Received: 4 January 2023 | Revised: 19 June 2023 | Accepted: 29 July 2023

DOI: 10.1002/sim.9870

RESEARCH ARTICLE

Statistics in Medicine WILEY

Forecasting emergency department waiting time using a state space representation

Kelly Trinh[1,2] | **Andrew Staib**[3,4] | **Anton Pak**[5,6]

[1]Data61, The Commonwealth Scientific and Industrial Research Organisation, Clayton, Victoria, Australia

[2]College of Science and Engineering, James Cook University, Douglas, Queensland, Australia

[3]Faculty of Medicine, University of Queensland, Brisbane, Woolloongabba, Queensland, Australia

[4]Emergency Department, Princess Alexandra Hospital, Woolloongabba, Queensland, Australia

[5]Centre for the Business and Economics of Health, The University of Queensland, Brisbane, Queensland, Australia

[6]Australian Institute of Tropical Health and Medicine, James Cook University, Douglas, Queensland, Australia

Correspondence
Kelly Trinh, Data61, The Commonwealth Scientific and Industrial Research Organisation, Clayton, VIC, Australia.
Email: kellytrinh89@gmail.com

Funding information
The Emergency Medicine Foundation (Australasia) Queensland Program, Grant/Award Number: EMJS-343R33-2020-STAIB; University of Queensland Health and Behavioural Science Faculty Early Career Academic Research Accelerator Award

The provision of waiting time information in emergency departments (ED) has become an increasingly popular practice due to its positive impact on patient experience and ED demand management. However, little scientific attention has been given to the quality and quantity of waiting time information presented to patients. To improve both aspects, we propose a set of state space models with flexible error structures to forecast ED waiting time for low acuity patients. Our approach utilizes a Bayesian framework to generate uncertainties associated with the forecasts. We find that the state-space models with flexible error structures significantly improve forecast accuracy of ED waiting time compared to the benchmark, which is the rolling average model. Specifically, incorporating time-varying and correlated error terms reduces the root mean squared errors of the benchmark by 10%. Furthermore, treating zero-recorded waiting times as unobserved values improves forecast performance. Our proposed model has the ability to provide patient-centric waiting time information. By offering more accurate and informative waiting time information, our model can help patients make better informed decisions and ultimately enhance their ED experience.

KEYWORDS
Bayesian state space model, emergency department waiting time, MCMC

1 | INTRODUCTION

Long waiting times in an emergency department (ED) have become of growing public policy concern over the recent years. These extended waits not only result in suboptimal health outcomes but also impact patients' satisfaction with the healthcare services they receive.[1,2] Furthermore, waiting times and queue length are associated with patients' decisions to leave without being seen, which can have negative health consequences.[3] To address waiting time uncertainty and improve patient experience, waiting time communication interventions, such as posting ED waiting time predictions

online or within the ED, have been proposed.[4] However, concerns remain regarding the quality of information reported and how patients perceive waiting time information.[5]

Providing accurate waiting time forecasts has been methodologically challenging due to the complex nature of clinical ED environment and significant variations in waiting time. Few hospitals provide waiting time predictions to patients, and those that do often rely on a simple rolling average, which provides relatively inaccurate forecasts.[6,7] Several studies have used statistical models and machine learning algorithms to improve the accuracy of individual waiting time predictions. For example, Lin et al[8] used queuing theory to model waiting time and evaluated impacts of queue growth and ED resource allocation on waiting times. Sun et al[9] used a quantile regression to account for large waiting time fluctuations in their forecasts. Ang et al[6] proposed a Lasso regression with variables from a fluid model, termed as "Q-Lasso," which outperformed the commonly used rolling average model by over 30% in mean squared error. Similarly, Pak et al[7] found that a Lasso-based model and a random forest approach generated more accurate waiting time predictions than the rolling average. Using queue-based features in non-linear models such as random forest can also benefit the forecast accuracy.[10]

While the previous studies have primarily focused on predicting individual waiting times for patients, particularly those with lower acuity levels, our study seeks to generate aggregated ED waiting time forecasts over a specific time window. This approach involves providing a single waiting time forecast with uncertainties to low acuity patients who arrive during a particular time frame. Patients in this category are typically seen in order of arrival, and are more likely to consider waiting time as a crucial factor in their decision-making process regarding where and when to seek healthcare services. Furthermore, providing an aggregated ED waiting time forecast offers an implementation advantage for hospitals. Displaying a single waiting time prediction in the ED waiting room is more feasible than delivering individualized predictions to each patient by an ED nurse.

In addition to forecasting aggregated ED waiting time, we make several methodological and practical contributions to the ED waiting time literature. First, it is often observed that many *zero-recorded* observations exist in aggregated ED waiting time dataset.* In practice, the *zero-recorded* observations within a time-window do not necessarily imply that patients will experience zero waiting time if they arrive to ED. The *zero-recorded* observations may reflect a situation where no patients were called in for treatment during that particular time window rather than an indication of no patients waiting. Therefore, treating *zero-recorded* observations as actual zeros not only misrepresents the actual waiting time, but also leads to significant fluctuations in ED waiting times. These fluctuations might affect the goodness-of-fit and out-of-sample forecast performance. To avoid these problems, we argue that treating these *zero-recorded* observations as unobserved values is a more appropriate approach.

Second, we present an alternative approach to the commonly used time dummy model for capturing temporal variations in waiting time such as daily, weekly, and monthly patterns. The time dummy approach often results in a large number of dummy variables if multiple periodic patterns are considered. This over-parameterization might lead to poor forecast performance. Furthermore, the marginal effects of time dummy models remain time-invariant. To overcome these limitations, we propose a set of state space (SS) models with flexible error structures, which effectively capture the dynamic behaviors of ED waiting times. These flexible error structures aim to mitigate extreme values and potential serial correlation among waiting times. Additionally, we introduce efficient Markov chain Monte Carlo (MCMC) algorithms for estimating these models.

Third, we make a significant contribution to the literature by generating forecasts for multiple periods ahead using SS models and evaluating their accuracy performance. Unlike the previous studies that solely concentrated on predicting individual patient waiting times for the next time interval (ie, one-period-ahead forecast), our modeling approach encompasses multiple-period-ahead forecasts. Multiple-period-ahead forecasts have important practical benefits in improving ED waiting time management and responding to the ED demand more effectively.

Lastly, we use a Bayesian approach to explore the predictive posterior distributions of the forecasted waiting times. This approach enables us to provide measures of uncertainty, such as probabilities. From a practical standpoint, relying solely on point forecasts of ED waiting times can be uninformative and potentially misleading. Hospitals are unable to guarantee that a patient will be seen within the predicted waiting time, and studies in the literature on patient experience with ED services have shown that patient dissatisfaction tends to increase when they wait longer than previously indicated.[11-13] Therefore, studying the distribution of waiting time forecasts is crucial, as ED waiting times are inherently uncertain and asymmetric due to the complex nature of clinical pathways and the unpredictability of arrivals of urgent and critically-ill patients, which require sudden diversion of significant ED resources. By generating and conveying uncertainties associated with the waiting time forecasts to patients, we can potentially provide better alternatives for hospitals and improve the overall patient experience.

This study is structured as follows. In Section 2, we present a set of model specifications with flexible error structures. In Section 3, we provide an overview of the data used in this article. In Section 4, we discuss the in-sample and out-of-sample forecast results. Finally, Section 5 concludes the article.

2 | STATE SPACE SPECIFICATION

Our model specification is based on a SS model, also known as a dynamic linear model. This modeling approach has proven to be a powerful tool in capturing unobserved processes and offers flexibility in handling missing values.[14] In fact, SS models have been implemented in various research areas such as electrical engineering,[15] physics,[16] business,[17] economics,[18] and ecology.[19] Moreover, the use of SS models in applied health research has been growing over the years. For instance, Nobre et al[20] compared the forecast performance of an SS model with a seasonal autoregressive integrated moving average (MA) model in predicting the incidence of hepatitis A and malaria cases. They found that both models performed similarly in short and long-term forecasts, although the SS model offered greater flexibility in capturing diverse types of time series patterns. Similarly, Fukaya et al[21] and Kawamori et al[22] utilized an SS model to measure base body temperature and forecast menstruation periods. Furthermore, Christensen et al[23] employed an SS model to capture seasonal variations in hospitalization rates for stroke.

Within a SS framework, we decompose the logarithm of waiting time y_t into a conditional mean process μ_t and an error process ε_t^y as follows:

$$y_t = \mu_t + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \sigma_y^2); \tag{1}$$

where y_t is comprised of the logarithm of observed waiting time, denoted as y_t^o, as well as the logarithm of "unobserved" waiting time, denoted as y_t^u. The error process ε_t^y is assumed to follow a normal distribution with a mean of 0 and a constant variance of σ_y^2. Later, we will discuss some flexible error structures for ε_t^y. Equation (1) is commonly referred to as an observation equation.

While the time-varying conditional mean μ_t in (1) is unspecified, the model (1) can incorporate several well-known specifications with an appropriate conditional mean process, such as:

1. A rolling average

$$\mu_t = \frac{1}{k} \sum_{i=t-k+1}^{t} y_i; \tag{2}$$

   where the conditional mean process is computed as an unweighted average of k waiting time observations. When σ_y^2 is constrained to 0, this model is referred to as a rolling average. The estimation of k is based on the smallest root mean-squared error within the in-sample data.
2. An unobserved component (UC) model:

$$\begin{aligned} \mu_t &= \tau_t, \\ \tau_t &= \tau_{t-1} + \varepsilon_t^\tau, \qquad \varepsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2); \end{aligned} \tag{3}$$

   where the initial value τ_0 is an unknown parameter and will be estimated. Equation (3) is commonly referred to as a state equation. In comparison to the rolling average model, the UC model offers greater flexibility by accounting for measurement errors in the data and potential model misspecifications. The conditional mean process τ_t is assumed to follow a random walk process. We select the UC model as the benchmark for the density forecast measure since the density forecast measure is not applicable to the rolling average model.
3. A time dummy (TD) model:

$$\mu_t = \gamma_0 + \mathbf{D}\boldsymbol{\gamma};$$

   where γ_0 represents the intercept, $\mathbf{D}$ is a $1 \times m$ matrix of dummy variables that could include days of the week and/or months of the year, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)'$ is a $m \times 1$ coefficient associated with $\mathbf{D}$. This model can accommodate a

Lasso regression if a penalty is applied to shrink some coefficients towards zero to deal with the over-parameterization, particularly when a large number of dummy variables are used.[7,9]

4. A time-varying auto-regressive model (TVAR)

$$\begin{aligned}\mu_t &= \beta_{0t} + y_{t-1}\beta_{1t} + y_{t-2}\beta_{2t} + \ \dots\ + y_{t-p}\beta_{pt},\\ &= \mathbf{X}_t\boldsymbol{\beta}_t,\\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\epsilon}_t^{\beta}, \qquad \boldsymbol{\epsilon}_t^{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\beta});\end{aligned} \tag{4}$$

where $\boldsymbol{\beta}_t = [\beta_{0t}, \beta_{1t}, \beta_{2t}, \dots, \beta_{pt}]'$, $\mathbf{X} = [1, y_{t-1}, \dots, y_{t-p}]$, and $\boldsymbol{\Sigma}_{\beta} = \text{diag}(\sigma^2_{\beta_0}, \sigma^2_{\beta_1}, \dots, \sigma^2_{\beta_p})$. The initial condition $\boldsymbol{\beta}_0 = [\beta_{00}, \beta_{10}, \dots, \beta_{p0}]$ is unknown and will be estimated. The TVAR model captures the conditional mean waiting time by utilizing past information of the waiting time.

If one sets $\boldsymbol{\Sigma}_{\beta}$ equal to 0, that is, $\sigma^2_{\beta_0} = \sigma^2_{\beta_1} = \dots = \sigma^2_{\beta_p} = 0$ in (4), a TVAR model becomes a constant auto-regressive (AR) model:

$$\mu_t = \beta_0 + \beta_1 y_{t-1} + \ \dots\ + \beta_p y_{t-p}.$$

Regarding the error process ϵ_t^y, imposing a normal distribution on ϵ_t^y can be restrictive, considering that waiting time exhibits significant fluctuations over time, which can be attributed to variations in both the mean process and the error process. In this article, we address this limitation by introducing various specifications for the error process ϵ_t^y.

It is well-known in the statistical literature that normal distributions have exponentially decaying tails, and therefore they have little mass for extreme values.[24,25] In the context of ED waiting time, patients, especially those who do not require urgent medical treatment, may experience significant waiting time, particularly when the ED is overwhelmed by an influx of emergency patients. To account for the potential extreme values, we incorporate a heavy-tailed distribution (ie, a t-distributed error) as the first specification for the error process, which can be expressed as follows:

$$\epsilon_t^y \sim \mathcal{N}(0, \lambda_t\sigma_y^2), \quad \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2). \tag{5}$$

The t-distributed error is expressed as a scale mixture of Gaussian distributions, which simplifies model estimation via data augmentation. When the scale mixture parameter λ_t is integrated out, the marginal distribution of the error term is t-distributed.[26] The scale mixture parameters $\lambda_1, \dots, \lambda_t$ are assumed to be independent and follow the inverse gamma distribution $\mathcal{IG}(\nu/2, \nu/2)$, where ν is the degrees of freedom parameter. The degrees of freedom is treated as an unknown parameter and has a uniform prior, that is, $\nu \sim \mathcal{U}(2, V)$. We restrict ν to be greater than 2 to ensure that the first and second moments of the t distribution exists.

To capture the high variability of ED waiting time, the second specification of the error process allows the variance of ϵ_t^y to vary over time using the stochastic volatility (SV) model[27†]:

$$\begin{aligned}\epsilon_t^y &\sim \mathcal{N}(0, e^{h_t}),\\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \epsilon_t^h, \quad \epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2);\end{aligned} \tag{6}$$

The log-volatility h_t follows a stationary AR(1) process with $|\phi_h| < 1$. The log-volatility process is initialized by drawing h_1 from a normal distribution, that is, $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2/(1-\phi_h^2))$. The prior on variance σ_h^2 follows a standard inverse gamma distribution, that is, $\sigma_h^2 \sim \mathcal{IG}(\alpha_h, \eta_h)$, while the prior on the mean log-volatility μ_h follows a normal distribution, that is, $\mathcal{N}(m_{\mu_h}, V_{\mu_h})$. To ensure the stationary process we assign a truncated normal prior to ϕ_h, that is, $\phi_h \sim \mathcal{N}(\phi_{h_0}, V_{\phi_h})\mathbf{1}(|\phi_h| < 1)$.

