Double Dynamic Bounded Full-integer-valued GARCH Model

Huaping Chen[1], Qianqian Zhao[1], Liugen Xue[1] and Fukang Zhu[2]

[1] *Henan University and* [2] *Jilin University*

Supplementary Materials

Contents

S1 The Derivatives of Log-Likelihood Function 2

S2 Proof 4

S3 Simulation Results 17

S3.1 Scenario I: Specified TDGV and TDGMV Models . . . 17

S3.2 Scenario II: Misspecified TDGV and TDGMV Models . 25

S3.3 Forecasting Performance Evaluation 25

S4 Some Auxiliary Results 35

S4.1 Auxiliary Lemma 35

S4.2 Auxiliary Discussion with regard to TD2 Distribution . 39

S1 The Derivatives of Log-Likelihood Function

In this part, we give the first and second derivatives of the log-likelihood function given in (3.12). For convenience, we denote $\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})$ with $\ell_t(\boldsymbol{\theta}) = \log \mathbb{P}_n(X_t = x_t|\mathcal{F}_{t-1})$, where

$$\mathbb{P}_n(x_t) = \mathbb{P}_n(X_t = x_t|\mathcal{F}_{t-1}) = \sum_{j=\max(0,-x_t)}^{[(n-x_t)/2]} \binom{n}{x_t+j, j} \alpha_1^{x_t+j} \alpha_2^j (1-\alpha_1-\alpha_2)^{n-x_t-2j}.$$

By Lemma 2 given in supplementary materials S4,

$$\frac{\partial \mathbb{P}_n(x_t)}{\partial \sigma_t^2} = \frac{1}{2}\left(\mathbb{P}_{n-1}(x_t-1) + \mathbb{P}_{n-1}(x_t+1) - 2\mathbb{P}_{n-1}(x_t)\right),$$
$$\frac{\partial \mathbb{P}_n(x_t)}{\partial \mu_t} = \left(\frac{\mu_t}{n} + \frac{1}{2}\right)\mathbb{P}_{n-1}(x_t-1) + \left(\frac{\mu_t}{n} - \frac{1}{2}\right)\mathbb{P}_{n-1}(x_t+1) - \frac{2\mu_t}{n}\mathbb{P}_{n-1}(x_t).$$

Therefore, $\forall \boldsymbol{\theta} \in \Theta$,

$$-1 \le \partial \mathbb{P}_n(x_t)/\partial \sigma_t^2 < 1 \text{ and } -9/2 \le \partial \mathbb{P}_n(x_t)/\partial \mu_t \le 9/2. \tag{S1.1}$$

In subsequent, we give the first derivatives of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is given as $\partial \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \sum_{t=1}^{T} \partial \ell_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$. Specially,

$$\begin{cases} \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_0} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \mu_t}, & \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \beta_1} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \mu_t}\mu_{t-1}, \\ \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \phi_0} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \sigma_t^2}, & \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \psi_1} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \sigma_t^2}\sigma_{t-1}^2, \\ \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_1} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \mu_t}X_{t-1}, & \dfrac{\partial \ell_t(\boldsymbol{\theta})}{\partial \phi_1} = \dfrac{1}{\mathbb{P}_n(x_t)}\dfrac{\partial \mathbb{P}_n(x_t)}{\partial \sigma_t^2}|X_{t-1}|. \end{cases} \tag{S1.2}$$

Therefore, $\forall i = 1, \cdots, 6, \forall j = 1, \cdots, 6$, by (S1.1) and Lemma 1,

$$|\partial \ell_t(\boldsymbol{\theta})/\partial \theta_i| \le 9n\varpi_2^{-1}/2 \text{ and } |\partial \ell_t(\boldsymbol{\theta})/\partial \theta_i \partial \ell_t(\boldsymbol{\theta})/\partial \theta_j| \le 81n^2\varpi_2^{-2}/4. \tag{S1.3}$$

To compute $\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$, we denote

$$
\begin{aligned}
f_1(x_t) &= -\frac{1}{\mathbb{P}_n^2(x_t)}\left(\frac{\partial\mathbb{P}_n(x_t)}{\partial\mu_t}\right)^2 + \frac{1}{\mathbb{P}_n(x_t)}\frac{\partial^2\mathbb{P}_n(x_t)}{\partial\mu_t^2},\\
f_2(x_t) &= -\frac{1}{\mathbb{P}_n^2(x_t)}\frac{\partial\mathbb{P}_n(x_t)}{\partial\mu_t}\frac{\partial\mathbb{P}_n(x_t)}{\partial\sigma_t^2} + \frac{1}{\mathbb{P}_n(x_t)}\frac{\partial^2\mathbb{P}_n(x_t)}{\partial\mu_t\partial\sigma_t^2},\\
f_3(x_t) &= -\frac{1}{\mathbb{P}_n^2(x_t)}\left(\frac{\partial\mathbb{P}_n(x_t)}{\partial\sigma_t^2}\right)^2 + \frac{1}{\mathbb{P}_n(x_t)}\frac{\partial^2\mathbb{P}_n(x_t)}{\partial\sigma_t^2\partial\sigma_t^2},
\end{aligned}
$$

where

- $\dfrac{\partial^2\mathbb{P}_n(x_t)}{\partial\mu_t^2} = \dfrac{1}{n}\mathbb{P}_{n-1}(x_t-1) + \dfrac{1}{n}\mathbb{P}_{n-1}(x_t+1) - \dfrac{2}{n}\mathbb{P}_{n-1}(x_t)$
  $+ \left(\dfrac{\mu_t}{n}+\dfrac{1}{2}\right)\dfrac{\partial\mathbb{P}_{n-1}(x_t-1)}{\partial\mu_t} + \left(\dfrac{\mu_t}{n}-\dfrac{1}{2}\right)\dfrac{\partial\mathbb{P}_{n-1}(x_t+1)}{\partial\mu_t} - \dfrac{2\mu_t}{n}\dfrac{\partial\mathbb{P}_{n-1}(x_t)}{\partial\mu_t}$,
- $\dfrac{\partial^2\mathbb{P}_n(x_t)}{\partial\mu_t\partial\sigma_t^2} = \dfrac{1}{2}\dfrac{\partial\mathbb{P}_{n-1}(x_t-1)}{\partial\mu_t} + \dfrac{1}{2}\dfrac{\partial\mathbb{P}_{n-1}(x_t+1)}{\partial\mu_t} - \dfrac{\partial\mathbb{P}_{n-1}(x_t)}{\partial\mu_t}$,
- $\dfrac{\partial^2\mathbb{P}_n(x_t)}{\partial\sigma_t^2\partial\sigma_t^2} = \dfrac{1}{2}\dfrac{\partial\mathbb{P}_{n-1}(x_t-1)}{\partial\sigma_t^2} + \dfrac{1}{2}\dfrac{\partial\mathbb{P}_{n-1}(x_t+1)}{\partial\sigma_t^2} - \dfrac{\partial\mathbb{P}_{n-1}(x_t)}{\partial\sigma_t^2}$.

