Omega^{-1}(\Omega - H)H^{-1}\right)\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}' \\
&\quad \times \left(H^{-1}ZCQC' - H^{-1}(\Omega - H)'\Omega^{-1}ZCQC'\right) \\
&= CQC' - CQC'Z'\Omega^{-1}ZCQC \\
&\quad + CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZCQC'.
\end{aligned}$$

Note also that

$$
\begin{aligned}
-F^{-1}BG^{-1} &= -\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'H^{-1}Z\left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right) \\
&= -\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\left(H^{-1}ZCQC' - H^{-1}ZCQC'Z'\Omega^{-1}ZCQC'\right) \\
&= -\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\left(H^{-1}ZCQC' - H^{-1}(\Omega - H)\Omega^{-1}ZCQC'\right) \\
&= -\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}\mathcal{I}'\Omega^{-1}ZCQC',
\end{aligned}
$$

and

$$
\begin{aligned}
-G^{-1}EF^{-1} &= -\left(CQC' - CQC'Z'\Omega^{-1}ZCQC'\right)Z'H^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1} \\
&= -\left(CQC'Z'H^{-1} - CQC'Z'\Omega^{-1}ZCQC'Z'H^{-1}\right)\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1} \\
&= -\left(CQC'Z'H^{-1} - CQC'Z'\Omega^{-1}(\Omega - H)H^{-1}\right)\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1} \\
&= -CQC'Z'\Omega^{-1}\mathcal{I}\left(\mathcal{I}'\Omega^{-1}\mathcal{I}\right)^{-1}.
\end{aligned}
$$

Therefore,

$$
Var(\beta|Y_T) = Var\left(\widehat{\beta}\right) - Cov\left(\widehat{\beta},\widehat{v}\right)Var(\widehat{v})^{-1}Cov\left(\widehat{\beta},\widehat{v}\right)'.
$$

Appendix C. TV-VAR(2) with Time-Varying Intercepts

VAR(2) Case: p = 2 (i.e., 2 Lags) and k = 3 (i.e., 3 Variables)

To make the matrix Z, first define

$$
\begin{aligned}
Z_t &= \underbrace{\left([1, y'_{t-1}, y'_{t-2}] \otimes I_k\right)}_{k \times (pk+1)k} \\
&= \left([1, y'_{t-1}, y'_{t-2}] \otimes I_k\right).
\end{aligned}
$$

Then,

$$
\underbrace{Z}_{k(T-p) \times (pk+1)k(T-p)} =
\begin{bmatrix}
Z_3 & & & 0 \\
& Z_4 & & \\
& & \ddots & \\
0 & & & Z_T
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
[1, y'_2, y'_1] \otimes I_k & & & 0 \\
& [1, y'_3, y'_2] \otimes I_k & & \\
& & \ddots & \\
0 & & & [1, y'_{T-1}, y'_{T-2}] \otimes I_k
\end{bmatrix}
$$

$$
= z \otimes I_k
$$

where

$$
\underbrace{z}_{(T-p) \times (kp+1)(T-p)} =
\begin{bmatrix}
[1, y'_2, y'_1] & & & 0 \\
& [1, y'_3, y'_2] & & \\
& & \ddots & \\
0 & & & [1, y'_{T-1}, y'_{T-2}]
\end{bmatrix}.
$$

For the regression:

$$
\underbrace{\begin{bmatrix} Y_T \\ -b_0^* \end{bmatrix}}_{k(T-p)(kp+2)\times 1} = \underbrace{\begin{bmatrix} Z \\ -C^{-1} \end{bmatrix}}_{k(T-p)(kp+2)\times(kp+1)k(T-p)} \underbrace{\beta^*}_{(T-p)k(kp+1)\times 1} + \underbrace{\begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}}_{k(T-p)(kp+2)\times 1},
$$

one needs to define

$$
X = \begin{bmatrix} Z \\ -C^{-1} \end{bmatrix}.
$$

Then,

$$
\underbrace{X'X}_{(kp+1)k(T-p)\times(kp+1)k(T-p)} = \begin{bmatrix} Z' & -C^{-1\prime} \end{bmatrix} \begin{bmatrix} Z \\ -C^{-1} \end{bmatrix}
$$

$$
= \begin{bmatrix} Z'Z + C^{-1\prime}C^{-1} \end{bmatrix},
$$

where

$$
\underbrace{C}_{(kp+1)k(T-p)\times(kp+1)k(T-p)} = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & I & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ I & I & I & I \end{bmatrix}
$$

$$
= \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{(T-p)\times(T-p)} \otimes I_{(kp+1)k}
$$

$$
= c \otimes I_{(kp+1)k}.
$$

Here,

$$
\underbrace{c}_{(T-p)\times(T-p)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.
$$