The third specification of the error process ϵ_t^y accounts for autocorrelation using a MA specification. It is widely acknowledged in the time series literature that neglecting serial correlation in model specifications can lead to imprecise estimates and inaccurate statistical inferences and predictions.[30,31] Without loss of generality, we model the error process using an MA(1) process:

$$\epsilon_t^y = u_t + \psi u_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma_y^2). \tag{7}$$

The $|\psi|$ is restricted to less than 1 to ensure the invertibility condition for the MA process. This representation is motivated by the Wold decomposition theorem, which establishes that any zero-mean covariance-stationary time series has an infinite MA representation.[32] In other words, such a process can be approximated arbitrarily well by a sufficiently high order of autoregressive MA model. We use a truncated Gaussian prior for ψ, specifically $\psi \sim \mathcal{N}(0, V_\psi)\mathbf{1}(|\psi| < 1)$, to ensure the stationary condition.

By combining these three key specifications of the error term process, we further generate four additional specifications for the error ε_t^y:

- tSV: incorporating t-distributed error in (5) with SV component in (6):

$$\begin{aligned} \varepsilon_t^y &\sim \mathcal{N}(0, \lambda_t e^{h_t}), & \lambda_t &\sim \mathcal{IG}(\nu/2, \nu/2), \\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2); \end{aligned} \tag{8}$$

- tMA: incorporating t-distributed error in (5) with a MA component in (7):

$$\varepsilon_t^y = u_t + \psi u_{t-1}, \qquad u_t \sim \mathcal{N}(0, \lambda_t \sigma_y^2), \qquad \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2); \tag{9}$$

- MASV: including both SV in (6) and MA in (7) components:

$$\begin{aligned} \varepsilon_t^y &= u_t + \psi u_{t-1}, & u_t &\sim \mathcal{N}(0, e^{h_t}), \\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2); \end{aligned} \tag{10}$$

- tMASV: including all three components in (5) to (7):

$$\begin{aligned} \varepsilon_t^y &= u_t + \psi u_{t-1}, & u_t &\sim \mathcal{N}(0, \lambda_t e^{h_t}), \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2), \\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2). \end{aligned} \tag{11}$$

Table 1 presents the list of candidate models that we consider in this article, which includes a rolling average model, as well as the models that specify the conditional time-varying mean process as TD, UC, and TVAR. For the error process, we consider a homoscedastic error (ie, a Gaussian distribution specified in Equation (1)), as well as seven additional specifications presented in Equations (5) to (11) for the TD, UC, and TVAR models. We also estimate the models where *zero-recorded* observations are treated as unknown, as well as the models where the observations are treated as *actual* zeros. In total, we estimate 41 models. The model parameters are estimated using MCMC algorithms, with details of the Bayesian estimation provided in Appendix B. The programming is conducted using MATLAB.

3 | DATA

Princess Alexandra Hospital (PAH) is a major tertiary urban hospital in Australia that provides care in all major adult specialities except for maternity. Since PAH is a public hospital, visits to the ED are covered by the universal public health insurance scheme called medicare, and patients are not required to pay any out-of-pocket expenses. The ED records contain detailed information on the date and time of each patient's progress throughout the ED, allowing us to calculate the actual waiting time for treatment and observe the level of activity in the ED at specific time points. Additionally, patient data such as triage category at the time of arrival are also recorded.

We restrict our analysis to patients who are classified as triage 3, 4, or 5 and assigned to the ED waiting room. These patients have a lower priority for medical treatment than more urgent cases (triage categories 1 and 2) due to the severity of their condition, which is determined by their presenting symptoms and pre-existing medical conditions. In our sample, patients allocated to the waiting room have largely homogeneous waiting times to treatment across triage categories, but their waiting times are longer than those who wait in the acute bed area.

The objective of this research is to generate hourly waiting time forecasts. To achieve this, we first calculate the waiting time to treatment for each patient as the difference between the time when they saw a doctor or a nurse and the time of triage. These individual waiting times are then averaged for each hourly time window in which patients arrived at the ED.

TABLE 1 A list of competing models.

Model	Description
Rolling average	Rolling average
UC	Unobserved component with a homoscedastic error
TD	Time dummy including days of a week with a homoscedastic error
TD-MA	TD with a moving average error
TD-SV	TD with a stochastic volatility error
TD-MASV	TD with a moving average and stochastic volatility error
TD-t	TD model with a *t*-distributed error
TD-tMA	TD-MA with a *t*-distributed error
TD-tSV	TD-SV with a *t*-distributed error
TD-tMASV	TD-MASV with a *t*-distributed error
UC-MA	UC with a moving average error
UC-SV	UC with a stochastic volatility error
UC-MASV	UC with a moving average and stochastic volatility error
UC-t	UC with a *t*-distributed error
UC-tMA	UC-MA with a *t*-distributed error
UC-tSV	UC-SV with a *t*-distributed error
UC-tMASV	UC-MASV with s *t*-distributed error
TVAR	Time-varying autoregressive AR(1) with a homoscedastic error
TVAR-MA	TVAR with a moving average error
TVAR-SV	TVAR with a stochastic volatility error
TVAR-MASV	TVAR with a moving average and stochastic volatility error
TVAR-t	TVAR with a *t*-distributed error
TVAR-tMA	TVAR-MA with a *t*-distributed error
TVAR-tSV	TVAR-SV with a *t*-distributed error
TVAR-tMASV	TVAR-MASV with a *t*-distributed error
TD-zero	TD considers zeros as an actual waiting time
TD-MA-zero	TD-MA considers zeros as an actual waiting time
TD-SV-zero	TD-SV considers zeros as an actual waiting time
TD-MASV-zero	TD-MASV considers zeros as an actual waiting time
TD-t-zero	TD-t considers zeros as an actual waiting time
TD-tMA-zero	TDt-MA considers zeros as actual waiting time
TD-tSV-zero	TD-tSV considers zeros as an actual waiting time
TD-tSV-zero	TD-tSV considers zeros as an actual waiting time
UC-zero	UC considers zeros as an actual waiting time
UC-MA-zero	UC-MA considers zeros as an actual waiting time
UC-SV-zero	UC-SV considers zeros as an actual waiting time
UC-MASV-zero	UC-MASV considers zeros as an actual waiting time
UC-t-zero	UC-t considers zeros as an actual waiting time
UC-tMA-zero	UC-tMA considers zeros as actual waiting time
UC-tSV-zero	UC-tSV considers zeros as an actual waiting time
UC-tMASV-zero	UC-tMASV considers zeros as an actual waiting time

For instance, if three patients arrived between 2 PM and 3 PM on a particular day, we use their actual waiting time to treatment to compute the mean waiting time for that interval. Our data comprise hourly waiting time windows for low acuity patients at PAH's ED from 1 January 2020 to 15 December 2020, which correspond to 8400 observations.

4 | APPLICATION TO PAH ED WAITING TIME FORECASTING

In this section, we present the full sample estimation results of some selected models listed in Table 1. The objective of this exercise is to examine whether there is any evidence suggesting the presence of time-varying variance and serial correlation in the error process. Second, we investigate the forecast performance of the models listed in Table 1 to understand the role of flexible structures in the error process in forecasting ED waiting time. Finally, we explore whether treating *zero-recorded* ED waiting time as an unobserved component can improve the forecast performance of these models.

4.1 | Priors

In order to complete the model specifications for the models listed in Table 1, we provide priors for the model parameters. To enable easy comparison, we impose the same priors for the common parameters across the models. For the mean process, we choose an independent Gaussian prior for each $\mu_0 = \{\tau_0, \gamma_0, \gamma_1, .., \gamma_m, \beta_{00}, \beta_{10}, \ \dots \ , \beta_{p0}\}$, that is, $\mu_0 \sim N(0, V_{\mu_0})$ where $V_{\mu_0} = 100$. Since the prior variance is relatively large, the prior of μ_0 is considered relatively non-informative. In addition, we assume independent inverse gamma priors for variances σ_y^2 and $\sigma_\Theta^2 = \left\{\sigma_\tau^2, \sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \ \dots \ .\sigma_{\beta_m}^2\right\}$:

$$\sigma_y^2 \sim \mathcal{IG}(\alpha_y, \eta_y), \qquad \sigma_\Theta^2 \sim \mathcal{IG}(\alpha_\Theta, \eta_\Theta).$$

For the prior hyperparameters of σ_y^2 and σ_Θ^2, we select a relatively small value for the shape $\alpha_y = \alpha_\Theta = 3$.[33-35] Choosing a small shape parameters serves two purposes. First, it ensures the existence of the mean and variance of σ_y^2 and σ_Θ^2.[‡] Second, it ensures that a precision-based algorithm is used to sample all states (time-varying) parameters simultaneously, rather than one-at-a-time. Jointly sampling all states, such as $\tau = (\tau_1, \ \dots \ , \tau_T)$, in one step leads to efficiency improvements compared to one-at-a-time Gibbs sampling (see Appendix B for further discussion).[33,34] The scale hyperparameter η_y is set to be $\alpha_y - 1$, which implies $\mathbb{E}(\sigma_y^2) = 1$. This choice is comparable to the models with SV, where the prior mean for the unconditional mean of log-volatility μ_h is equal to zero ($\exp(\mu_h) = 1$). For the scale hyperparameter of each σ_Θ, we set $\eta_\Theta = 0.1^2(\alpha_\Theta - 1)$, which implies $\mathbb{E}(\sigma_\Theta^2) = 0.1^2$. The chosen prior mean reflects the desired smoothness of the corresponding state transition associated with the conditional mean process.

For the SV component in the error process, we use a normal prior for the unconditional mean of the log-volatility μ_h, given by $\mu_h \sim N(0, V_{\mu_h})$, where $V_{\mu_h} = 100$, and a truncated normal prior $\phi_h \sim \mathcal{N}(\phi_{h_0}, V_{\phi_h})\mathbf{1}(|\phi_h| < 1)$ where $\phi_{h_0} = 0.95$ and $V_{\phi_h} = 100$. As for σ_h^2, the prior is an inverse gamma, $\sigma_h^2 \sim \mathcal{IG}(\alpha_h, \eta_h)$ with $\alpha_h = 3$ and the scale hyperparameter $\eta_h = 0.1^2(\alpha_h - 1)$, which implies $\mathbb{E}(\sigma_h^2) = 0.1^2$ and reflects the smoothness of the corresponding transition associated with the log-volatility.

The degrees of freedom parameter (ν) in the SS model with t-distributed error is treated as an unknown parameter and has a uniform prior, that is, $\nu \sim \mathcal{U}(2{,}100)$. We restrict ν to be greater than 2 to ensure the first and second moments of the t distribution exist. For the MA component of the error process, we impose a truncated Gaussian prior for ψ to guarantee the invertibility of the MA process. Specifically, we set $\psi \sim \mathcal{N}(0, V_\psi)\mathbf{1}(|\psi| < 1)$ with $V_\psi = 100$. Additionally, we assume an unobserved waiting time y_t^u follows a normal prior, that is, $y_t^u \sim \mathcal{N}(0, V_{y^u})$ with a large variance $V_{y^u} = 100$.

4.2 | Full sample estimation results

Table 1 presents the full estimation results (01/01/2020 00:00:00 to 15/12/2020 23:00:00) of some selected models, which include TD-tMA, UC-tMA, TVAR-tMA, TD-tMA, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV. The objective is to investigate the presence of time-varying variance and serial correlation in the error process. All estimates presented in this section are based on 25 000 draws from MCMC samplers, following a burn-in period of 5000.

A key parameter of interest is the MA coefficient ψ. If the posterior density of ψ tends to concentrate near zero, it suggests that the inclusion of the MA component may not be necessary. The posterior distributions of ψ depicted in Figure 1 demonstrate that the majority of the distribution mass for the MA coefficient, across the model specifications, is concentrated in regions far from zero. For example, the posterior mean of ψ is approximately 0.1 under the UC-tMASV model. This finding implies that even after accounting for time-variation in variances, there is a notable autocorrelation in the observed waiting time, lending support to the presence of an MA component.

Another interesting point is that under the TD-tMA, TD-tMASV, UC-tMA, and UC-tMASV models, the MA coefficient ψ is estimated to be positive, while it is estimated to be negative under the TVAR-tMA and TVAR-tMASV models. The difference might not be surprising. In the TD and UC model variants, the autocorrelation of the waiting time is modeled through either time dummy variables or a random walk in the conditional mean process, which might not be sufficient to capture all the observed positive autocorrelation. In the TVAR-tMA and TVAR-tMASV models, past waiting time information is directly captured in the conditional mean. The results suggest that the autoregressive components, based on the waiting time with one lag, seem to "over-capture" the observed positive autocorrelation, resulting in a negative autocorrelation in the residuals.

As an example, Figure 2 reports the estimated posterior means and the 90% credible intervals for the standard error $\exp(h_t/2)$ during the period from 10/12/2020 00:00:00 to 15/12/2020 23:00:00. This particular subset of data is presented to provide a clearer visualization of the dynamic movements in the standard error. The full 1-year hourly dataset is too dense to observe the pattern effectively. Across the model specifications, we observe substantial time-variation in the standard error estimates for the waiting time, emphasizing the importance of accounting for time-variation in the error process.

To evaluate the convergence and efficiency of the MCMC algorithms in estimating the models, we compute the convergence diagnostic proposed by Geweke.[37] For visual inspections, we provide box plots of inefficiency factors and trace plots for specific parameters.[38] Further details can be found in Appendix C.