Therefore,

$$
\begin{aligned}
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0^2} = f_1(x_t),\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\beta_1} = f_1(x_t)\mu_{t-1},\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\phi_0} = f_2(x_t),\\
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\psi_1} = f_2(x_t)\sigma_{t-1}^2,\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\alpha_1} = f_1(x_t)X_{t-1},\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\phi_1} = f_2(x_t)|X_{t-1}|,\\
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\beta_1^2} = f_1(x_t)\mu_{t-1}^2,\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\beta_1\partial\phi_0} = f_2(x_t)\mu_{t-1},\ \ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\beta_1\partial\psi_1} = f_2(x_t)\sigma_{t-1}^2\mu_{t-1},\\
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\beta_1\partial\alpha_1} = f_1(x_t)X_{t-1}\mu_{t-1},\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\beta_1\partial\phi_1} = f_2(x_t)\mu_{t-1}|X_{t-1}|,\ \ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\phi_0^2} = f_3(x_t),\\
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\phi_0\partial\psi_1} = f_3(x_t)\sigma_{t-1}^2,\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\phi_0\partial\alpha_1} = f_2(x_t)X_{t-1},\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\phi_0\partial\phi_1} = f_3(x_t)|X_{t-1}|,\\
&\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\psi_1^2} = f_3(x_t)\sigma_{t-1}^4,\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\psi_1\partial\alpha_1} = f_2(x_t)X_{t-1}\sigma_{t-1}^2,\ \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\psi_1\partial\phi_1} = f_2(x_t)\sigma_{t-1}^2|X_{t-1}|,
\end{aligned}
$$

$$\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \alpha_1^2} = f_1(x_t) X_{t-1}^2, \ \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \alpha_1 \partial \phi_1} = f_2(x_t) X_{t-1} |X_{t-1}|, \ \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \phi_1^2} = f_3(x_t) |X_{t-1}^2|.$$

Similarly, one can compute $\partial^3 \ell_t(\boldsymbol{\theta}) / \partial \theta_i \theta_j \theta_k, \forall (i,j,k) \in \{1, 2, \cdots, 6\}^3$. Here, we omit them.

S2 Proof

In this part, we give the proofs of Propositions 1-3, Remark 2, Lemma 1 and Theorems 1-2, respectively.

Proposition 1

The first item holds by Proposition 3 in Omair et al. (2016), and it is easy to see the third item holds. Here, we only illustrate the second item. If $\{Z_i\}$ is an independent random variable sequence and $Z_i \sim \text{TD}(1, \alpha_1, \alpha_2)$, then $\{|Z_i|\}$ is a sequence of mutually independent and identically distributed random variables with $\mathbb{P}(|Z_1| = 1) = \alpha_1 + \alpha_2 = 1 - \mathbb{P}(|Z_1| = 0)$. Therefore,

$$\sum_{i=1}^{n} |Z_i| \sim \text{Bin}(n, \alpha_1 + \alpha_2).$$

The proof is complete.

Proposition 2

Denote $r_1 = \alpha_1 + \alpha_2$ and $r_2 = \beta_1 + \beta_2$. By (2) in Gibbs and Su (2002),

$$\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2)) \tag{S2.1}$$
$$= 0.5\sum_{k=-n}^{n}\left|\left(\sum_{j=w_1}^{w_2} C(k,j)\alpha_1^{k+j}\alpha_2^{j}(1-r_1)^{w_3} - \sum_{j=w_1}^{w_2} C(k,j)\beta_1^{k+j}\beta_2^{j}(1-r_2)^{w_3}\right)\right|,$$

where $w_1 = \max(0,-k), w_2 = [(n-k)/2]$, $w_3 = n-k-2j$ and $C(k,j) = \binom{n}{k+j,j}$. If $n=1$,

$$\mathrm{TV}(\mathrm{TD1}(1,\alpha_1,\alpha_2),\mathrm{TD1}(1,\beta_1,\beta_2)) = 0.5\left|\big(r_1-r_2\big)+\big((1-r_1)-(1-r_2)\big)\right|$$
$$\leq \|\tau_1-\tau_2\| = 1-(1-\|\tau_1-\tau_2\|).$$

For all $n>1$, we assume $\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2)) \leq 1-(1-\|\tau_1-\tau_2\|)^n$ and denote $w_1^{'} = \max(0,-k)$ and $w_2^{'} = [(n+1-k)/2]$. It is easy to see that $C_{n+1}(k,j) = C_n(k-1,j) + C_n(k+1,j-1) + C_n(k,j)$. Then we denote $d = n+1-k-2j$ and consider the case of $n+1$ based on (S2.1).

$$\begin{aligned}
&2\mathrm{TV}(\mathrm{TD1}(n+1,\alpha_1,\alpha_2),\mathrm{TD1}(n+1,\beta_1,\beta_2))\\
&= \sum_{k=-(n+1)}^{n+1}\left|\sum_{j=w_1^{'}}^{w_2^{'}} C_{n+1}(k,j)\left(\alpha_1^{k+j}\alpha_2^{j}(1-r_1)^{d} - \beta_1^{k+j}\beta_2^{j}(1-r_2)^{d}\right)\right|\\
&\leq \sum_{k=-(n+1)}^{n+1}\left|\sum_{j=w_1^{'}}^{w_2^{'}} C_{n}(k-1,j)\left(\alpha_1^{k+j}\alpha_2^{j}(1-r_1)^{d} - \beta_1^{k+j}\beta_2^{j}(1-r_2)^{d}\right)\right|\\
&+ \sum_{k=-(n+1)}^{n+1}\left|\sum_{j=w_1^{'}}^{w_2^{'}} C_{n}(k+1,j-1)\left(\alpha_1^{k+j}\alpha_2^{j}(1-r_1)^{d} - \beta_1^{k+j}\beta_2^{j}(1-r_2)^{d}\right)\right|
\end{aligned}$$

$$+\sum_{k=-(n+1)}^{n+1}\left|\sum_{j=w_1'}^{w_2'}C_n(k,j)\left(\alpha_1^{k+j}\alpha_2^j(1-r_1)^d-\beta_1^{k+j}\beta_2^j(1-r_2)^d\right)\right|$$

$$\leq \alpha_1\times 2\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2))+|\alpha_1-\beta_1|$$

$$+\alpha_2\times 2\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2))+|\alpha_2-\beta_2|$$

$$+(1-r_2)\times 2\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2))+\|\tau_1-\tau_2\|.$$

Hence,

$$\mathrm{TV}(\mathrm{TD1}(n+1,\alpha_1,\alpha_2),\mathrm{TD1}(n+1,\beta_1,\beta_2))$$

$$\leq (1-(r_1-r_2))\times\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2))+\|\tau_1-\tau_2\|$$

$$=1-(1-\|\tau_1-\tau_2\|)^{n+1}.$$

To sum up, for all $n\geq 1$, $\mathrm{TV}(\mathrm{TD1}(n,\alpha_1,\alpha_2),\mathrm{TD1}(n,\beta_1,\beta_2))\leq 1-(1-\|\tau_1-\tau_2\|)^n$.