The rest of the matrices needed for GLS are:

$$
\underbrace{Y_T}_{(T-p)k\times 1} =
\begin{bmatrix}
y_{p+1} \\
y_{p+2} \\
\vdots \\
y_T
\end{bmatrix},
$$

$$
\underbrace{Z}_{k(T-p)\times(pk+1)k(T-p)} =
\begin{bmatrix}
Z_{p+1} & & & 0 \\
& Z_{p+2} & & \\
& & \ddots & \\
0 & & & Z_T
\end{bmatrix},
$$

$$
\underbrace{\beta}_{(T-p)k(kp+1)\times 1} =
\begin{bmatrix}
\beta_{p+1} \\
\beta_{p+2} \\
\vdots \\
\beta_T
\end{bmatrix}, \quad
\underbrace{\varepsilon}_{(T-p)k\times 1} =
\begin{bmatrix}
\varepsilon_{p+1} \\
\varepsilon_{p+2} \\
\vdots \\
\varepsilon_T
\end{bmatrix},
$$

$$
\underbrace{\eta}_{(T-p)k(kp+1)\times 1} =
\begin{bmatrix}
\eta_{p+1} \\
\eta_{p+2} \\
\vdots \\
\eta_T
\end{bmatrix},
$$

$$
\underbrace{C}_{(kp+1)kp(T-p)\times(kp+1)k(T-p)} =
\begin{bmatrix}
I & 0 & \cdots & 0 \\
I & I & & \vdots \\
\vdots & \vdots & \ddots & 0 \\
I & I & I & I
\end{bmatrix},
$$

$$
\underbrace{Q}_{(T-p)k(kp+1)\times(T-p)k(kp+1)} =
\begin{bmatrix}
Q_{p+1} & & & 0 \\
& Q_{p+2} & & \\
& & \ddots & \\
0 & & & Q_T
\end{bmatrix},
$$

$$
\underbrace{H}_{(T-p)k\times(T-p)k} =
\begin{bmatrix}
H_{p+1} & & & 0 \\
& H_{p+2} & & \\
& & \ddots & \\
0 & & & H_T
\end{bmatrix},
$$

$$
\underbrace{b_0^*}_{(T-p)k(kp+1)\times 1} =
\begin{bmatrix}
b_0 \\
0 \\
\vdots \\
0
\end{bmatrix}, \quad
P_0^* =
\begin{bmatrix}
P_0 & 0 & & 0 \\
0 & 0 & & \\
\vdots & & \ddots & \\
0 & 0 & \cdots & 0
\end{bmatrix}.
$$

Notes

[1] An alternative to those two methods is the approach presented by Cooley and Prescott (1976), who use the likelihood method to estimate the unknown parameters rather than Kalman filtering.

[2] Related to our approach of not using Kalman filtering, McCausland et al. (2011) develop and propose a new simulation smoothing approach which is more computationally efficient than the approach based on Kalman filtering. While we pay little attention to computational efficiency in this paper, evaluating computation costs along with estimation accuracy should be further investigated in later studies.

[3] Ito et al. (2014, 2016) do not formally prove that their regression-based approach generates estimates that are equivalent to Kalman-smoothed estimates.

[4] As our model include unknown parameters such as the variances of the error terms, we must rely on feasible GLS (FGLS), which may not be equivalent to GLS.

[5] In this paper, we focus on the case where ε_t and η_t are mutually uncorrelated. Relaxing this assumption poses a great challenge.

6 This assumption does not change our conclusions below. The main difference is that $Var(\beta) = C(P_0^* + Q)C'$ and $Var(Y_T) = ZC(P_0^* + Q)C'Z' + H = \Omega$. An exception is when the diffuse prior is used and the likelihood function is computed excluding the first few observations. In such a case, the estimates of the unknown intercept parameters under the two approaches would differ.

7 By contrast, Duncan and Horn (1972) assume that matrix F is known, which renders their estimation impractical. The original form of Maddala and Kim (1998, pp. 469–70) is similar to ours, but it is a general form for a scalar y_t. Hence, it does not aim to deal with the autoregressive part of time-varying parameter models nor consider vector processes.

8 For the Bayesian approach, we focus on the posterior mean from MCMC. Since our simulations are based on Primiceri's (2005) model, we use the same priors as his. The Matlab codes provided by D. Korobilis are used, which can be downloaded from: https://drive.google.com/file/d/1pYNP96FeGgBH1KpnDEEdXGqZ62ZPw_PQ/view, accessed on 14 March 2022.