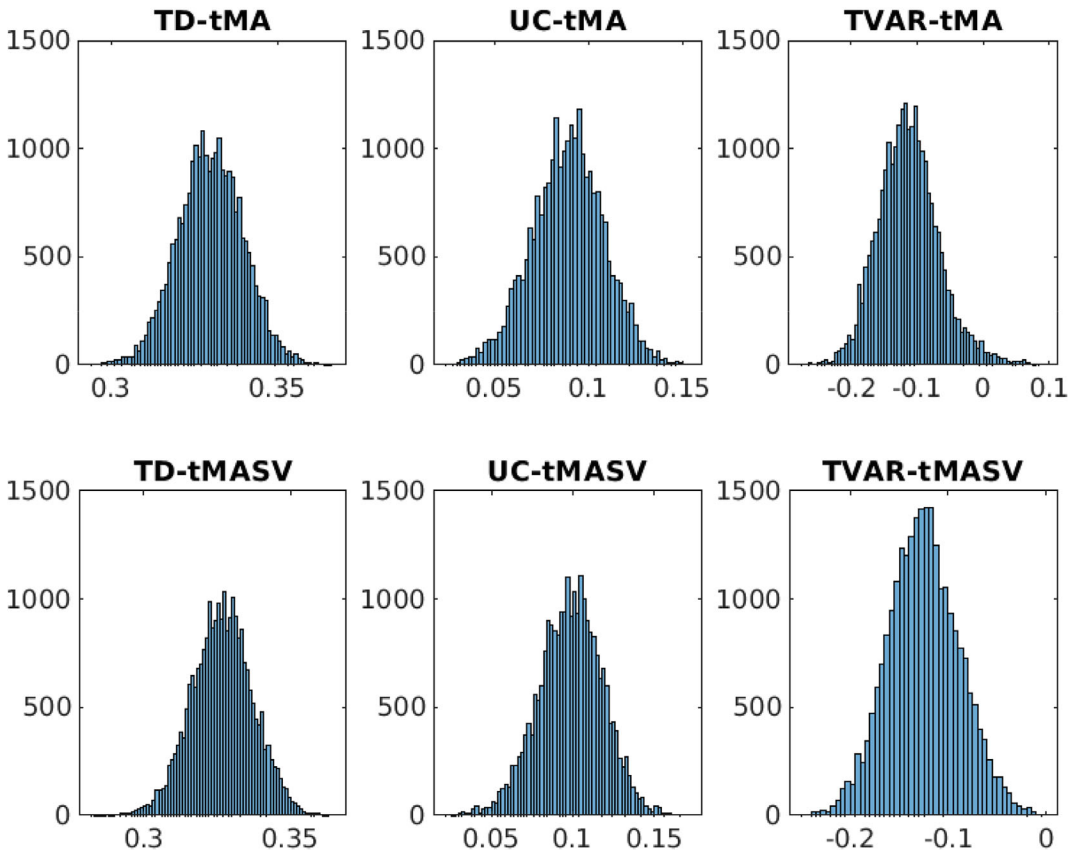

FIGURE 1 Estimated posterior distributions of ψ under the TD-tMA, UC-tMA, TVAR-tMA, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

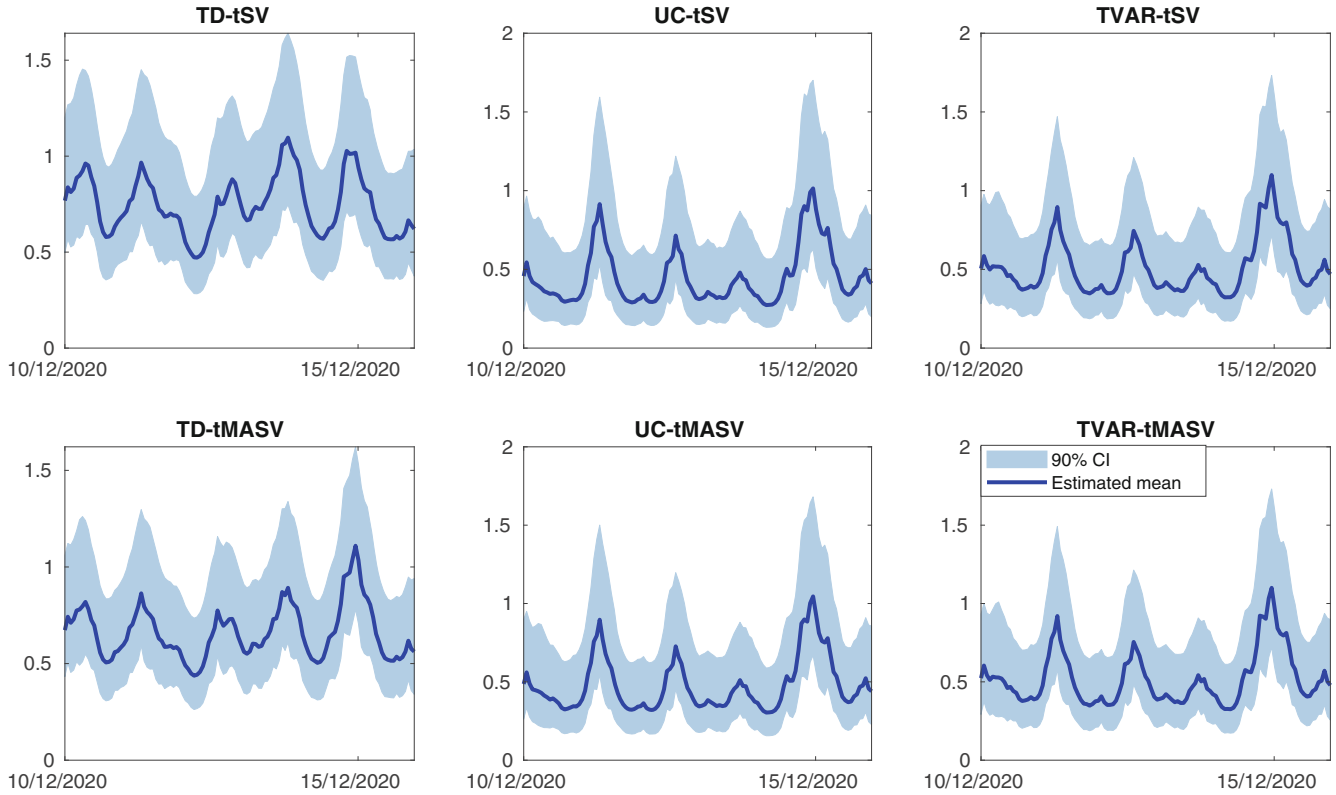

FIGURE 2 Estimated posterior means and the 90% credible intervals for the standard error $\exp(h_t/2)$ during the period from 10/12/2020 00:00:00 to 15/12/2020 23:00:00 under the TD-tSV, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

4.3 | Forecast setup and forecast measures

In order to evaluate the performance of the models listed in Table 1, we conduct a recursive out-of-sample forecasting exercise. The primary objective of this exercise is to investigate whether incorporating flexible error structures enhances the accuracy of waiting time forecasts. Additionally, we explore whether treating *zero-recorded* ED waiting times as unobserved values improves the forecast performance of the models.

The recursive exercise involves dividing the data into three sub-samples. The first part is the *initialization period*, which contains the first p observations that are used to initialize TVAR(p) models. This ensures that UC, TD, and TVAR model variants have the same initial observations. For this exercise, we choose the lags of TVAR models (p) equal to 1. The second part is the *training period*, denoted as $\mathbf{y}_{1:t}$, which represents an expanding window of observations. We first start with the training period up until t (10/12/2020 23:00:00). We then employ MCMC samplers, as described in Appendix B, to obtain posterior draws of model parameters for each model based on $\mathbf{y}_{1:t}$. The point forecast for the waiting time observations, $\mathbb{E}(y_{t+h}|\mathbf{y}_{1:t})$, is computed as the predictive mean for h-period-ahead forecasts, where $h = 1, 2, \ldots, 8$ (ie, 11/12/2020 00:00:00 to 11/12/2020 07:00:00). Additionally, we compute $p(y_{t+h}|\mathbf{y}_{1:t})$ as the density forecast. This procedure is repeated by iteratively incorporating additional data from 11/12/00:00:00 to 15/12/2020 22:00:00 into the *training period*. The third part of the data is the *evaluation period*, consisting of the remaining observations, which are used to assess the forecast of each model. In total, we generate 168 recursive predictions for $h = 1$ and 161 recursive predictions for $h = 8$.

To compare the performance of the models, we use both point forecast measures, such as the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE), and the density forecast measure—the log predictive likelihood. Utilizing the density forecast measure offers an insight into the entire predictive distribution of ED waiting times, enabling a more comprehensive evaluation of the models' forecasting performance.

To ensure a fair comparison, we exclude *zero-recorded* observations when computing the forecast measures. This exclusion is necessary because we do not observe the actual waiting time within the time windows recorded as zero. The point forecast measures, namely RMSFE and MAFE, are defined as follows:

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=t_0}^{T-h-n_0}(y_{t+h}^0 - \mathbb{E}(y_{t+h}|\mathbf{y}_{1:t}))^2}{T-h-t_0-n_0+1}},$$
$$\text{MAFE} = \frac{\sum_{t=t_0}^{T-h-n_0}|y_{t+h}^0 - \mathbb{E}(y_{t+h}|\mathbf{y}_{1:t})|}{T-h-t_0-n_0+1};$$

where y_{t+h}^0 is the observed value of y_{t+h}, t_0 denotes the start of the evaluation period, and n_0 indicates the number of zero-recorded observations in the evaluation dataset. A smaller forecast error corresponds to a smaller RMSFE and MAFE.

To evaluate the density forecast $p(y_{t+h}|\mathbf{y}_{1:t})$ we use the predictive likelihood $p(y_{t+h} = y_{t+h}^0|\mathbf{y}_{1:t})$, which represents the predictive density of y_{t+h} evaluated at the observed value y_{t+h}^0. A high value of the predictive likelihood indicates that the observed value y_{t+h}^0 is likely to be within the density forecast, whereas a low value suggests otherwise. We evaluate the density forecasts by computing the average log predictive likelihood (ALPL).

$$\text{ALPL} = \frac{1}{T-h-t_0-n_0+1}\sum_{t=t_0}^{T-h-n_0}\log p_{t+h}(y_{t+h} = y_{t+h}^0|\mathbf{y}_{1:t}).$$

To facilitate the comparison of performance among the models, we standardize the forecast measures relative to the benchmark models. If the ratio between the model's RMSFE and the benchmark's RMSFE is smaller than 1, it suggests that the model outperforms the benchmark, and vice versa. Similarly, for ALPL, we standardize the values by subtracting the ALPL of each model from the ALPL of the benchmark model. A positive relative ALPL indicates that other models perform better than the benchmark, while a negative value suggests the opposite.

4.4 | Forecast results

Tables 2–4 provide a summary of the relative RMSFE, MAFE, and ALPL forecast measures for the competing models compared to the benchmark models at different forecast horizons: $h = 1$ (one-hour-ahead forecast), $h = 2$ (two-hour-ahead forecast), … , and $h = 8$ (eight-hour-ahead forecast). To determine whether the differences in forecast accuracy between the benchmark model and the alternative models are statistically significant, we conduct a one-sided test of equal predictive accuracy as described by Diebold and Mariano.[39] The differences in accuracy that are statistically different from zero are denoted with one, two, or three asterisks in Tables 2–4, corresponding to significance levels of 10%, 5%, and 1%, respectively.

A notable finding from the forecasting results is that treating *zero-recorded* observations as unobserved values significantly improves the forecast performance.[§] Tables 2–4 reveal that the models treating *zero-recorded* observations as unobserved values generally exhibit lower relative RMSFE and MAFE values, and higher relative ALPL values compared to their counterparts where *zero-recorded* observations are treated as *actual zero* values. For example, the UC-tMASV model achieves the smallest relative RMSFE of 0.914 for the one-hour-ahead forecast, while the UC-tMASV-zero model yields a relative RMSFE of 1.039. The key takeaway is that appropriately handling *zero-recorded* observations is necessary before modeling or forecasting ED waiting time.

We observe that no single model dominates in forecasting ED waiting time across the forecast horizons. However, models employing UC and TVAR specifications in the conditional mean process tend to yield better forecasts compared to the rolling average and TD models, particularly for short-term forecasts up to 4 h ahead. The RMSFE results presented in Table 2 demonstrate that the UC-tMASV model outperforms the benchmark rolling average model by nearly 10% in terms of the RMSFE criterion. Moreover, the 68% credible interval of the predicted waiting times obtained from the UC-tMASV model encompasses the majority of the observed values (refer to Figure A1 in Appendix A).

The ALPL density forecast measure indicates that the UC-SV and TVAR-MASV models are superior in forecasting ED waiting time for one-hour-ahead-forecast and three-hour-ahead-forecast, respectively. These results align with the findings from the estimation using the full sample in Section 4.2. Notably, incorporating both time-variation and serial correlation in the model specifications leads to improved in-sample fitness and out-of-sample forecast performance compared to the model specifications with constant variances.

While the findings indicate that the TD models do not perform well in short-term forecasting, they exhibit favorable performance in long-term forecasting ($h > 4$). The TD and TD-SV models achieve a reduction in RMSFE and MAFE of

TABLE 2 Relative RMSFE.

Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$
Rolling average	0.744	0.788	0.827	0.854	0.890	0.922	0.978	1.009
UC	0.923	0.929	0.923	0.937	0.947	0.933*	0.921***	0.919**
TD	1.055	1.007	0.966	0.924	0.892	0.864*	**0.813****	**0.788*****
TD-MA	0.933	1.001	0.961	0.928	**0.889**	**0.86****	**0.813****	**0.788*****
TD-SV	1.058	1.001	0.961	0.928	0.895	0.864	0.817**	0.792**
TD-MASV	0.922	1.007	0.967	0.926	0.895	0.867	0.816**	0.79**
TD-t	1.059	1.006	0.964	0.929	**0.889**	0.864*	0.813**	0.792***
TD-tMA	0.934	1.009	0.964	0.930	0.898	0.868*	0.82**	0.793*
TD-tSV	1.057	1.007	0.967	0.926	0.894	0.873	0.816**	0.793**
TD-tMASV	0.925	1.011	0.961	0.926	0.892	0.869**	0.822**	0.792**
UC-MA	0.92*	0.946	0.929	**0.923**	0.918	0.895***	0.872***	0.863***
UC-SV	0.918***	0.93**	0.924**	0.943**	0.959	0.943***	0.939***	0.94*
UC-MASV	0.919***	**0.928***	0.92*	0.942	0.955	0.94**	0.931***	0.923**
UC-t	0.931**	0.937	0.929*	0.958	0.971	0.948**	0.948**	0.942**
UC-tMA	0.93**	0.939	0.925	0.947	0.965	0.942**	0.93**	0.925
UC-tSV	0.917**	0.935***	0.917**	0.941**	0.961	0.938**	0.929***	0.931*
UC-tMASV	**0.914****	**0.928****	**0.916***	0.943	0.965	0.938**	0.939***	0.934***
TVAR	0.948**	0.943*	0.926	0.951	0.970	0.949	0.952**	0.968
TVAR-MA	0.94*	0.936	0.923	0.943	0.958	0.937**	0.939	0.941**
TVAR-SV	0.939**	0.948**	0.925	0.955*	0.985*	0.967*	0.970	0.980
TVAR-MASV	0.944**	0.942**	0.925**	0.949*	0.978	0.956**	0.966	0.976
TVAR-t	0.945**	0.949	0.932*	0.958	0.986	0.956	0.973*	0.981
TVAR-tMA	0.947**	0.944	0.924	0.954	0.972	0.951	0.952*	0.968
TVAR-tSV	0.936***	0.943	0.924*	0.957*	0.985	0.960	0.964*	0.985
TVAR-tMASV	0.939**	0.939***	0.93*	0.960	0.981*	0.962	0.97*	0.980
TD-zero	2.023	1.882	1.836	1.700	1.676	1.613	1.522	1.433
TD-MA-zero	2.235	1.903	1.800	1.744	1.666	1.624	1.489	1.454
TD-SV-zero	1.055	1.028	0.982	0.947	0.912	0.881	0.833**	0.811*
TD-MASV-zero	1.488	1.037	0.961	0.930	0.919	0.891	0.837**	0.799**
TD-t-zero	1.061	1.001	0.977	0.934	0.898	0.877	0.827**	0.798**
TD-tMA-zero	1.059	1.017	0.971	0.935	0.904*	0.872	0.821**	0.795**
TD-tSV-zero	1.070	1.020	0.982	0.929	0.920	0.871	0.835***	0.803**
TD-tMASV-zero	1.058	1.024	0.977	0.948	0.908	0.877*	0.836***	0.805**
UC-zero	2.900	2.738	2.598	2.520	2.450	2.406	2.252	2.142
UC-MA-zero	2.891	2.578	2.471	2.407	2.270	2.214	2.081	2.050
UC-SV-zero	5.189	5.031	4.692	4.751	4.594	4.522	4.474	4.348
UC-MASV-zero	4.827	4.669	4.386	4.455	4.272	4.223	4.181	4.106
UC-t-zero	0.932	0.943	0.944	0.952	0.969	0.947*	0.938*	0.925**
UC-tMA-zero	0.937	0.938	0.927	0.959	0.967	0.944*	0.936	0.926**
UC-tSV-zero	1.071	1.018	0.969	0.995	1.037	0.965**	0.971	0.993
UC-tMASV-zero	1.039	1.004*	0.950	0.976	1.029	0.997	0.942***	0.976

Note: The results presented for the benchmark (rolling average) model represent the actual RMSFE, while the results for the other models are presented relative to the benchmark. The best-performing model is indicated in bold. The subscript symbols ***, **, and * denote the significance levels of 1%, 5%, and 10%, respectively, indicating the presence of a significant difference in predictive accuracy between the alternative models and the benchmark. These significance levels are determined using the asymptotic test proposed by Diebold and Mariano.[39]

TABLE 3 Relative MAFE.

Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$
Rolling average	0.571	0.623	0.662	0.679	0.726	0.767	0.811	0.828
UC	0.924	0.935	0.936	0.949	0.967	0.934**	0.919***	0.914***
TD	1.126	1.038	0.980	0.949	0.895	0.854*	**0.803****	**0.781*****
TD-MA	0.948	1.032	0.981	0.951	**0.892**	**0.846***	**0.803*****	**0.781*****
TD-SV	1.121	1.036	0.981	0.947	0.902	0.848	0.805**	0.785**
TD-MASV	0.923	1.036	0.987	0.951	0.900	0.852	0.806**	0.787**
TD-t	1.125	1.037	0.981	0.955	**0.892**	0.85*	0.805***	0.785***
TD-tMA	0.932	1.042	0.978	0.951	0.906	0.856*	0.811**	0.79*
TD-tSV	1.123	1.038	0.984	0.949	0.897	0.860	0.803*	0.787**
TD-tMASV	0.923	1.041	0.980	0.946	0.895	0.855*	0.811**	0.788**
UC-MA	0.927	0.979	0.955	0.955	0.931	0.884***	0.861***	0.848**
UC-SV	0.9**	0.911***	0.917*	0.941**	0.970	0.947***	0.94***	0.945**
UC-MASV	**0.899****	0.919	0.924	0.954	0.971	0.949**	0.930***	0.927***
UC-t	0.907***	0.924	0.928	0.961	0.985	0.961**	0.943**	0.955*
UC-tMA	0.907**	0.940	0.944	0.963	0.988	0.948*	0.932***	0.930
UC-tSV	0.902**	0.921***	0.92**	0.949***	0.980	0.946***	0.927***	0.934**
UC-tMASV	**0.899****	**0.908****	**0.910****	0.939	0.978	0.948**	0.939***	0.941***
TVAR	0.932*	0.93*	0.927	0.939	0.963	0.932	0.915	0.917***
TVAR-MA	0.928**	0.935	0.928	0.942	0.960	0.915*	0.905*	0.899*
TVAR-SV	0.929**	0.919**	0.912*	0.928	0.956	0.939**	0.916***	0.917*
TVAR-MASV	0.93**	0.91**	0.916	**0.921****	0.962	0.929**	0.92***	0.922*
TVAR-t	0.926**	0.924	0.922	0.943*	0.971	0.931***	0.926*	0.930
TVAR-tMA	0.931**	0.931	0.925	0.939	0.962	0.922	0.911***	0.918**
TVAR-tSV	0.92**	0.915*	0.91*	0.929	0.963	0.934	0.915***	0.924**
TVAR-tMASV	0.929**	0.911*	0.922	0.933	0.965	0.934**	0.922***	0.921**
TD-zero	2.136	1.944	1.848	1.746	1.700	1.550	1.517	1.425
TD-MA-zero	2.246	1.966	1.869	1.788	1.675	1.602	1.461	1.440
TD-SV-zero	1.116	1.051	0.992	0.967	0.908	0.864*	0.819**	0.802**
TD-MASV-zero	1.295	1.040	0.992	0.963	0.911	0.866*	0.818**	0.792***
TD-t-zero	1.123	1.040	1.000	0.958	0.905	0.859	0.814*	0.791**
TD-tMA-zero	1.112	1.051	0.992	0.964	0.903*	0.859*	0.81**	0.791**
TD-tSV-zero	1.136	1.050	0.992	0.964	0.915	0.863	0.82**	0.799**
TD-tMASV-zero	1.112	1.058	0.994	0.970	0.917	0.864*	0.821***	0.799**
UC-zero	3.324	3.036	2.834	2.759	2.603	2.497	2.372	2.293
UC-MA-zero	3.143	2.887	2.699	2.649	2.395	2.339	2.227	2.183
UC-SV-zero	3.076	3.170	3.232	3.593	3.503	3.481	3.726	3.709
UC-MASV-zero	2.525	2.571	2.413	2.519	2.524	2.414	2.477	2.416
UC-t-zero	2.490	2.547	2.344	2.468	2.432	2.310	2.431*	2.419**
UC-tMA-zero	0.930	0.939	0.948	0.977	0.995	0.954	0.937	0.933**
UC-tSV-zero	0.937*	0.943	0.944	0.968	1.015	0.973***	0.963**	0.971
UC-tMASV-zero	0.923*	0.926**	0.944	0.97*	1.005	0.978*	0.963***	0.984**

Note: The results presented for the benchmark (rolling average) model represent the actual MAFE, while the results for the other models are presented relative to the benchmark. The best-performing model is indicated in bold. The subscript symbols ***, **, and * denote the significance levels of 1%, 5%, and 10%, respectively, indicating the presence of a significant difference in predictive accuracy between the alternative models and the benchmark. These significance levels are determined using the asymptotic test proposed by Diebold and Mariano.[39]

TABLE 4 Relative ALPL.

Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$
UC	−1.042	−1.107	−1.151	−1.196	−1.252	−1.272	−1.317	−1.347
TD	−0.137***	−0.081***	−0.042***	0.011***	0.058***	**0.077*** **	**0.123*** **	**0.155*** **
TD-MA	−0.013***	−0.081***	−0.042***	0.01***	**0.06*** **	**0.077*** **	**0.123*** **	**0.155*** **
TD-SV	−0.13***	−0.086***	−0.053***	−0.007***	0.038***	0.053***	0.096***	0.131***
TD-MASV	0.021***	−0.08***	−0.053***	−0.009***	0.037***	0.047***	0.094***	0.13***
TD-t	−0.14***	−0.083***	−0.045***	0.008***	0.057***	0.074***	0.119***	0.152***
TD-tMA	−0.009***	−0.092***	−0.053***	0***	0.048***	0.064***	0.107***	0.14***
TD-tSV	−0.134***	−0.09***	−0.055***	−0.009***	0.037***	0.052***	0.097***	0.131***
TD-tMASV	0.017***	−0.082***	−0.053***	−0.01***	0.034***	0.048***	0.093***	0.126***
UC-MA	0.001	−0.021	−0.006	**0.019**	0.036***	0.044***	0.056***	0.069***
UC-SV	**0.064*** **	**0.031*** **	0.01*	−0.004	−0.021	−0.020	−0.036	−0.032
UC-MASV	0.057***	0.026**	0.005**	−0.006	−0.021	−0.020	−0.027	−0.015
UC-t	0.015***	0.002***	−0.005**	−0.025	−0.032	−0.026	−0.040	−0.039
UC-tMA	0.012***	−0.006	−0.010	−0.020	−0.026	−0.019	−0.024	−0.017
UC-tSV	0.056***	0.019***	0.008**	−0.006	−0.017	−0.018	−0.026	−0.024
UC-tMASV	0.06***	0.026***	0.012***	−0.007	−0.025	−0.023	−0.034	−0.028
TVAR	−0.007	0.003	0.005	−0.003	−0.004	−0.002	−0.003*	−0.002***
TVAR-MA	−0.002	0.000	0.004	0.011	0.010	0.016	0.013*	0.016***
TVAR-SV	0.048***	0.027***	0.011***	0.001***	−0.013***	−0.025***	−0.031***	−0.027***
TVAR-MASV	0.043***	**0.031*** **	**0.014*** **	0.006***	−0.012***	−0.015***	−0.022***	−0.026***
TVAR-t	0.01***	0.002**	0.001	−0.012	−0.019	−0.018***	−0.026***	−0.024***
TVAR-tMA	0.009**	0.000	0.001	−0.008	−0.011	−0.013***	−0.013**	−0.016***
TVAR-tSV	0.052***	0.026***	0.012***	−0.001***	−0.018***	−0.023***	−0.031***	−0.027***
TVAR-tMASV	0.043***	0.026***	0.007***	−0.005***	−0.019***	−0.022***	−0.033***	−0.022***
TD-zero	−1.768	−1.703	−1.658	−1.612	−1.556	−1.536	−1.491	−1.460
TD-MA-zero	−1.772	−1.704	−1.659	−1.614	−1.557	−1.537	−1.492	−1.461
TD-SV-zero	−0.316	−0.266	−0.245	−0.199	−0.162	−0.145	−0.099	−0.062
TD-MASV-zero	−0.297	−0.263	−0.252	−0.200	−0.164	−0.149	−0.103	−0.069
TD-t-zero	−0.180	−0.127	−0.087	−0.038*	0.015**	0.03***	0.073***	0.107***
TD-tMA-zero	−0.160	−0.124	−0.087**	−0.036***	0.013***	0.03***	0.077***	0.107***
TD-tSV-zero	−0.234	−0.183	−0.154	−0.104	−0.062	−0.041	−0.002	0.038*
TD-tMASV-zero	−0.205	−0.182	−0.154	−0.105	−0.062	−0.044	0.005	0.037
UC-zero	−1.791	−1.726	−1.683	−1.637	−1.582	−1.562	−1.518	−1.487
UC-MA-zero	−1.792	−1.724	−1.680	−1.635	−1.579	−1.560	−1.515	−1.484
UC-SV-zero	−2.023	−1.797	−1.791	−1.834	−1.859	−1.893	−1.930	−1.951
UC-MASV-zero	−1.743	−1.662	−1.680	−1.718	−1.744	−1.789	−1.813	−1.841
UC-t-zero	−0.019	−0.024	−0.035	−0.033	−0.037	−0.038	−0.036	−0.023
UC-tMA-zero	−0.010	−0.019	−0.026	−0.034	−0.029	−0.035	−0.033	−0.025
UC-tSV-zero	−0.043	−0.100	−0.094	−0.119	−0.131	−0.164	−0.180	−0.177
UC-tMASV-zero	−0.029	−0.087	−0.089	−0.115	−0.133	−0.139	−0.136	−0.161

Note: The results presented for the benchmark (UC) model represent the actual ALPL, while the results for the other models are presented relative to the benchmark. The best-performing model is indicated in bold. The subscript symbols ***, **, and * denote the significance levels of 1%, 5%, and 10%, respectively, indicating the presence of a significant difference in predictive accuracy between the alternative models and the benchmark. These significance levels are determined using the asymptotic test proposed by Diebold and Mariano.[39]

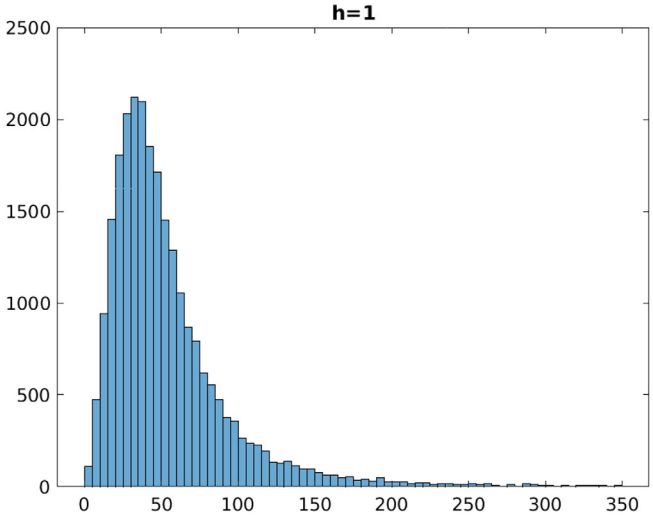

FIGURE 3 Predictive posterior distribution of ED waiting time at $h = 1$ using the UC-tMASV model.

approximately 22% compared to the benchmark rolling average model for an eight-hour-ahead-forecast. Furthermore, the relative density forecast measure ALPL yields significantly positive values, indicating that these models provide better predictions of waiting time compared to the benchmark UC model. Figure A1 in Appendix A displays the out-of-sample forecasts of the UC-tMASV and TD models for $h = 1$ and $h = 8$. In comparison to the UC-tMASV model, the forecasts based on the TD model exhibit less volatility. This suggests that incorporating features that capture dynamic movements of waiting time and flexible error structures improve short-term forecasts. However, the benefits of incorporating the features are less pronounced in improving long-term forecasts. Additionally, it is worth noting that the TD models with flexible error specifications do not appear to outperform the TD model with a constant variance for long-term forecasts.

Hospitals often face challenges in meeting fixed waiting time predictions for patients in the ED, rendering point forecasts of ED waiting time less informative. This observation is consistent with findings in the waiting time literature, where studies have shown that patient dissatisfaction tends to increase when waiting times exceed previous indications.[11,12] As a result, conveying information about the uncertainties associated with ED waiting time forecasts can provide valuable practical benefits. ED waiting times are inherently uncertain and asymmetric due to the complex nature of ED clinical pathways and the unpredictability of arrivals of urgent and critically-ill patients, necessitating the sudden redirection of substantial ED resources.