Note that $\|\tau_1-\tau_2\|=|\alpha_1-\beta_1|+|\alpha_2-\beta_2|=(\alpha_1-\beta_1)+(\alpha_2-\beta_2)=(\alpha_1+\alpha_2)-(\beta_1+\beta_2)=r_1-r_2$ for all $\alpha_1\geq\beta_1>n/(2\delta)>0$ and $\alpha_2\geq\beta_2>n/(2\delta)>0$. In addition,

$$r_1-r_2\leq\frac{r_1-r_2}{r_1}=1-\frac{r_2}{r_1}.$$

Hence, $1-(1-\|\tau_1-\tau_2\|)^n=1-(1-(r_1-r_2))^n\leq 1-(1-(1-r_2/r_1))^n=1-(r_2/r_1)^n$. Furthermore, for all $\alpha_1\geq\beta_1>\delta/(2n)>0$ and $\alpha_2\geq\beta_2>$

$n/(2\delta) > 0$, it follows $r_1 \geq r_2 \geq n/\delta > 0$. Thus,

$$\frac{r_1}{r_2} = 1 + \frac{r_1 - r_2}{r_2} \leq \exp\left(\frac{|r_1 - r_2|}{r_2}\right) \leq \exp\left(\frac{|r_1 - r_2|}{n/\delta}\right)$$

by using $\exp(x) \geq 1 + x, \forall x \geq 0$. Therefore,

$$1 - \left(\frac{r_2}{r_1}\right)^n \leq 1 - \exp\left(-n\frac{|r_1 - r_2|}{n/\delta}\right) = 1 - \exp(-\delta|r_1 - r_2|).$$

Thus, $\mathrm{TV}(\mathrm{TD1}(n, \alpha_1, \alpha_2), \mathrm{TD1}(n, \beta_1, \beta_2)) \leq 1 - \exp(-\delta|r_1 - r_2|) = 1 - \exp(-\delta\|\tau_1 - \tau_2\|)$. The proof is complete.

Theorem 1

(1) Let $\boldsymbol{Z_t} = (X_t, \mu_t, \sigma_t^2)^\top$ with

$$\mu_t = \alpha_0 + \alpha_1 X_{t-1} + \beta_1 \mu_{t-1} \text{ and } \sigma_t^2 = \phi_0 + \phi_1 |X_{t-1}| + \psi_1 \sigma_{t-1}^2. \quad \text{(S2.2)}$$

Under Assumption 1, we have $|\beta_1| < 1$ and $|\psi_1| < 1$. Iterating the recursions in (S2.2) backward yields

$$\mu_t = \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=0}^{\infty} \beta_1^i X_{t-1-i} \text{ and } \sigma_t^2 = \frac{\phi_0}{1 - \psi_1} + \phi_1 \sum_{i=0}^{\infty} \psi_1^i |X_{t-1-i}|. \quad \text{(S2.3)}$$

It is easy to see that $\{\mu_t\}$ and $\{\sigma_t^2\}$ converge absolutely by $|\beta_1| < 1$, $0 < \psi_1 < 1$ and the finiteness of $\{X_t\}$,

Let $\{s_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence of Uniform$(0, 1)$ random variables such that $X_t = F^{-1}_{\mu_t, \sigma_t^2}(s_t)$, where F^{-1}_{μ,σ^2} denotes the quantile function of the conditional distribution. Substituting $X_{t-1-i} = F^{-1}_{\mu_{t-1-i}, \sigma^2_{t-1-i}}(s_{t-1-i})$

into (S2.3), there exist some measurable functions $\tilde{\mu}$ and $\tilde{\sigma}^2$ such that $\mu_t = \tilde{\mu}(\ldots, s_{t-2}, s_{t-1})$ and $\sigma_t^2 = \tilde{\sigma}^2(\ldots, s_{t-2}, s_{t-1})$. Thus, $\{X_t\}$ is reexpressed as $X_t = F^{-1}_{\tilde{\mu}(s_{t-1},s_{t-2},\ldots),\,\tilde{\sigma}^2(s_{t-1},s_{t-2},\ldots)}(s_t) = G(s_t, s_{t-1}, s_{t-2}, \ldots)$ for some measurable function G. Therefore,

$$\boldsymbol{Z_t} = (X_t, \mu_t, \sigma_t^2)^\top = \big(G(s_t, s_{t-1}, \ldots),\, \tilde{\mu}(s_{t-1}, \ldots),\, \tilde{\sigma}^2(s_{t-1}, \ldots)\big)^\top,$$

i.e., $\{\boldsymbol{Z_t}\}$ admits a Bernoulli-shift representation with i.i.d. innovations $\{s_t\}$.

(2) By using Theorem 12.0.1 in Meyn and Tweedie (2009), it suffices to show that $\{\boldsymbol{\lambda}_t\}$ is bounded in probability on average. First, $\forall \boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2 = \boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{\lambda}_1 + \boldsymbol{B}\boldsymbol{Y}_1$ by (2.7). By Proposition 1, $\mathrm{E}(\boldsymbol{Y}_1|\boldsymbol{\lambda}_1) = (\mu_1, \mu_1^2/n + \sigma_1^2)^\top$, $\boldsymbol{B}\mathrm{E}(\boldsymbol{Y}_1|\boldsymbol{\lambda}_1) = \boldsymbol{B}\lambda_1 + (0, \phi_1\mu_1^2/n)^\top \le \boldsymbol{B}\lambda_1 + (0, n\phi_1)^\top := \boldsymbol{B}\lambda_1 + \boldsymbol{\delta}_2$. Hence, $\mathrm{E}(\boldsymbol{\lambda}_2|\boldsymbol{\lambda}_1) = \boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{\lambda}_1 + \boldsymbol{B}\mathrm{E}(\boldsymbol{Y}_1|\boldsymbol{\lambda}_1) \le \boldsymbol{\delta} + (\boldsymbol{A}+\boldsymbol{B})\boldsymbol{\lambda}_1$, where $\boldsymbol{\delta} = \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2$. By introduction,

$$\begin{aligned}\mathrm{E}(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_1) &\le \boldsymbol{\delta} + (\boldsymbol{A}+\boldsymbol{B})\mathrm{E}\big(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_1\big) \\ &\le \big(\boldsymbol{I} + (\boldsymbol{A}+\boldsymbol{B}) + \cdots + (\boldsymbol{A}+\boldsymbol{B})^{t-1}\big)\boldsymbol{\delta} + (\boldsymbol{A}+\boldsymbol{B})^t\boldsymbol{\lambda}_1.\end{aligned}$$