9 In addition, we can compute the values of the log-likelihood function to evaluate whether the repeated use of FGLS improves estimation accuracy. Our simulation tends to show that 2FGLS has a higher likelihood value than 1FGLS.

10 After computing $\widehat{\beta}_{t,i,n}$, we discard first 40 ($t = 1, \dots, 40$) of them in accordance with Korobilis's Matlab codes for the Bayesian MCMC method. Hence, we use $T = length\ of\ the\ data\ -40$ to compute the sample moments of $\widehat{\beta}$ by (11) through (18).

11 Throughout this simulation study, we use bold numbers to highlight the best (the smallest median *dist* and the median *rat* closest to one) estimation method of the four approaches (OLS, 1FGLS, 2FGLS, 2FGLS' and Primiceri).

12 We use the data and MATLAB codes provided by Koop and Korobilis (2010).

13 Note that the impulse responses of Primiceri's (2005) VAR vary largely over time. This is not because the time-varying parameters (β_t) are very volatile over time, but mainly because the variance of the shocks are time-dependent and vary greatly, as shown in Figure 1 of Primiceri (2005, p. 832) and as discussed in the conclusion thereof.

References

Bernanke, Ben S., and Ilian Mihov. 1998. Measuring monetary policy. *Quarterly Journal of Economics* 113: 869–902. [CrossRef]

Chan, Joshua C. C., and Ivan Jeliazkov. 2009. Efficient simulation and integrated likelihood estimationin state space models. *International Journal of Mathematical Modelling and Numerical Optimisation* 1: 101–20. [CrossRef]

Cogley, Timothy F., and Thomas J. Sargent. 2001. Evolving post-world war II u.s. inflation dynamics. *NBER Macroeconomics Annual* 16: 331–73. [CrossRef]

Cogley, Timothy F., and Thomas J. Sargent. 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* 8: 262–302. [CrossRef]

Cogley, Timothy F., and Edward C. Prescott. 1976. Estimation in the presence of stochastic parameter variation. *Econometrica* 44: 167–84.

Duncan, David B., and Susan D. Horn. 1972. Linear dynamic recursive estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association* 67: 815–21. [CrossRef]

Durbin, James, and Siem J. Koopman. 2012. *Time Series Analysis by State Space Methods*, 2nd ed. Oxford: Oxford University Press.

Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–84. [CrossRef]

Hansen, Bruce E. 1992. Testing for parameter instability in linear models. *Journal of Policy Modeling* 14: 517–33. [CrossRef]

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge and New York: Cambridge University Press.

Ito, Mikio, Akihiko Noda, and Tatsuma Wada. 2014. International stock market efficiency: A non-bayesian time-varying model approach. *Applied Economics* 46: 2744–754. [CrossRef]

Ito, Mikio, Akihiko Noda, and Tatsuma Wada. 2016. The evolution of stock market efficiency in the us: A non-bayesian time-varying model approach. *Applied Economics* 48: 621–35. [CrossRef]

Ito, Mikio, Akihiko Noda, and Tatsuma Wada. 2021. Time-varying comovement of foreign exchange markets: A GLS-based time-varying model approach. *Mathematics* 9: 849.

Koop, Gary, and Dimitris Korobilis. 2010. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* 3: 267–358. [CrossRef]

Koopman, Siem J. 1997. Exact initial kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* 92: 1630–638. [CrossRef]

Maddala, Gangadharrao S., and In-Moo Kim. 1998. *Unit Roots, Cointegration, and Structural Change*. Cambridge, New York: Cambridge University Press.

McCausland, William J., Shirley Miller, and Denis Pelletier. 2011. Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics & Data Analysis* 55: 199–212.

Perron, Pierre. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57: 1361–401. [CrossRef]

Perron, Pierre, and Tatsuma Wada. 2009. Let's take a break: Trends and cycles in us real gdp. *Journal of Monetary Economics* 56: 749–65. [CrossRef]

Primiceri, Giorgio E. 2005. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* 72: 821–52. [CrossRef]

Sant, Donald T. 1977. Generalized least squares applied to time varying parameter models. *Annals of Economic and Social Measurement* 6: 301–14.

Sargan, Dennis J., and Alok Bhargava. 1983. Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. *Econometrica* 51: 799–820. [CrossRef]

Shephard, Neil G., and Andrew C. Harvey. 1990. On the probability of estimating a deterministic component in the local level model. *Journal of Time Series Analysis* 11: 339–347. [CrossRef]

Tanaka, Katsuto. 2017. *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, 2nd ed. Hoboken: John Wiley & Suns, Inc.

Taylor, Stephen J. 2007. *Modelling Financial Time Series*, 2nd ed. Hackensack: World Scientific.