Using Bayesian estimation for the models enables us to examine the predictive distribution of ED waiting time, thereby improving our understanding of the associated uncertainties in the forecasts. Figure 3 illustrates the predictive posterior distribution of the predicted waiting time at $h = 1$ using the UC-tMASV model. By leveraging the predictive distribution, we can readily compute the x% credible interval for the one-period-ahead waiting time prediction. For example, the 90% credible interval indicates a 90% probability that the waiting time for the next time window will range from 14 to 132 min. In contrast, the 50% credible interval suggests that the waiting time for the next time window is expected to fluctuate between 28 and 67 min. Such information provides practitioners with a degree of confidence in the accuracy of waiting time forecasts.

5 | CONCLUSION

In this article, we propose a set of models within a state space framework for forecasting ED waiting times. The state space framework provides a parsimonious specification that captures the dynamic variations of ED waiting time and provides flexibility in handling *zero-recorded* waiting times. Our findings indicate that treating *zero-recorded* waiting times as unobserved values improves forecast performance. Additionally, incorporating flexible error structures in the model specifications enhances forecast accuracy, especially for short-term forecasts. Based on our empirical findings, the UC-tMASV model is superior for short term forecasts while the TD model provides better forecasts for longer term horizons. The

UC-tMASV model outperforms the rolling average model by 10% for forecasts up to four hours ahead. On the other hand, both the TD and TD-SV models reduce the RMSFE of the benchmark model by nearly 22% for eight-hour-ahead forecasts.

As a future research direction, it would be valuable to incorporate other explanatory variables, if available, to further understand their impacts on waiting time and investigate whether the inclusion of exogenous variables can enhance forecast performance. Another avenue is to combine forecasts from several different models using Bayesian model averaging. A challenge is to develop an efficient and accurate algorithm to compute marginal likelihoods for the SS models with flexible error structures.

ACKNOWLEDGEMENTS

We would like to thank the editor Dr. Robert Platt, an anonymous associated editor and two anonymous referees for their insightful comments and constructive suggestions, which greatly improved the article. This research was supported by a Grant from the Emergency Medicine Foundation (Australasia) Queensland Program, EMJS-343R33-2020-STAIB. Dr. Anton Pak acknowledges the support from the University of Queensland Health and Behavioural Science Faculty Early Career Academic Research Accelerator Award.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Princess Alexandra Hospital, Brisbane, Queensland, Australia. Restrictions apply to the availability of these data, which were used under license for this study. Data are available with Dr. Andrew Staib given the permission of Princess Alexandra Hospital.

ENDNOTES

*From now onwards, aggregated ED waiting time and ED waiting time are used interchangeably.

†Another way to accommodate the time-variation in variance is the generalized autoregressive conditional heteroskedasticity model.[28,29]

‡The inverse gamma distribution has the probability density function as follows:[36]

$$f(z;\alpha,\eta)=\frac{\eta^{\alpha}}{\Gamma(\alpha)}\frac{1}{y^{\alpha+1}}\exp\left(-\frac{\eta}{z}\right),\quad z>0.$$

The mean of z is $E(z)=\frac{\eta}{\alpha-1}$ for $\alpha>1$, and $var(z)=\frac{\eta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha>2$.

§For the computation purpose, we add a small positive number (ie, ϵ = 1e−3, 1e−5, and 1e−7) to the zero waiting time observations before taking a logarithm for model estimation. We find that the main conclusions are robust across different choices of ϵ.

ORCID

Kelly Trinh https://orcid.org/0000-0002-1814-6847

REFERENCES

1. Asplin B, Magid D, Rhodes K, Solberg L, Lurie N, Camargo C. A conceptual model of emergency department crowding. *Ann Emerg Med*. 2003;42(2):173-180.
2. Soremekun O, Takayesu J, Bohan S. Framework for analyzing wait times and other factors that impact patient satisfaction in the emergency department. *J Emerg Med*. 2011;41(6):686-692.
3. Johnson M, Myers S, Wineholt J, Pollack M, Kusmiesz A. Patients who leave the emergency department without being seen. *J Emerg Nurs*. 2009;35(2):105-108.
4. Yip A, McLeod S, McRae A, Xie B. Influence of publicly available online wait time data on emergency department choice in patients with noncritical complaints. *Can J Emerg Med*. 2012;14(4):237-246.
5. Weiner SG. Advertising emergency department wait times. *West J Emerg Med*. 2013;14(2):77-78.
6. Ang E, Kwasnick S, Bayati M, Plambeck E, Aratow M. Accurate emergency department wait time prediction. *Manuf Serv Oper Manag*. 2016;18(1):141-156.
7. Pak A, Gannon B, Staib A. Predicting waiting time to treatment for emergency department patients. *Int J Med Inform*. 2021;145:104303.
8. Lin D, Patrick J, Labeau F. Estimating the waiting time of multi-priority emergency patients with downstream blocking. *Health Care Manag Sci*. 2014;17(1):88-99.
9. Sun Y, Teow KL, Heng B, Ooi C, Tay SY. Real-time prediction of waiting time in the emergency department, using quantile regression. *Ann Emerg Med*. 2012;60:299-308.
10. Benevento E, Aloini D, Squicciarini N, Dulmin R, Mininno V. Queue-based features for dynamic waiting time prediction in emergency department. *Meas Bus Excell*. 2019;23(4):458-471.

11. Thompson D, Yarnold P. Relating patient satisfaction to waiting time perceptions and expectations: the disconfirmation paradigm. *Acad Emerg Med*. 1995;2(12):1057-1062.
12. Hedges J, Trout A, Magnusson R. Satisfied patients exiting the emergency department (SPEED) study. *Acad Emerg Med*. 2002; 9(1):15-21.
13. Boudreaux E, O'Hea E. Patient satisfaction in the emergency department: a review of the literature and implications for practice. *J Emerg Med*. 2004;26(1):13-26.
14. Harvey A. Trends and cycles in macroeconomic time series. *J Bus Econ Stat*. 1985;3(3):216-227.
15. Roesser R. A discrete state-space model for linear image processing. *IEEE Trans Automat Contr*. 1975;20:1-10.
16. Uhlmann A. The "transition probability" in the state space of a *-algebra. *Rep Math Phys*. 1976;9(2):273-279.
17. Watson M. Univariate detrending methods with stochastic trends. *J Monet Econ*. 1986;18(1):49-75.
18. Grant A, Chan J. Reconciling output gaps: unobserved components model and Hodrick–Prescott filter. *J Econ Dyn Control*. 2017;75:114-121.
19. Newman K, Buckland S, Morgan B, et al. *Modelling Population Dynamics: Model Formulation, Fitting and Assessment Using State-Space Methods*. New York: Springer; 2014.
20. Nobre F, Monteiro A, Telles P, Williamson D. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Stat Med*. 2001;20(20):3051-3069.
21. Fukaya K, Kawamori A, Osada Y, Kitazawa M, Ishiguro M. The forecasting of menstruation based on a state-space modeling of basal body temperature time series. *Stat Med*. 2017;36(21):3361-3379.
22. Kawamori A, Fukaya K, Kitazawa M, Ishiguro M. A self-excited threshold autoregressive state-space model for menstrual cycles: forecasting menstruation and identifying within-cycle stages based on basal body temperature. *Stat Med*. 2019;38(12): 2157-2170.
23. Christensen A, Lundbye-Christensen S, Overvad K, Rasmussen L, Dethlefsen C. Modeling gradually changing seasonal variation in count data using state space models: a cohort study of hospitalization rates of stroke in atrial fibrillation patients in Denmark from 1977 to 2011. *BMC Med Res Methodol*. 2011;12:174.
24. Jeffreys H. *The Theory of Probability*. Oxford: Oxford University Press; 1998.
25. Casella G, Berger R. *Statistical Inference*. Boston, MA: Cengage Learning; 2021.
26. Geweke J. Bayesian treatment of the independent student-t linear model. *J Appl Economet*. 1993;8:S19-S40.
27. Kim S, Shephard N, Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev Econ Stud*. 1998;65(3):361-393.
28. Engle R. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom. *Econometrica*. 1982;50(4):987-1007.
29. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *J Econom*. 1986;31(3):307-327.
30. Montgomery D, Jennings C, Kulahci M. *Introduction to Time Series Analysis and Forecasting*. Hoboken, NJ: John Wiley & Sons; 2015.
31. Hamilton JD. *Time Series Analysis*. Princeton, NJ: Princeton University Press; 2020.
32. Box G, Jenkins G. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day; 1976.
33. Frühwirth-Schnatter S. Data augmentation and dynamic linear models. *J Time Ser Anal*. 1994;15(2):183-202.
34. Chan J, Jeliazkov I. Efficient simulation and integrated likelihood estimation in state space models. *Int J Math Model Numer Optim*. 2009;1(1-2):101-120.
35. Chib S, Jeliazkov I. Inference in semiparametric dynamic models for binary longitudinal data. *J Am Stat Assoc*. 2006;101(474): 685-700.
36. Triantafyllopoulos K. *Bayesian Inference of State Space Models*. Cham: Springer; 2021.
37. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo J, Smith A, Dawid A, Berger JE, eds. *Bayesian Statistics 4*. Oxford: Oxford University Press; 1992:169-193.
38. Chib S. Markov chain Monte Carlo methods: computation and inference. *Handb Econom*. 2001;5:3569-3649.
39. Diebold F, Mariano R. Comparing predictive accuracy. *J Bus Econ Stat*. 2002;20(1):134-144.
40. Harvey A. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press; 1990.
41. Carter C, Kohn R. On Gibbs sampling for state space models. *Biometrika*. 1994;81(3):541-553.
42. McCausland W, Miller S, Pelletier D. Simulation smoothing for state–space models: a computational efficiency analysis. *Comput Stat Data Anal*. 2011;55(1):199-212.
43. Chan J. Moving average stochastic volatility models with application to inflation forecast. *J Econom*. 2013;176(2):162-172.

How to cite this article: Trinh K, Staib A, Pak A. Forecasting emergency department waiting time using a state space representation. *Statistics in Medicine*. 2023;42(24):4458-4483. doi: 10.1002/sim.9870

APPENDIX A. FIGURES

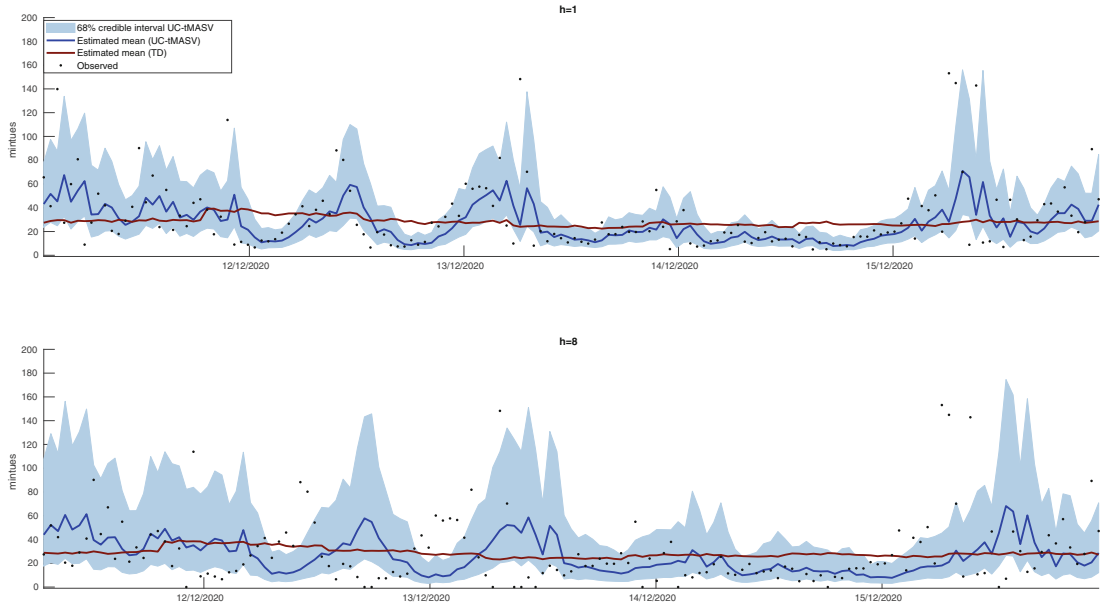

FIGURE A1 The observed waiting time, mean predicted waiting time and its 68% credible interval at $h = 1$ and 8 from the UC-tMSV, and TD models.

APPENDIX B. MCMC ALGORITHMS

In this section, we present MCMC samplers for estimating a set of state space (SS) models with flexible error structures described in Section 2. Without loss of generality, we keep the conditional mean as stated in Equation (3):

$$\mu_t = \mu_{t-1} + \varepsilon_t^{\mu}, \varepsilon_t^{\mu} \sim \mathcal{N}(0, \sigma_{\mu}^2).$$

We present MCMC samplers for the models with normal, t-distributed, SV, and MA errors. The estimation of their variants where errors are the combination of the specifications should be readily implemented.

B.1 SS with a normal distribution

For convenience, we provide a partial representation of the state space model with a normal distribution here. A detailed explanation of the notation can be found in Section 2.