By $\rho(\boldsymbol{A}+\boldsymbol{B}) < 1$, then $\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B}$ is nonsingular, $\big[\boldsymbol{I} - (\boldsymbol{A}+\boldsymbol{B})\big]^{-1} = \sum_{t=0}^{+\infty}(\boldsymbol{A}+\boldsymbol{B})^t$ and $(\boldsymbol{A}+\boldsymbol{B})^t \to 0$, as $t \to \infty$. Therefore,

$$\mathrm{E}(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_1) \le \big[\boldsymbol{I} - (\boldsymbol{A}+\boldsymbol{B})\big]^{-1}\boldsymbol{\delta} + (\boldsymbol{A}+\boldsymbol{B})^t\boldsymbol{\lambda}_1.$$

Subsequently,

$$\begin{aligned}
&\mathbb{P}^t(\boldsymbol{\lambda}_{t+1} \in \mathcal{C}|\boldsymbol{\lambda}_1) = \mathbb{P}(\lambda_{1,t+1} \leq k_1, \lambda_{2,t+1} \leq k_2|\boldsymbol{\lambda}_1) \\
&\geq 1 - \mathbb{P}(\lambda_{1,t+1} > k_1|\boldsymbol{\lambda}_1) - \mathbb{P}(\lambda_{2,t+1} > k_2|\boldsymbol{\lambda}_1) \\
&\geq 1 - \frac{\mathrm{E}(\lambda_{1,t+1}|\boldsymbol{\lambda}_1)}{k_1} - \frac{\mathrm{E}(\lambda_{2,t+1}|\boldsymbol{\lambda}_1)}{k_2} \\
&= 1 - (k_1^{-1}, k_2^{-1})(\mathrm{E}(\lambda_{1,t+1}|\boldsymbol{\lambda}_1), \mathrm{E}(\lambda_{2,t+1}|\boldsymbol{\lambda}_1))^\top \\
&= 1 - \nu^\top \mathrm{E}(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_1) \geq 1 - \nu^\top \left[(\boldsymbol{I} - (\boldsymbol{A} + \boldsymbol{B}))^{-1}\boldsymbol{\delta} + (\boldsymbol{A} + \boldsymbol{B})^t \boldsymbol{\lambda}_1\right].
\end{aligned}$$

It is easy to see that there exist $k_1 \in \mathbb{R}$ and $k_2 \in \mathbb{R}$ large enough such that $\nu^\top \mathrm{E}(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_1) < \epsilon$. Hence, $\mathbb{P}^t(\boldsymbol{\lambda}_{t+1} \in \mathcal{C}|\boldsymbol{\lambda}_1) \geq 1 - \epsilon$, i.e., $\forall k \geq 1, k^{-1}\sum_{t=1}^k \mathbb{P}^t(\boldsymbol{\lambda}_{t+1} \in \mathcal{C}|\boldsymbol{\lambda}_1) \leq 1 - \epsilon$. Therefore, $\{\boldsymbol{\lambda}_t\}$ is boundedness in probability on average. Thus, there exists at least one stationary distribution of $\{\boldsymbol{\lambda}_t\}$.

In the following, we illustrate the uniqueness of its stationary distribution. By Assumption 1, $\|\boldsymbol{A}\|_p \leq 1, p \in [1, \infty)$. For all $l \geq 1$, $\{\boldsymbol{\lambda}_t\}$ is a reachable. In subsequent, we prove that $\{\boldsymbol{\lambda}_t\}$ is an e-chain, i.e., $\|\mathbb{P}_{x_1}^k f - \mathbb{P}_{z_1}^k f\| < \epsilon$, if $\|x_1 - z_1\| < \eta$, for all $k \geq 1$. Let $|f| \leq 1$.

First, we consider $k = 1$, using Proposition 2 and $\|\boldsymbol{AB}\|_p \leq \|\boldsymbol{A}\|_p \|\boldsymbol{B}\|_p$,

$$\begin{aligned}
&\|\mathbb{P}_{\boldsymbol{x}_1} f - \mathbb{P}_{\boldsymbol{z}_1} f\| \\
&= \|\sum_m \sum_n \left(f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top)\mathbb{P}(m,n|\boldsymbol{x}_1) - f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)\mathbb{P}(m,n|\boldsymbol{z}_1)\right)\|
\end{aligned}$$

$$\leq \|f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top)\| \| \sum_m \sum_n \|\mathbb{P}(m,n|\boldsymbol{x}_1) - \mathbb{P}(m,n|\boldsymbol{z}_1)\|$$

$$+ \| \sum_m \sum_n \mathbb{P}(m,n|\boldsymbol{z}_1)\| \|f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top) - f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)\|$$

$$\leq \sum_m \sum_n \|\mathbb{P}(m,n|\boldsymbol{x}_1) - \mathbb{P}(m,n|\boldsymbol{z}_1)\| + \|f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^\top) - f(\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^\top)\|$$

$$\leq 1 - \exp(-\zeta\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|) + \|\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{z}_1)\|_p$$

$$\leq 1 - \exp(-\zeta\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|) + \|\boldsymbol{A}\|_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p$$

$$\leq 1 - \exp(-\zeta\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|) + \epsilon^{'}.$$

Second, we consider the case of $k = 2$, $\boldsymbol{x}_2 = \boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{B}(m,n)^{'}$, $\boldsymbol{z}_2 = \boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{z}_1 + \boldsymbol{B}(m,n)^{'}$, $\|\boldsymbol{x}_2 - \boldsymbol{z}_2\| \leq \|\boldsymbol{A}\|_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p$, then

$$\|\mathbb{P}^2_{\boldsymbol{x}_1} f - \mathbb{P}^2_{\boldsymbol{z}_1} f\| = \| \sum_m \sum_n \left(\mathbb{P}(m,n|\boldsymbol{x}_1)\mathbb{P}_{\boldsymbol{x}_2} f - \mathbb{P}(m,n|\boldsymbol{z}_1)\mathbb{P}_{\boldsymbol{z}_2} f\right) \|$$

$$\leq \| \sum_m \sum_n \mathbb{P}(m,n|\boldsymbol{x}_1) \left(\mathbb{P}_{\boldsymbol{x}_2} f - \mathbb{P}_{\boldsymbol{z}_2} f\right) \| + \| \sum_m \sum_n \left(\mathbb{P}(m,n|\boldsymbol{x}_1) - \mathbb{P}(m,n|\boldsymbol{z}_1)\right) \mathbb{P}_{\boldsymbol{z}_2} f\|$$

$$\leq \|\mathbb{P}_{\boldsymbol{x}_2} f - \mathbb{P}_{\boldsymbol{z}_2} f\| + \| \sum_m \sum_n \left(\mathbb{P}(m,n|\boldsymbol{x}_1) - \mathbb{P}(m,n|\boldsymbol{z}_1)\right) \|$$

$$\leq \epsilon^{'} + [1 - \exp(-\zeta\|\boldsymbol{x}_2 - \boldsymbol{z}_2\|)] + [1 - \exp(-\zeta\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|)]$$

$$\leq \epsilon^{'} + [1 - \exp(-\zeta\|A\|_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p)] + [1 - \exp(-\zeta\|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p)].$$