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t^{y}, \quad \varepsilon_t^{y} \sim \mathcal{N}(0, \sigma_y^2) \\ \mu_t &= \mu_{t-1} + \varepsilon_t^{\mu}, \quad \varepsilon_t^{\mu} \sim \mathcal{N}(0, \sigma_{\mu}^2). \end{aligned}$$

To facilitate the discussion on estimation, we stack $y_t, \mu_t, \varepsilon_t^{y}$, and ε_1^{μ} over time, that is, $\mathbf{y} = (y_1, \dots, y_T)'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$, $\boldsymbol{\varepsilon}^{y} = (\varepsilon_1^{y}, \dots, \varepsilon_T^{y})'$, $\boldsymbol{\varepsilon}^{\mu} = (\varepsilon_1^{\mu}, \dots, \varepsilon_T^{\mu})'$. The model can be expressed in a concise matrix form as follows:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}^{y}, \quad \boldsymbol{\varepsilon}^{y} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2 \mathbf{I}_T), \tag{B1}$$

$$\mathbf{H}_{\mu}\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}^{\mu}, \quad \boldsymbol{\varepsilon}^{\mu} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mu}^2 \mathbf{I}_T); \tag{B2}$$

where $\mathbf{I}_T$ is a $T \times T$ identity matrix, $\mathbf{H}_{\mu} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & 1 \end{bmatrix}$, $\boldsymbol{\mu}_0 = \begin{bmatrix} \mu_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

It is worth noting that the logarithm of the waiting time $\mathbf{y}$ can be decomposed into the observed waiting time $\mathbf{y}^o$ and the unobserved waiting time $\mathbf{y}^u$ as follows

$$\mathbf{y} = \mathbf{M}_o\mathbf{y}^o + \mathbf{M}_u\mathbf{y}^u,$$

where $\mathbf{y}^o$ is a $T_o \times 1$ vector, and $\mathbf{y}^u$ is a $T_u \times 1$ vector. Matrices $\mathbf{M}_o$ and $\mathbf{M}_u$, containing values of zeros and ones, are the $T \times T_o$ and $T \times T_u$ matrices associated with $\mathbf{y}^o$ and $\mathbf{y}^u$, and $T = T_u + T_o$. For example, suppose that $\mathbf{y}$ contains three waiting time periods, and the waiting time at period 2 is unobserved. In this case, the matrices $\mathbf{M}_o$ and $\mathbf{M}_u$ take the following forms:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} y_2.$$

The posterior draws of all model parameters can be obtained by sampling sequentially from:

1. $p(\boldsymbol{\mu}|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2) = p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2)$,
2. $p(\mu_0|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \sigma_\mu^2, \sigma_y^2) = p(\mu_0|\boldsymbol{\mu}, \sigma_\mu^2)$,
3. $p(\sigma_\mu^2|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_y^2) = p(\sigma_\mu^2|\boldsymbol{\mu}, \mu_0)$,
4. $p(\sigma_y^2|, \boldsymbol{\mu}, \mathbf{y}^o, \mathbf{y}^u \mu_0, \sigma_\mu^2) = p(\sigma_y^2|\mathbf{y}, \boldsymbol{\mu})$,
5. $p(\mathbf{y}^u|\mathbf{y}^o, \sigma_y^2, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2) = p(\mathbf{y}^u|\mathbf{y}^o, \sigma_y^2, \boldsymbol{\mu})$.

Sampling $\boldsymbol{\mu}$

A conventional approach for estimating the state vector μ_t is the Kalman filter.[40] However, in this article, we adopt the precision-based algorithm proposed by Chan and Jeliazkov.[34] This algorithm has several advantages. First, it samples all states ($p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2)$) in one step instead of one-at-a-time ($p(\mu_t|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2, \{\mu_j\}_{j\neq t})$), which improves the efficiency of the MCMC sampler.[33,34,41] Given the time periods of 8400 in our empirical study, the recursive forecast exercise would become excessively time-consuming if we sample one-at-a time for the state vectors. Second, this algorithm takes advantage of the sparse and banded structure of matrices, particularly $\mathbf{H}_\mu$, which speeds up computation and reduces storage costs.[42]

The prior of $\boldsymbol{\mu}$ is obtained through the state Equation (B2) by pre-multiplying both sides with $\mathbf{H}_\mu$, resulting in:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{H}_\mu^{-1}\boldsymbol{\mu}_0 + \mathbf{H}_\mu^{-1}\boldsymbol{\varepsilon}^\mu, & \boldsymbol{\varepsilon}^\mu &\sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2\mathbf{I}_T), \\ &= \tilde{\boldsymbol{\mu}}_0 + \tilde{\boldsymbol{\varepsilon}}^\mu, & \tilde{\boldsymbol{\varepsilon}}^\mu &\sim \mathcal{N}(0, \boldsymbol{\Omega}_\mu); \end{aligned} \tag{B3}$$

where $\tilde{\boldsymbol{\mu}}_0 = \mathbf{H}_\mu^{-1}\boldsymbol{\mu}_0$, $\tilde{\boldsymbol{\varepsilon}}^\mu = \mathbf{H}_\mu^{-1}\boldsymbol{\varepsilon}^\mu$, and $\boldsymbol{\Omega}_\mu^{-1} = \sigma_\mu^2\mathbf{H}_\mu'\mathbf{H}_\mu$.

Equation (B3) implies that the prior density of $\boldsymbol{\mu}$ is a normal distribution, that is, $\boldsymbol{\mu} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_0, \boldsymbol{\Omega}_\mu)$ with the log-prior as follows:

$$\log p(\boldsymbol{\mu}|\mu_0, \sigma_\mu^2) = c_1 - \frac{T}{2}\log\sigma_\mu^2 - \frac{1}{2}(\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_0)'\boldsymbol{\Omega}_\mu^{-1}(\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_0); \tag{B4}$$

where c_1 is a normalization term independent of $\boldsymbol{\mu}$.

The log-likelihood function is obtained from the mean Equation (B1), which is:

$$\log p(\mathbf{y}|\boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_y^2) = c_2 - \frac{T}{2}\log(\sigma_y^2) - \frac{1}{2\sigma_y^2}(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{y} - \boldsymbol{\mu}); \tag{B5}$$

where c_2 is a normalization term independent of $\boldsymbol{\mu}$ and other model parameters.

Applying Bayes' theorem, the log-posterior distribution is

$$\log p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2) = c_3 - \frac{1}{2}\left(\boldsymbol{\mu}'(\sigma_y^{-2}\mathbf{I}_T + \boldsymbol{\Omega}_\mu^{-1})\boldsymbol{\mu} - 2\boldsymbol{\mu}'(\sigma_y^{-2}\mathbf{y} + \boldsymbol{\Omega}_\mu^{-1}\tilde{\boldsymbol{\mu}}_0)\right); \tag{B6}$$

where c_3 is a normalization term independent of $\boldsymbol{\mu}$. Equation (B6) indicates that the posterior of $\boldsymbol{\mu}$ follows a normal distribution $\mathcal{N}(\mathbf{m}_\mu, \mathbf{K}_\mu^{-1})$ where:

$$\mathbf{K}_\mu = (\sigma_y^{-2}\mathbf{I}_T + \boldsymbol{\Omega}_\mu^{-1}); \quad \mathbf{m}_\mu = \mathbf{K}_\mu^{-1}(\sigma_y^{-2}\mathbf{y} + +\boldsymbol{\Omega}_\mu^{-1}\tilde{\boldsymbol{\mu}}_0).$$

Sampling $\boldsymbol{\mu_0}$

The log-likelihood function is obtained from the state equation at $t = 1$, which is:

$$\log p(\mu_1|\mu_0, \sigma_\mu^2) = -\frac{1}{2}\log(2\pi\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2}(\mu_1 - \mu_0)'(\mu_1 - \mu_0).$$

Given the prior $\mu_0 \sim N(0, V_\mu)$, the posterior density of μ_0 is a normal distribution $\mathcal{N}(m_0, K_0^{-1})$ where

$$K_0 = \sigma_u^{-2} + V_\mu^{-1}; \quad m_0 = K_0^{-1}\sigma_\mu^{-2}\mu_1.$$

Sampling σ_μ^2

The log-likelihood function is obtained from the state Equation (B2):

$$\log p(\boldsymbol{\mu}|\mu_0, \sigma_\mu^2) = -\frac{T}{2}\log(2\pi\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2}\sum_{t=1}^{T}(\mu_t - \mu_{t-1})^2.$$

Given the prior of σ_μ^2 is an inverse gamma $\sigma_\mu^2 \sim \mathcal{IG}(\alpha_\mu, \eta_\mu)$:

$$\log(\sigma_\mu^2) = c_4 - (\alpha_\mu - 1)\log(\sigma_\mu^2) - \frac{\eta_\mu}{\sigma_\mu^2};$$

where c_4 is a normalization term independent of σ_μ^2.

Applying Bayes' theorem, the posterior density of σ_μ^2 follows an inverse gamma distribution $\mathcal{IG}(\hat{\alpha}_\mu, \hat{\eta}_\mu)$, where:

$$\hat{\alpha}_\mu = \alpha_\mu + \frac{T}{2}; \qquad \hat{\eta}_\mu = \eta_\mu + \frac{1}{2}\sum_{t=1}^{T}(\mu_t - \mu_{t-1})^2.$$

Sampling σ_y^2

Similar to σ_μ^2, the posterior of σ_y^2 is an inverse gamma $\mathcal{IG}(\hat{\alpha}_y, \hat{\eta}_y)$ where

$$\hat{\alpha}_y = \alpha_y + \frac{T}{2}; \quad \hat{\eta}_y = \eta_y + \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{y} - \boldsymbol{\mu}).$$

Sampling $\mathbf{y}^u$

The log-likelihood of the observed data from Equation (B1) can be re-parameterized in terms of the observed data $\mathbf{y}$ and the missing values $\mathbf{y}^*$ as follows

$$\begin{aligned}\log p(\mathbf{y}|\boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_y^2) &= p(\mathbf{y}^u, \mathbf{y}^o|\boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_y^2)\\ &= c_2 - \frac{T}{2}\log(\sigma_y^2) - \frac{1}{2\sigma_y^2}\left((\mathbf{M}_o\mathbf{y}^o + \mathbf{M}_u\mathbf{y}^u) - \boldsymbol{\mu}\right)'\left((\mathbf{M}_o\mathbf{y}^o + \mathbf{M}_u\mathbf{y}^u) - \boldsymbol{\mu}\right)\\ &= c_2 - \frac{T}{2}\log(\sigma_y^2) - \frac{1}{2\sigma_y^2}(\boldsymbol{\mu} - \mathbf{M}_o\mathbf{y}^o)'(\boldsymbol{\mu} - \mathbf{M}_o\mathbf{y}^o) - \frac{1}{2\sigma_y^2}\left(\mathbf{y}^{u'}\mathbf{M}_u'\mathbf{M}_u\mathbf{y}^u - 2\mathbf{y}^{u'}\mathbf{M}_u'(\boldsymbol{\mu} - \mathbf{M}_o\mathbf{y}^o)\right).\end{aligned}$$

Given the prior of $\mathbf{y}^u$ is $\mathcal{N}(\boldsymbol{\theta}_y, \mathbf{V}_y)$, where the posterior distribution is a normal distribution, $\mathcal{N}(\mathbf{m}_y, \mathbf{K}_y^{-1})$, with

$$\mathbf{K}_y = \sigma_y^2\mathbf{M}_u'\mathbf{M}_u + \mathbf{V}_y^{-1}; \qquad \mathbf{m}_y = \mathbf{K}_y^{-1}\left(\sigma_y^{-2}\mathbf{M}_u'(\boldsymbol{\mu} - \mathbf{M}_o\mathbf{y}^o) + \mathbf{V}_y^{-1}\theta_y\right).$$

B.2 SS with a *t* distributed error

$$y_t = \mu_t + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \lambda_t \sigma_y^2), \quad \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2),$$
$$\mu_t = \mu_{t-1} + \varepsilon_t^\mu, \varepsilon_t^\mu \sim \mathcal{N}(0, \sigma_\mu^2).$$

Similar to the SS model with a normal error, we represent a SS model with a *t* distributed error in a compact matrix form as follows:

$$\mathbf{y} = \mathbf{M}_o \mathbf{y}^o + \mathbf{M}_u \mathbf{y}^u = \boldsymbol{\mu} + \boldsymbol{\varepsilon}^y, \qquad \boldsymbol{\varepsilon}^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_y),$$
$$\mathbf{H}_\mu \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}^\mu, \qquad \boldsymbol{\varepsilon}^\mu \sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{I}_T);$$

where $\boldsymbol{\Omega}_y = \text{diag}(\sigma_y^2\lambda_1, \ldots, \sigma_y^2\lambda_T) = \text{diag}(\sigma_y^2\boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$.

The model parameters are obtained via a MCMC sampler as follows:

1. $p(\boldsymbol{\mu}|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\lambda}, \nu) = p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\lambda}, \nu)$,
2. $p(\mu_0|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\lambda}, \nu) = p(\mu_0|\boldsymbol{\mu}, \sigma_\mu^2)$,
3. $p(\sigma_\mu^2|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_y^2, \boldsymbol{\lambda}, \nu) = p(\sigma_\mu^2|\boldsymbol{\mu}, \mu_0)$,
4. $p(\sigma_y^2|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \boldsymbol{\lambda}, \nu) = p(\sigma_y^2|\mathbf{y}, \boldsymbol{\mu})$,
5. $p(\mathbf{y}^u|\mathbf{y}, \sigma_y^2, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \boldsymbol{\lambda}, \nu) = p(\mathbf{y}^u|\mathbf{y}, \sigma_y^2, \boldsymbol{\mu})$,
6. $p(\boldsymbol{\lambda}|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\mu}, \nu) = \prod_{t=1}^T p(\lambda_t|\nu, \mathbf{y}_t, \sigma_y^2, \mu_t)$,
7. $p(\nu|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\mu}, \boldsymbol{\lambda}) = p(\nu|\boldsymbol{\lambda})$.

We now present the sampling technique for two additional parameters in block 6 and block 7 of the MCMC sampler, that is, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ and τ. Note that $(\lambda_1, \ldots \lambda_T)$ are conditionally independent, and we sample each of λ_t sequentially from the posterior density $\mathcal{IG}(\widehat{\alpha}_{\eta_t}, \widehat{\beta}_{\eta_t})$, where

$$\widehat{\alpha}_{\lambda_t} = \frac{\nu+1}{2}, \quad \widehat{\beta}_{\lambda_t} = \frac{1}{2}\left(\nu + \frac{1}{\sigma_y^2}(y_t - \mu_t)^2\right).$$

Applying Bayes' theorem, we can obtain the log-posterior density of ν as follows:

$$\log p(\nu|\boldsymbol{\lambda}) \propto \log p(\nu) + \log p(\boldsymbol{\lambda}|\nu)$$
$$\propto \frac{T\nu}{2}\log\left(\frac{\nu}{2}\right) - T\log\Gamma(\nu/2) - \left(\frac{\nu}{2}+1\right)\sum_{t=1}^T \log(\lambda_t) - \frac{\nu}{2}\sum_{t=1}^T \lambda_t^{-1};$$

for $2 < \nu < 100$, $\Gamma(.)$ is a gamma function. Obtaining the first and second derivatives of the log-density with respect to ν is straightforward and can be computed as follows:

$$\frac{d\log p(\nu|\boldsymbol{\nu})}{d\nu} = \frac{T}{2}\log(\nu/2) + \frac{T}{2} - \frac{T}{2}\Psi(\nu/2) - \frac{1}{2}\sum_{t=1}^T \log\lambda_t - \frac{1}{2}\sum_{t=1}^T \lambda_t^{-1}$$
$$\frac{d^2\log p(\nu|\boldsymbol{\nu})}{d\nu^2} = \frac{T}{2\nu} - \frac{T}{4}\Psi'(\nu/2),$$

where $\Psi(z) = \frac{d}{dz}\log\Gamma(z)$ and $\Psi'(z)\frac{d}{dz}\log\Psi(z)$ are the digamma and trigamma functions, respectively. As the first and second derivatives can be evaluated efficiently, we can maximize $\log p(\nu|\Lambda)$ using Newton-Raphson method to obtain the mode, denoted as $\widehat{\nu}$, and the negative Hessian evaluated at the mode, denoted as K_ν^{-1}. We then use an independence-chain Metropolis-Hastings step with the proposal distribution $\mathcal{N}(\widehat{\nu}, K_\nu^{-1})$, denoted as $q(\nu)$. In this step, given the current draw ν, a candidate draw $\nu^c \sim \mathcal{N}(\widehat{\nu}, K_\nu^{-1})$ is accepted with probability

$$\min\left\{1, \frac{p(\nu^c|\boldsymbol{\lambda})}{p(\nu|\boldsymbol{\lambda})} \times \frac{q(\nu)}{q(\nu^c)}\right\},$$

otherwise we stay at the current draw ν.