Similarly, for all $k \geq 1$,

$$\|\boldsymbol{x}_k - \boldsymbol{z}_k\| = \|\boldsymbol{A}(\boldsymbol{x}_{k-1} - \boldsymbol{z}_{k-1})\| = \cdots = \|\boldsymbol{A}^s(\boldsymbol{x}_{k-s} - \boldsymbol{z}_{k-s})\|$$

$$\leq \|\boldsymbol{A}^s\|_p \|\boldsymbol{x}_{k-s} - \boldsymbol{z}_{k-s}\|, s = 1, 2, \cdots, k.$$

Using $\exp(-x) \geq 1 - x, x < 0$,

$$\begin{aligned}\|\mathbb{P}^k_{\boldsymbol{x}_1} f - \mathbb{P}^k_{\boldsymbol{z}_1} f\| &\leq \epsilon^{'} + \sum\nolimits_{s=0}^{k-1} [1 - \exp(-\zeta \|\boldsymbol{A}^s\|_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p)] \\ &\leq \epsilon^{'} + \sum\nolimits_{s=0}^{k-1} \zeta \|\boldsymbol{A}^s\|_p \|\boldsymbol{x}_1 - \boldsymbol{z}_1\|_p = \epsilon^{'} + \eta\zeta \|\boldsymbol{A}\|_p^{-1}.\end{aligned}$$

Therefore, $\{\boldsymbol{\lambda}_t\}$ is an e-chain. Thus, its stationary distribution is unique.

(3) Denote $\boldsymbol{\lambda} = (\mu, \sigma^2)^\top$, $F^{-1}_{\boldsymbol{\lambda}}(\boldsymbol{s}) = (F^{-1}_{\mu}(s_1), F^{-1}_{\mu^2/n+\sigma^2}(s_2))$ with $F^{-1}_x(s) = \inf\{t \geq 0, F_x(t) \geq s\}$, $\boldsymbol{s} = (s_1, s_2) \in (0,1)^2$. By Proposition 2,

$$\begin{aligned}\int \|F_{\boldsymbol{\lambda}}(\boldsymbol{s}) - F_{\boldsymbol{\lambda}_0}(\boldsymbol{s})\|_1 d\boldsymbol{s} &= \|(\mu, \mu^2/n + \sigma^2)^\top - (\mu_0, \mu_0^2/n + \sigma_0^2)^\top\|_1 \\ &= \|\boldsymbol{D}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)\|_1 \leq \|\boldsymbol{D}\|_1 \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_1 \leq 3\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_1,\end{aligned}$$

where

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 \\ (\mu + \mu_0)/n & 1 \end{pmatrix}$$

and $\|\boldsymbol{D}\|_1 = 1 + |\mu + \mu_0|/n \leq 3$, $\forall (\mu, \mu_0) \in [-n, n]^2$.

Hence,

$$\begin{aligned}&\mathrm{E}\|\boldsymbol{\lambda}_1(\boldsymbol{\lambda}) - \boldsymbol{\lambda}_1(\boldsymbol{\lambda}_0)\|_p = \int \|\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{\lambda} + \boldsymbol{B}F^{-1}_{\boldsymbol{\lambda}}(\boldsymbol{s}) - (\boldsymbol{\delta}_1 + \boldsymbol{A}\boldsymbol{\lambda}_0 + \boldsymbol{B}F^{-1}_{\boldsymbol{\lambda}_0}(\boldsymbol{s}))\|_p d\boldsymbol{s} \\ &= \int \|\boldsymbol{A}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) + \boldsymbol{B}(F^{-1}_{\boldsymbol{\lambda}}(\boldsymbol{s}) - F^{-1}_{\boldsymbol{\lambda}_0}(\boldsymbol{s}))\|_p d\boldsymbol{s} \\ &\leq \|\boldsymbol{A}\|_p \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_p + \|\boldsymbol{B}\|_p \int \|F^{-1}_{\boldsymbol{\lambda}}(\boldsymbol{s}) - F^{-1}_{\boldsymbol{\lambda}_0}(\boldsymbol{s})\|_p d\boldsymbol{s} \\ &\leq \|\boldsymbol{A}\|_p \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_p + \|\boldsymbol{B}\|_p \int \|F^{-1}_{\boldsymbol{\lambda}}(\boldsymbol{s}) - F^{-1}_{\boldsymbol{\lambda}_0}(\boldsymbol{s})\|_1 d\boldsymbol{s}\end{aligned}$$

$$\leq \|\boldsymbol{A}\|_p \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_p + (1 + |\mu + \mu_0|/n)\|\boldsymbol{B}\|_p \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_1$$

$$= \big(\max(|\beta_1|, \psi_1) + 3 \times 2^{1-1/p}(1 + |\mu + \mu_0|/n) \max(|\alpha_1|, \phi_1) \big) \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_p$$

By Assumption 1, $\max(|\beta_1|, \psi_1) + 3 \times 2^{1-1/p}(1 + |\mu + \mu_0|/n) \max(|\alpha_1|, \phi_1) < 1$. Therefore, $\{\boldsymbol{\lambda}_t\}$ is geometric moment contraction process by Proposition 4.2.1 in Liu (2012). Therefore, there exists a unique stationary distribution. According to Theorem 6.4.1 in Doukhan and Wintenberger (2008), it is also τ-weakly dependent process and constitutes an ergodic causal Bernoulli shift, thereby completing this proof.

Remark 2

Denote $\boldsymbol{Z_t} = (X_t, \sigma_t^2)$. It is easy to see $\{\boldsymbol{Z_t}\}$ is a time-homogeneous Markov chain. In the following, we first denote $Y_t = |X_t|$, then $\{\sigma_t^2\}$ is rewritten as $\sigma_t^2 := g(Y_{t-1}, \sigma_{t-1}^2) = \phi_0 + \phi_1 Y_{t-1} + \psi_1 \sigma_{t-1}^2$. Thus, for all $y_1 < y_2, \sigma_1^2 < \sigma_2^2$,

$$|g(y_1, \sigma_1^2) - g(y_2, \sigma_2^2)| \leq \phi_1 |y_1 - y_2| + \psi_1 |\sigma_1^2 - \sigma_2^2|. \tag{S2.4}$$

Hence, (A1) in Doukhan et al. (2021) holds by (S2.4), if $\phi_0 > 0$, $\phi_1 \in (0, 1), \psi \in (0, 1)$ and $0 < \phi_1 + \psi_1 < 1$.

Second, if $X \sim \text{TD2}(n, 0, \sigma^2/n)$, then $\text{E}(X^2) = \sigma^2$. Subsequently, we denote $Q_b(x) = \mathbb{P}(X \leq x)$ and $g(b) = b$, where $Q_b(x)$ is the cumulative distribution of the $\text{TD2}(n, 0, \sigma^2/n)$ with $\int x^2 dQ_b(x) = b$. Hence,

$g(b)$ maps from $(0,n)$ to $(0,n)$ and is a strictly monotonic function with $\int x^2 d\mathrm{TD2}(n,0,g(b)/n) = b$. Assume $U_1, U_2, \cdots, U_n$ be independent and uniformly distributed on [0,n]. Let $b, b^{'} \in (0,n)$, then $X := \sum_{i=1}^{n} I(U_i \leq g(b)) \sim Q_b$ and $X^{'} = \sum_{i=1}^{n} I(U_i \leq g(b^{'})) \sim Q_{b'}$. Thus, $X \leq X^{'}$ with probability one, if $b \leq b^{'}$, i.e., X is stochastically less than or equal to $X^{'}$. Therefore, (A2) in Doukhan et al. (2021) holds.

To sum up, Theorem 2.4 and Theorem 2.6 in Doukhan et al. (2021) hold. Hence, the first two items of Theorem 1 hold for the TDGV(1,1) model. By Doukhan (1994), β-mixing implies strong mixing, so $\{X_t\}$ is stationary and strongly mixing at a geometric rate, i.e., $\{X_t\}$ is ergodic. Similar to Proposition 2 in Davis and Liu (2016), $\{\sigma_t^2\}$ is also ergodic. Thus, the third item of Theorem 1 holds for the TDGV(1,1) model. The proof is closed.

Proposition 3

By Definition 2, $\mathrm{E}(X_t|\mathcal{F}_{t-1}) = \mu_t, \mathrm{E}(X_t) = \mathrm{E}(\mu_t) := \mu_X = \alpha_0/(1-\alpha_1-\beta_1)$, $\mathrm{Var}(X_t|\mathcal{F}_{t-1}) = \sigma_t^2$ and

$$\mathrm{E}(\sigma_t^2) = \phi_0 + \phi_1 \mathrm{E}|X_t| + \psi_1 \mathrm{E}(\sigma_t^2) = (\phi_0 + \phi_1 \mathrm{E}|X_t|)/(1-\psi_1). \tag{S2.5}$$

By Proposition 1, $|X_t||\mathcal{F}_{t-1} \sim \mathrm{Bin}(n, p_t^2 + \nu_t)$, thus,

$$\mathrm{E}(|X_t|) = \mathrm{E}(\mathrm{E}(|X_t||\mathcal{F}_{t-1})) = \mathrm{E}\mu_t^2/n + \mathrm{E}\sigma_t^2. \tag{S2.6}$$

Substituting (S2.6) into (S2.5) yields $n(1-\psi_1)\mathrm{E}\sigma_t^2 = n\phi_0 + \phi_1\mathrm{E}(\mu_t^2) + n\phi_1\mathrm{E}\sigma_t^2$. Thus,

$$\mathrm{E}\sigma_t^2 = \frac{\phi_0}{1-\phi_1-\psi_1} + \frac{\phi_1}{n(1-\phi_1-\psi_1)}\mathrm{E}(\mu_t^2). \tag{S2.7}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(X_t) &= \mathrm{E}\big(\mathrm{Var}(X_t|\mathcal{F}_t)\big) + \mathrm{Var}(\mathrm{E}(X_t|\mathcal{F}_t)) = \mathrm{E}(\sigma_t^2) + \mathrm{Var}(\mu_t) \\
&\overset{\text{(S2.7)}}{=} \frac{\phi_0}{1-\phi_1-\psi_1} + \frac{\phi_1}{n(1-\phi_1-\psi_1)}\big(\mathrm{Var}(\mu_t) + \mu_X^2\big) + \mathrm{Var}(\mu_t), \\
n(1-\phi_1-\psi_1)\mathrm{Var}(X_t) &= n\phi_0 + \phi_1\mu_X^2 + \big(\phi_1 + n(1-\phi_1-\psi_1)\big)\mathrm{Var}(\mu_t), \\
n(1-\phi_1-\psi_1)(1-\beta_1^2)\mathrm{Var}(X_t) &= (n\phi_0 + \phi_1\mu_X^2)(1-\beta_1^2) + \alpha_1^2\big(\phi_1 + n(1-\phi_1-\psi_1)\big)\mathrm{Var}(X_t),
\end{aligned}$$

hence,

$$\mathrm{Var}(X_t) = \frac{n(1-\beta_1^2)\phi_0 + \phi_1(1-\beta_1^2)\mu_X^2}{n(1-\phi_1-\psi_1)(1-\alpha_1^2-\beta_1^2) - \phi_1\alpha_1^2}.$$

The proof is closed.

Lemma 1

By Definition 2, if there exists a $t \geq 1$ such that $X_t(\boldsymbol{\theta}) = X_t(\tilde{\boldsymbol{\theta}})$, $\mathbb{P}_n(x|\mathcal{F}_{t-1})_{\boldsymbol{\theta}}$ a.s., then $\mathrm{E}_{\boldsymbol{\theta}}(X_t|\mathcal{F}_{t-1}) = \mathrm{E}_{\tilde{\boldsymbol{\theta}}}(X_t|\mathcal{F}_{t-1})$ and $\mathrm{Var}_{\boldsymbol{\theta}}(X_t|\mathcal{F}_{t-1}) = \mathrm{Var}_{\tilde{\boldsymbol{\theta}}}(X_t|\mathcal{F}_{t-1})$, i.e.,

$$\begin{aligned}
\alpha_0 + \alpha_1 X_{t-1} + \beta_1\mu_{t-1} &= \tilde{\alpha}_0 + \tilde{\alpha}_1 X_{t-1} + \tilde{\beta}_1\mu_{t-1}, \\
\phi_0 + \phi_1|X_{t-1}| + \psi_1\sigma_{t-1}^2 &= \tilde{\phi}_0 + \tilde{\phi}_1|X_{t-1}| + \tilde{\psi}_1\sigma_{t-1}^2,
\end{aligned}$$

if and only if $\alpha_0 = \tilde{\alpha}_0, \alpha_1 = \tilde{\alpha}_1, \beta_1 = \tilde{\beta}_1, \phi_0 = \tilde{\phi}_0, \phi_1 = \tilde{\phi}_1, \psi_1 = \tilde{\psi}_1$. Hence, the TDGMV((1,1),(1,1)) model is identifiable.