B.3 SS with stochastic volatility

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t^y, & \varepsilon_t^y &\sim \mathcal{N}(0, e^{h_t}),\\ \mu_t &= \mu_{t-1} + \varepsilon_t^\mu, & \varepsilon_t^\mu &\sim \mathcal{N}(0, \sigma_\mu^2),\\ h_t &= \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2). \end{aligned}$$

We re-parameterize a SS model with SV in a compact matrix such as

$$\begin{aligned} \mathbf{y} = \mathbf{M}_o\mathbf{y}^o + \mathbf{M}_u\mathbf{y}^u = \boldsymbol{\mu} + \boldsymbol{\varepsilon}^y, \quad & \boldsymbol{\varepsilon}^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_y),\\ \mathbf{H}_\mu\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}^\mu, \quad & \boldsymbol{\varepsilon}^\mu \sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2\mathbf{I}_T),\\ \mathbf{H}_\mathbf{h}\mathbf{h} = \mathbf{h}_0 + \boldsymbol{\varepsilon}^\mathbf{h} \quad & \boldsymbol{\varepsilon}^\mathbf{h} \sim N(0, \boldsymbol{\Omega}_h); \end{aligned}$$

where $\boldsymbol{\Omega}_y = \text{diag}(e^{h_1}, \ldots\ e^{h_T})$, $\boldsymbol{\Omega}_h = \text{diag}(\sigma_h^2/(1-\sigma_h^2), \sigma_h^2, \ldots\ \sigma_h^2)$, $\mathbf{H}_\mathbf{h} = \begin{bmatrix} 1 & 0 & \ldots & 0\\ -\phi_h & 1 & \ldots & 0\\ \vdots & \ddots & \ddots & \vdots\\ 0 & \ldots & -\phi_h & 1 \end{bmatrix}$, $\mathbf{h}_0 = \begin{bmatrix} \mu_h\\ \mu_h(1-\phi_h)\\ \vdots\\ \mu_h(1-\phi_h) \end{bmatrix}$.

The posterior distribution of model parameters are sampled from

1. $p(\boldsymbol{\mu}|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \mathbf{h}, \sigma_h^2, \mu_h, \phi_h) = p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2)$,
2. $p(\mu_0|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \sigma_\mu^2, \sigma_y^2, \mathbf{h}, \mu_h, \phi_h) = p(\mu_0|\boldsymbol{\mu}, \sigma_\mu^2)$,
3. $p(\sigma_\mu^2|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \mathbf{h}, \sigma_h^2, \mu_h, \phi_h) = p(\sigma_\mu^2|\boldsymbol{\mu}, \mu_0)$,
4. $p(\mathbf{y}^u|\mathbf{y}^o, \mathbf{h}, \sigma_h^2, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \mu_h, \phi_h) = p(\mathbf{y}^u|\mathbf{y}, \sigma_y^2, \boldsymbol{\mu})$,
5. $p(\mathbf{h}|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_h^2, \mu_h, \phi_h) = p(\mathbf{h}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2, \mu_h, \phi_h)$,
6. $p(\mu_h|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_h^2, \mathbf{h}, \phi_h) = p(\mu_h|\mathbf{h}, \mu_h, \phi_h)$,
7. $p(\phi_h|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \sigma_h^2, \mathbf{h}, \mu_h) = p(\phi_h|\mathbf{h}, \mu_h, \sigma_h^2)$,
8. $p(\sigma_h^2|\boldsymbol{\mu}, \mathbf{y}^o, \mathbf{y}^u\mu_0, \mathbf{h}, \sigma_\mu^2) = p(\sigma_y^2|\mathbf{h}, \mu_h, \phi_h)$.

The additional blocks of the MCMC are for SV process (block 5-8). The log-likelihood function of observed data $\mathbf{y}$ can be represented:

$$\begin{aligned} \log p(\mathbf{y}|\boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \mathbf{h}) &= c_2 - \frac{1}{2}\sum_{t=1}^{T} h_t - \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Omega}_y^{-1}(\mathbf{y}-\boldsymbol{\mu});\\ &= c_2 - \frac{1}{2}\sum_{t=1}^{T} h_t - \frac{1}{2}\sum_{t=1}^{T}(\tilde{y}_t - \mu_t)'\mathrm{e}^{-h_t}(\tilde{y}_t - \mu_t). \end{aligned}$$

The conditional posterior distribution for the log-volatility, $\mathbf{h}$, is not standard. To address this, we adopt the auxiliary mixture sampling approach proposed by Kim et al for block 5.[27] The key idea of this approach is to transform the mean equation into a linear model in terms of h_t, resulting in the following formulation:

$$\log(\tilde{y} - \mu_t)^2 = \log(\tilde{y})^* = h_t + \log\tilde{\varepsilon}_t \quad \tilde{\varepsilon}_t \sim \mathcal{N}(0, 1).$$

Kim et al[27] demonstrated that the likelihood of $\log(\tilde{y})^*$ follows a log-normal distribution which can be approximated by using the mixture of seven normal distribution with mean and variables defined in Table 4 of their article. After using the auxiliary mixture sampling to approximate the likelihood, we apply the precision-based algorithm by Chan and Jeliazkov to sample log-volatility $\mathbf{h}$.

Sampling μ_h and σ_h^2 (blocks 6 and 8) from the their posteriors are straightforward as both of the posteriors are standard. The posterior of μ_h is a normal distribution, that is, $\mathcal{N}(\hat{m}_{\mu_h}, K_{\mu_h}^{-1})$ with

$$K_{\mu_h} = V_{\mu_h}^{-1} + \mathbf{X}'_{\mu_h}\boldsymbol{\Omega}_h^{-1}\mathbf{X}_{\mu_h}, \quad \hat{m}_{\mu_h} = K_{\mu_h}^{-1})(V_{\mu_h}^{-1}m_{\mu_h} + \mathbf{X}'_{\mu_h}\boldsymbol{\Omega}_h^{-1}\mathbf{Z}_{\mu_h});$$

where $\mathbf{X}'_{\mu_h} = (1, 1-\phi_h, \ldots, 1-\phi_h)'$, $\mathbf{Z}_{\mu_h} = (h_1, h_2 - \phi_h h_1, \ldots, h_T - \phi_h h_{T-1})'$.

The posterior of σ_h^2 is an inverse gamma, that is, $\mathcal{IG}(\hat{\alpha}_h, \hat{\eta}_h)$ where

$$\hat{\alpha}_h = \alpha_h + \frac{T}{2}; \quad \hat{\eta}_h = \eta_h + \frac{[(1-\phi_h^2)(h_1-\mu_h)^2 + \sum_{t=2}^{T}(h_t - \mu_h - \phi_h(h_{t-1}-\mu_h))^2]}{2}.$$

For block 7—sampling ϕ_h, the posterior of ϕ_h is given by

$$p(\sigma_y^2|\mathbf{h}, \mu_h, \phi_h) \propto p(\phi_h)g(\phi_h)\exp\left(-\frac{1}{2\sigma_h^2}(h_t - \mu_h - \phi_h(h_{t-1}-\mu_h))^2\right);$$

where $p(\phi)$ is the truncated normal prior given in Section 2, $g(\phi_h) = (1-\phi_h^2)^{1/2}\exp\left(-\frac{1}{2\sigma_h^2}(1-\phi_h^2)(h_1-\mu_h)^2\right)$. Since the posterior of ϕ_h is non-standard, we use an independence-chain Metropolis-Hastings step with a proposal truncated normal distribution $\mathcal{N}(\hat{\phi}_h, K_{\phi_h}^{-1})\mathbf{1}(|\phi_h|<1)$ with $K_{\phi_h} = V_{\phi_h}^{-1} + \mathbf{X}'_{\phi_h}\mathbf{X}_{\phi_h}/\sigma_h^2$, $\hat{\phi}_h = K_{\phi_h}^{-1}(V_{\phi_h}^{-1}\phi_{h_0} + \mathbf{X}'_{\phi_h}\mathbf{Z}_{\phi_h}/\sigma_h^2)$, where $\mathbf{X}_{\phi_h} = (h_1-\mu_h, \ldots, h_{T-1}-\mu_h)'$ and $\mathbf{Z}_{\phi_h} = (h_2-\mu_h, \ldots, h_T-\mu_h)'$. Given the current draw ϕ_h, a proposal ϕ_h^* from the proposal truncated normal distribution is accepted with probability

$$\min\left\{1, \frac{g(\phi_h^*)}{g(\phi_h)}\right\};$$

otherwise we stay at the current state ϕ_h.

B.4 SS with moving average

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t^y, & \varepsilon_t^y &\sim \mathcal{N}(0, \sigma_y^2),\\ \epsilon_t^y &= u_t + \psi u_{t-1}, & u_t &\sim \mathcal{N}(0, \sigma_y^2),\\ \mu_t &= \mu_{t-1} + \varepsilon_t^\mu, & \varepsilon_t^\mu &\sim \mathcal{N}(0, \sigma_\mu^2). \end{aligned}$$

We re-parametrize the model specification of SS-MA in a compact form of matrices:

$$\mathbf{y} = \mathbf{M}_o\mathbf{y}^o + \mathbf{M}_u\mathbf{y}^u = \boldsymbol{\mu} + \mathbf{H}_{\boldsymbol{\psi}}\mathbf{u} \qquad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\mathbf{I}_T), \tag{B7}$$

$$\mathbf{H}_{\boldsymbol{\mu}}\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathbf{0}} + \boldsymbol{\varepsilon}^{\boldsymbol{\mu}} \qquad \boldsymbol{\varepsilon}^{\boldsymbol{\mu}} \sim \mathcal{N}(0, \sigma_\mu^2\mathbf{I}_T), \tag{B8}$$

where $\mathbf{H}_{\boldsymbol{\psi}} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \psi & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \psi & 1 \end{bmatrix}$

1. $p(\boldsymbol{\mu}|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2, \psi) = p(\boldsymbol{\mu}|\mathbf{y}, \mu_0, \sigma_\mu^2, \sigma_y^2, \psi)$,
2. $p(\mu_0|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \sigma_\mu^2, \sigma_y^2, \psi) = p(\mu_0|\boldsymbol{\mu}, \sigma_\mu^2)$,
3. $p(\sigma_\mu^2|\mathbf{y}^o, \mathbf{y}^u, \boldsymbol{\mu}, \mu_0, \sigma_y^2, \psi) = p(\sigma_\mu^2|\boldsymbol{\mu}, \mu_0)$,
4. $p(\sigma_y^2|\boldsymbol{\mu}, \mathbf{y}^o, \mathbf{y}^u\mu_0, \sigma_\mu^2, \psi) = p(\sigma_y^2|\tilde{\mathbf{y}}, \boldsymbol{\mu})$,
5. $p(\mathbf{y}^u|\mathbf{y}^o, \sigma_y^2, \boldsymbol{\mu}, \mu_0, \sigma_\mu^2, \psi) = p(\mathbf{y}^u|\mathbf{y}^o, \sigma_y^2, \boldsymbol{\mu}, \psi)$
6. $p(\psi|\mathbf{y}^o, \mathbf{y}^u, \mu_0, \sigma_\mu^2, \sigma_y^2, \boldsymbol{\mu}) = p(\psi|\mathbf{y}, \sigma_y^2, \boldsymbol{\mu})$.

To efficiently estimate the SS model with MA, we employ the algorithm proposed by Chan.[43] The first step involves transforming the mean Equation (B7) in such a way that the error process becomes uncorrelated. This is achieved by pre-multiplying both sides of Equation (B7) with $\mathbf{H}_{\boldsymbol{\psi}}^{-1}$. As a result, Equation (B7) can be expressed as:

$$\hat{\mathbf{y}} = \hat{\boldsymbol{\mu}} + \mathbf{u} \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\mathbf{I}_T),$$

where $\hat{\mathbf{y}} = \mathbf{H}_{\boldsymbol{\psi}}^{-1}\mathbf{y}$, $\hat{\boldsymbol{\mu}} = \mathbf{H}_{\boldsymbol{\psi}}^{-1}\boldsymbol{\mu}$.

We now sample $\hat{\boldsymbol{\mu}}$ instead of $\boldsymbol{\mu}$ in block 1. Once we have a draw of $\hat{\boldsymbol{\mu}}$, we can obtain a draw for $\boldsymbol{\mu}$ easily by pre-multiplying $\hat{\boldsymbol{\mu}}$ by $\mathbf{H}_{\boldsymbol{\psi}}$. As mentioned in the SS with a normal error, the prior of $\boldsymbol{\mu}$ is a normal distribution, that is, $\boldsymbol{\mu} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_0, \boldsymbol{\Omega}_{\mu})$. By a simple change of variable, we have $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\mathbf{H}_{\boldsymbol{\psi}}^{-1}\tilde{\boldsymbol{\mu}}_0, \mathbf{H}_{\boldsymbol{\psi}}^{-1}\boldsymbol{\Omega}_{\mu}\mathbf{H}_{\boldsymbol{\psi}}^{-1'})$. A similar approach is applied for block 5 to sample $\mathbf{y}^{\mu}$.

Given the prior $p(\psi)$, the log posterior of ψ in block 6 is as follows:

$$\begin{aligned}\log p(\psi|\mathbf{y}, \boldsymbol{\mu}, \sigma_y^2) &= \log p(\psi) \log(\mathbf{y}|\boldsymbol{\psi}, \sigma_y^2) \\ &= c_5 + \log p(\psi) - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{H}_{\boldsymbol{\psi}}^{-1}\boldsymbol{\Omega}_{\mu}\mathbf{H}_{\boldsymbol{\psi}}^{-1'})(\mathbf{y} - \boldsymbol{\mu}).\end{aligned}$$

Since the dimension of ψ is typically low, we can use numerical optimization routines to obtain the mode and negative Hessian of $\log p(\psi|\mathbf{y}, \boldsymbol{\mu}, \sigma_y^2)$ evaluated at the mode. These values are denoted as $\hat{\psi}$ and $\mathbf{K}_{\psi}$, respectively. Then, draws from $p(\psi|\mathbf{y}, \boldsymbol{\mu}, \sigma_y^2)$ can be obtained using an independence-chain Metropolis-Hastings step with the proposal of $\mathcal{N}(\hat{\psi}, \mathbf{K}_{\psi}^{-1})$, denoted as $q(\psi)$. In this step, given the current draw ψ, a candidate draw $\psi^c \sim \mathcal{N}(\hat{\psi}, \mathbf{K}_{\psi}^{-1})$ is accepted with probability

$$\min\left\{1, \frac{p(\psi^c|\mathbf{y}, \boldsymbol{\mu}, \sigma_y^2)}{p(\psi|\mathbf{y}, \boldsymbol{\mu}, \sigma_y^2)} \times \frac{q(\psi)}{q(\psi^c)}\right\},$$

otherwise we stay at the current draw ψ.

APPENDIX C. EVIDENCE ON CONVERGENCE AND EFFICIENCY OF THE MCMC ALGORITHMS

To evaluate the convergence and efficiency of the MCMC algorithms, we compute the Geweke[37] convergence diagnostic (CD). Additionally, we provide visual inspections through traceplots of selected parameters and boxplots of inefficiency factors.

The *P*-values of the Geweke[37] convergence diagnostic for selected parameters are presented in Table C1. Due to the large number of states for time-varying parameters such as volatility (h_t), we present the results for a few selected states (eg, h_7, h_{4010}, h_{8383}) as illustrations. A similar strategy is applied to the "unobserved" waiting time (y_t^{μ}). A *P*-value of the CD statistic greater than 0.05 indicates that the null hypothesis of convergence to the posterior distribution is not rejected. This indicates that there is no significant evidence against the convergence of the MCMC samples to the posterior distribution.

To evaluate the efficiency of the MCMC samplers in drawing the states over the sample period, we generate boxplots of inefficiency factors for all h_t and y_t^{μ}. The inefficiency factor is the inverse of the relative numerical efficiency measure

TABLE C1 *P*-values of Geweke convergence diagnostic statistic.

Parameters	TD-tMA	UC-tMA	TVAR-tMA	TD-tSV	UC-tSV	TVAR-tSV	TD-tMASV	UC-tMASV	TVAR-tMASV
ψ	0.317	0.305	0.346	-	-	-	0.318	0.300	0.306
ϕ_h	-	-	-	0.317	0.319	0.317	0.317	0.319	0.315
μ_h	-	-	-	0.317	0.318	0.318	0.318	0.320	0.320
σ_h^2	-	-	-	0.322	0.316	0.318	0.319	0.309	0.306
ν	0.313	0.315	0.318	0.319	0.286	0.309	0.327	0.321	0.333
h_7	-	-	-	0.274	0.322	0.310	0.321	0.319	0.307
h_{4010}	-	-	-	0.583	0.880	0.339	0.362	0.351	0.282
h_{8383}	-	-	-	0.309	0.311	0.316	0.310	0.309	0.319
y_7^{μ}	0.317	0.321	0.323	0.325	0.320	0.314	0.323	0.314	0.321
y_{4010}^{μ}	0.318	0.316	0.319	0.329	0.312	0.317	0.329	0.318	0.317
y_{8383}^{μ}	0.324	0.326	0.322	0.321	0.321	0.328	0.327	0.322	0.319

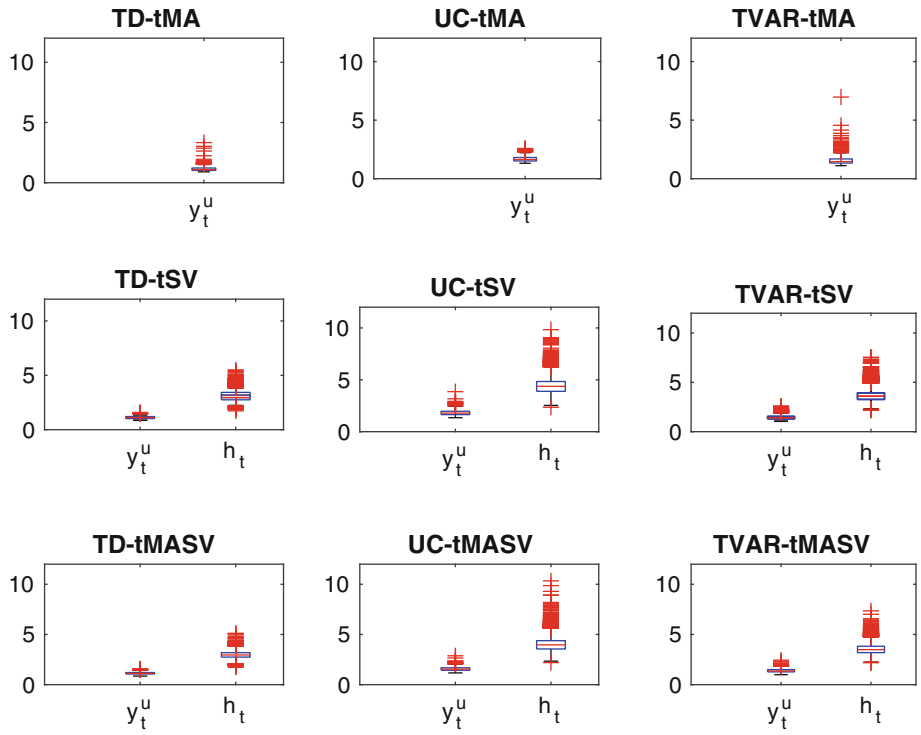

FIGURE C1 Inefficiency factors of y_t^u and h_t under the TD-tMA, UC-tMA, TVAR-tMA, TD-tSV, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

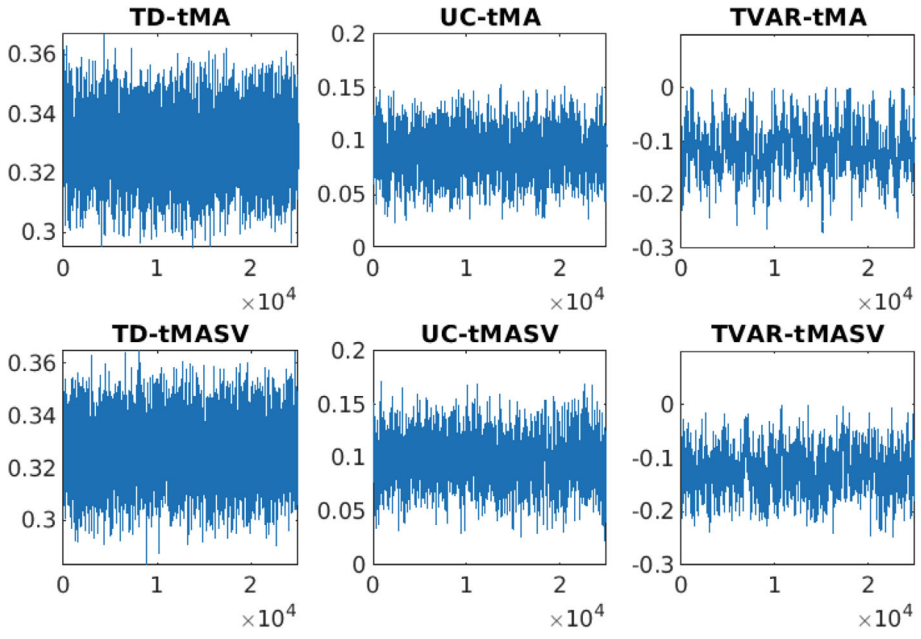

FIGURE C2 Trace plots of the estimated moving average parameter ψ under the TD-tMA, UC-tMA, TVAR-tMA, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

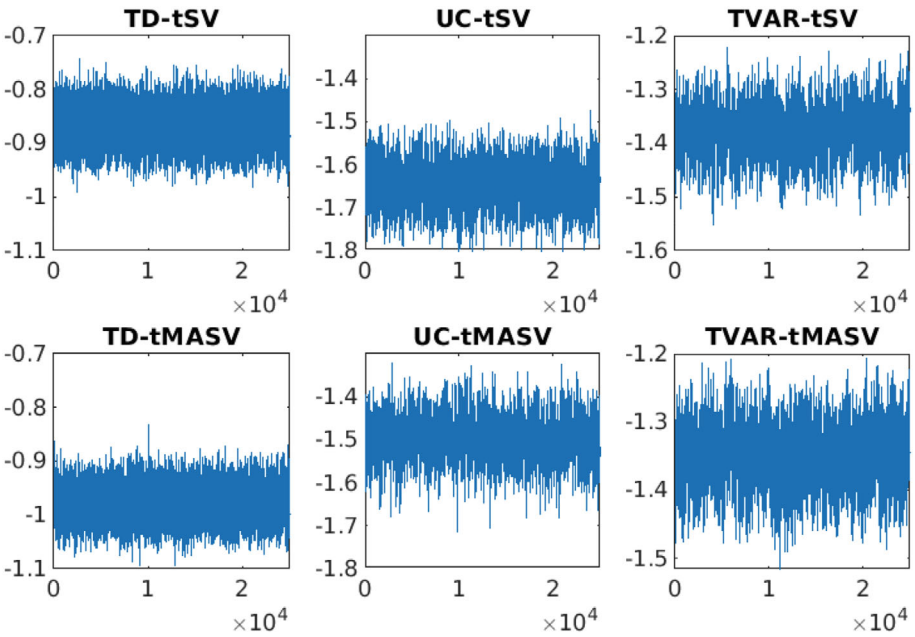

FIGURE C3 Trace plots of the estimated moving average parameter μ_h under the TD-tSV, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

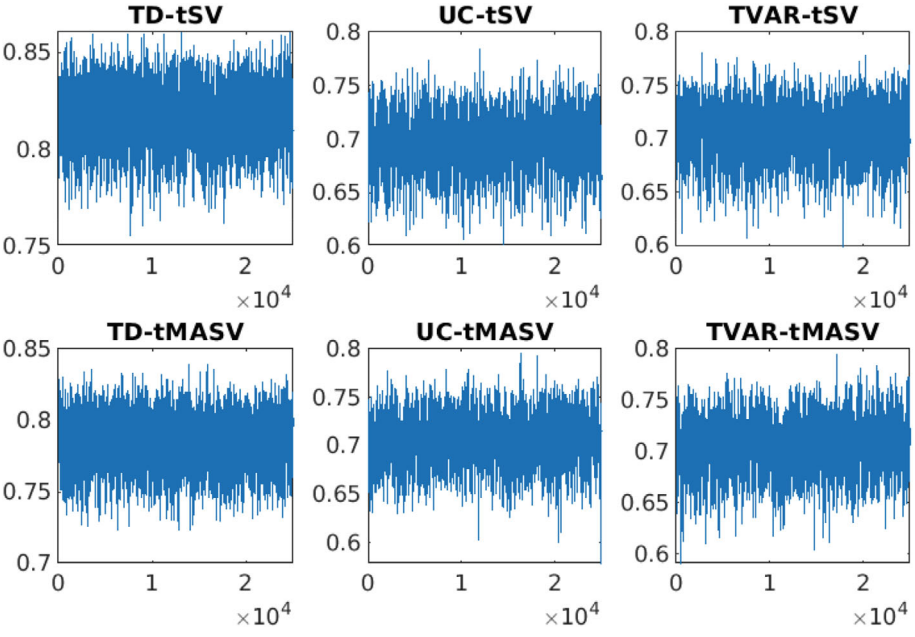

FIGURE C4 Trace plots of the estimated moving average parameter ϕ_h under the TD-tSV, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

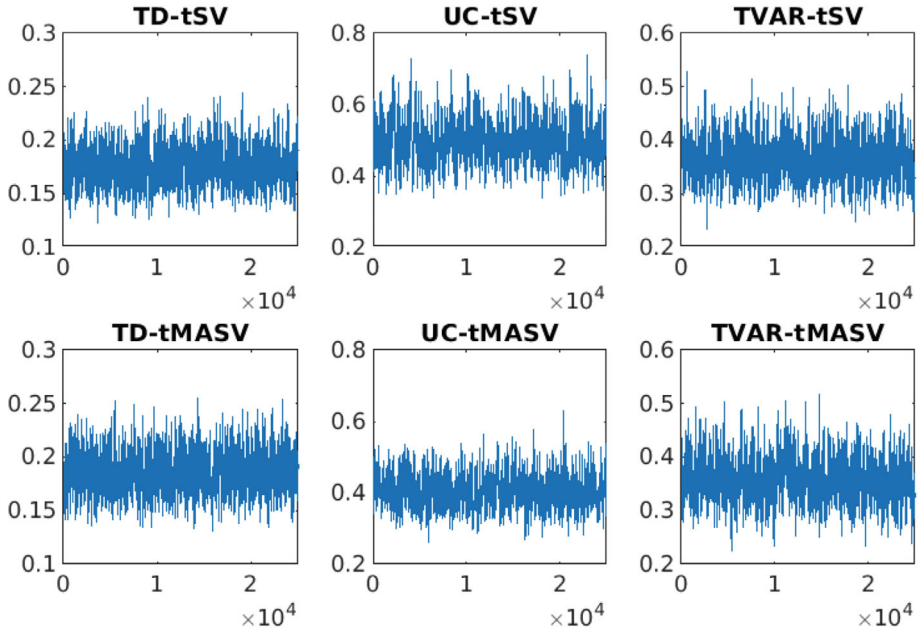

FIGURE C5 Trace plots of the estimated moving average parameter σ_h^2 under the TD-tSV, UC-tSV, TVAR-tSV, TD-tMASV, UC-tMASV, and TVAR-tMASV models.

proposed by Chib,[38] which is defined as

$$1 + 2\sum_{s=1}^{\infty}\rho_s,$$

where ρ_s is the sample autocorrelation at lag s. The inefficiency factors measure the efficiency of the MCMC samplers. They quantity additional draws need to be taken to achieve the same precision as independent draws. For example, an inefficiency factor of 50 after 50 000 draws for a parameter is equivalent to obtaining the same level of precision as 10 000 independent draws from the posterior. Figure C1 reveals that the inefficiency factor is low, less than 10, which is viewed as satisfactory.

The trace plots Figures C2–C5, for selected parameters such as $\psi, \mu_h, \phi_h, \sigma_h^2$ under different model specifications (ie, TD-tSV, UC-tSV, TVAR-tSV, TD-tMA, UC-tMA, TVAR-tMA, TD-tMASV, UC-tMASV, and TVAR-tMASV), exhibit stability.

Overall, the analysis indicates convergence and good mixing of the MCMC samplers. The MCMC samplers are effectively exploring the posterior distributions and obtaining reliable estimates for the parameters of interest.