Theorem 2

By Assumption 1, $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0$ is an interior points in $\boldsymbol{\Theta}$. By supplementary material S1, $l_t(\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ for each X_t in an open and convex neighbourhood $N(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$. Thus, $l_t(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ for each X_t. By Lemma 3, $\mathrm{E}\left(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} |l_t(\boldsymbol{\theta})|\right) < \infty$ and $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\sum_{t=2}^{T} l_t(\boldsymbol{\theta})/T - \mathrm{E}l_t(\boldsymbol{\theta})\right| = o_p(1)$, and $\mathrm{E}l_t(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0$. Consequently, we have established all the conditions for consistency of Theorem 4.1.2 in Amemiya (1985), thus the first item of Theorem 2 holds.

The proof of the second item rests on the Taylor series expansion of the score vector around $\boldsymbol{\theta}_0$, i.e.,

$$\begin{aligned}\mathbf{0} &= T^{-1/2}\partial\ell(\hat{\boldsymbol{\theta}}_T)/\partial\boldsymbol{\theta} \\ &= T^{-1/2}\partial\ell(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} + \left(T^{-1}\partial^2\ell(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\right)\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0), \qquad \text{(S2.8)}\end{aligned}$$

where $\boldsymbol{\theta}^\star$ lies in between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$. According to Theorem 4.1.3 of Amemiya (1985), we need to show that

$$T^{-1/2}\frac{\partial\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}(\boldsymbol{\theta}_0)) \text{ and } \frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \xrightarrow{p} -\boldsymbol{M}(\boldsymbol{\theta}_0). \qquad \text{(S2.9)}$$

The former one holds by Lemma 4 and Lemma 5, and $T^{-1}\partial^2\ell(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top \xrightarrow{p} -\boldsymbol{M}(\boldsymbol{\theta}_0)$ by Lemma 6.

By supplementary materials S1, it is easy to see $\forall(i,j,k) \in \{1,2,\cdots,6\}^3$, $\partial^3\ell_t(\boldsymbol{\theta})/\partial\theta_i\theta_j\theta_k$ exists and is continuous with respect to $\boldsymbol{\theta}$. By Lemma 1 given in supplementary materials S4 and (S1.1), there exists some constant c such that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\partial^3\ell_t(\boldsymbol{\theta})/\partial\theta_i\theta_j\theta_k| < c$.

Let $h(\tau) = \boldsymbol{\nu}_1^\top\partial^2 l_t(\boldsymbol{\theta}_0 + \tau(\boldsymbol{\theta}-\boldsymbol{\theta}_0))/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\boldsymbol{\nu}_2$, where $\boldsymbol{\nu}_i \in \mathbb{R}^6, i = 1,2$. By using mean value theorem and Lemma 1 in Jensen and Rahbek (2004),

$$\begin{aligned} |h(1)-h(0)| &= \left|\nu_1^\top\left(\partial^2 l_t(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top - \partial^2 l_t(\boldsymbol{\theta}_0))/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\right)\nu_2\right| \\ &\le \|\nu_1\|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\|\nu_1\|c \to 0, \quad as\ \boldsymbol{\theta}\to\boldsymbol{\theta}_0. \end{aligned} \tag{S2.10}$$

Thus, by using Lemma 6 and (S2.10)

$$\begin{aligned} \|\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - -\boldsymbol{M}(\boldsymbol{\theta}_0)\| &\le \|\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\| + \|\frac{1}{T}\frac{\partial^2\ell(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \boldsymbol{M}(\boldsymbol{\theta}_0)\| \\ &\to 0. \end{aligned}$$

Therefore, $T^{-1}\partial^2\ell(\boldsymbol{\theta}^\star)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top \xrightarrow{p} -\boldsymbol{M}(\boldsymbol{\theta}_0)$. Hence, by using (S2.8), (S2.9) and Theorem 4.1.3 in Amemiya (1985), it is direct to obtain that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{M}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{M}^{-1}(\boldsymbol{\theta}_0))$.

S3 Simulation Results

This section starts with the simulation results for the specified TDGV(1,1) and TDGMV ((1,1),(1,1)). Next, we examine the model performance under misspecified settings. Finally, we discuss the performance of the forecasting.

S3.1 Scenario I: Specified TDGV and TDGMV Models

Without loss of generality, we perform the analysis with $n = 10$ and sample size $T = 300$ and 1000. Within each scenario, we simulate 1000 replications and employ the `constrOptim` function from the stats package in `R` to optimize the function given by equation (3.12), yielding $\hat{\boldsymbol{\theta}}$. Performance metrics, including the empirical bias, ESD and ASD, are summarized in Tables T1–T5.

In addition, to provide additional graphical information, we present their boxplots of (B1) and (C1) in Figures F1 and F2, and their quantile-quantile plots (qqplots) with $T = 1000$ in Figures F3 and F4. Other graphical results are similar, and we omit the details for brevity.

Table T1: Simulation results of (B1)-(B4) for the specified TDGV model.

	ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000
			(B1) $= (0.5, 0.2, 0.4)$			
Bias	0.0756	0.0048	−0.0706	0.0315	0.0013	−0.0287
ESD	0.2651	0.1194	0.2582	0.2253	0.0678	0.2198
ASD	0.6462	0.1284	0.6318	0.3083	0.0734	0.3052
			(B2) $= (0.5, 0.2, 0.5)$			
Bias	0.0632	0.0478	0.0094	0.0051	−0.0532	−0.0389
ESD	0.3167	0.2694	0.1377	0.0781	0.2751	0.2257
ASD	0.6836	0.3250	0.1373	0.0776	0.5500	0.2640
			(B3) $= (0.5, 0.3, 0.4)$			
Bias	0.0619	0.0402	0.0060	0.0002	−0.0526	−0.0333
ESD	0.2761	0.1971	0.1430	0.0820	0.2561	0.1796
ASD	0.4629	0.2139	0.1456	0.0793	0.4156	0.1923
			(B4) $= (0.5, 0.3, 0.5)$			
Bias	0.0786	0.0343	0.0094	−0.0003	−0.0561	−0.0233
ESD	0.3001	0.2277	0.1598	0.0885	0.2337	0.1741
ASD	0.5307	0.2536	0.1588	0.0865	0.3870	0.1866

Table T2: Simulation results of (B5)-(B8) for the specified TDGV model.

	ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000
			(B5) $= (0.5, 0.4, 0.4)$			
Bias	0.0686	0.0215	−0.0047	−0.0003	−0.0454	−0.0146
ESD	0.2811	0.1802	0.1602	0.0891	0.2339	0.1487
ASD	0.3750	0.1771	0.1571	0.0857	0.3027	0.1460
			(B6) $= (0.5, 0.4, 0.3)$			
Bias	0.0216	0.0148	−0.0041	−0.0038	−0.0101	−0.0100
ESD	0.2295	0.1450	0.1517	0.0791	0.2252	0.1447
ASD	0.3223	0.1519	0.1510	0.0799	0.3157	0.1494
			(B7) $= (1, 0.3, 0.4)$			
Bias	−0.2179	−0.1636	−0.0182	−0.0159	0.1060	0.0805
ESD	0.3025	0.2342	0.1834	0.1041	0.1762	0.1270
ASD	1.0201	0.5872	0.2034	0.1233	0.5211	0.2996
			(B8) $= (1, 0.4, 0.3)$			
Bias	−0.1997	−0.1319	−0.0191	−0.0118	0.1029	0.0698
ESD	0.2674	0.1976	0.1891	0.1059	0.1797	0.1220
ASD	0.8561	0.3963	0.2184	0.1137	0.4955	0.2220

Table T3: Simulation results of (C1)-(C3) for the specified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
						(C1) $= (0.5, -0.1, 0.1, 0.5, 0.2, 0.3)$						
Bias	−0.0054	−0.0137	−0.0054	−0.0014	0.0203	0.0248	0.0027	0.0162	0.0024	−0.0006	−0.0083	−0.0195
ESD	0.2353	0.1652	0.0648	0.0351	0.4538	0.3247	0.2434	0.1715	0.1047	0.0571	0.2721	0.1988
ASD	0.2234	0.1365	0.0576	0.0329	0.4219	0.2633	0.4123	0.1916	0.1012	0.0553	0.4778	0.2205
						(C2) $= (-0.5, 0.1, 0.2, 0.5, 0.2, 0.3)$						
Bias	−0.0469	−0.0300	−0.0034	−0.0007	−0.0631	−0.0414	−0.0155	0.0119	0.0159	0.0005	−0.0010	−0.0116
ESD	0.2845	0.2099	0.0658	0.0344	0.4212	0.3005	0.2504	0.1655	0.1179	0.0498	0.2712	0.1865
ASD	0.3119	0.1871	0.0575	0.0327	0.4385	0.2660	0.3183	0.1628	0.0948	0.0477	0.3611	0.1829
						(C3) $= (-0.5, -0.1, 0.1, 0.5, 0.2, 0.3)$						
Bias	0.0309	0.0072	−0.0050	−0.0002	0.0637	0.0150	0.0130	0.0155	0.0001	−0.0004	−0.0212	−0.0165
ESD	0.2421	0.1664	0.0652	0.0335	0.4761	0.3231	0.2319	0.1761	0.1044	0.0570	0.2628	0.2048
ASD	0.1985	0.1440	0.0565	0.0333	0.3719	0.2773	0.4102	0.2002	0.1033	0.0579	0.4756	0.2318

Table T4: Simulation results of (C4)-(C6) for the specified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
(C4) $= (-0.5, -0.1, 0.2, 0.5, 0.2, 0.3)$												
Bias	0.0130	0.0006	−0.0155	−0.0019	0.0413	0.0043	−0.0089	0.0072	0.0101	−0.0028	−0.0008	−0.0062
ESD	0.2447	0.1757	0.0629	0.0337	0.4248	0.3077	0.2487	0.1678	0.1105	0.0546	0.2687	0.1942
ASD	0.2063	0.1497	0.0571	0.0326	0.3483	0.2587	0.3720	0.1773	0.1009	0.0522	0.4300	0.2032
(C5) $= (-0.5, 0.1, 0.1, 0.5, 0.2, 0.3)$												
Bias	−0.0380	−0.0256	0.0001	−0.0021	−0.0636	−0.0399	−0.0095	0.0101	−0.0018	−0.0017	0.0052	−0.0105
ESD	0.2676	0.2030	0.0664	0.0350	0.4400	0.3316	0.2365	0.1695	0.0986	0.0527	0.2666	0.1942
ASD	0.2649	0.1720	0.0583	0.0331	0.4146	0.2756	0.3671	0.1813	0.0954	0.0511	0.4246	0.2066
(C6) $= (-0.5, -0.1, -0.1, 0.5, 0.2, 0.3)$												
Bias	0.0178	0.0128	−0.0058	−0.0016	0.0502	0.0312	0.0026	0.0141	−0.0046	−0.0005	−0.0036	−0.0175
ESD	0.2044	0.1390	0.0637	0.0337	0.4648	0.3231	0.2327	0.1758	0.1023	0.0616	0.2696	0.2080
ASD	0.1791	0.1182	0.0570	0.0334	0.3880	0.2706	0.3767	0.2124	0.1037	0.0590	0.4451	0.2462

Table T5: Simulation results of (C7)-(C9) for the specified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
						$(C7) = (0.5, -0.1, -0.2, 0.5, 0.2, 0.3)$						
Bias	−0.0376	−0.0142	−0.0065	−0.0018	0.0940	0.0384	0.0245	0.0087	−0.0039	−0.0015	−0.0269	−0.0101
ESD	0.1859	0.1245	0.0631	0.0337	0.4590	0.3130	0.2290	0.1733	0.1131	0.0604	0.2662	0.2030
ASD	0.1695	0.1106	0.0570	0.0332	0.3979	0.2713	0.4793	0.2109	0.1083	0.0602	0.5608	0.2443
						$(C8) = (0.5, -0.1, -0.1, 0.5, 0.2, 0.3)$						
Bias	−0.0261	−0.0177	−0.0056	0.0011	0.0651	0.0368	0.0073	0.0210	−0.0071	0.0008	−0.0025	−0.0243
ESD	0.2005	0.1457	0.0642	0.0335	0.4693	0.3407	0.2417	0.1737	0.1070	0.0588	0.2777	0.2000
ASD	0.1771	0.1218	0.0569	0.0333	0.3864	0.2796	0.4088	0.2206	0.1058	0.0597	0.4794	0.2538
						$(C9) = (-0.5, -0.1, -0.2, 0.5, 0.2, 0.3)$						
Bias	0.0265	0.0152	−0.0042	−0.0006	0.0903	0.0453	0.0097	0.0187	−0.0078	−0.0014	−0.0089	−0.0186
ESD	0.1907	0.1242	0.0616	0.0336	0.4579	0.3136	0.2359	0.1783	0.1020	0.0597	0.2720	0.2046
ASD	0.1796	0.1086	0.0569	0.0334	0.4136	0.2648	0.4420	0.2331	0.1084	0.0601	0.5217	0.2673

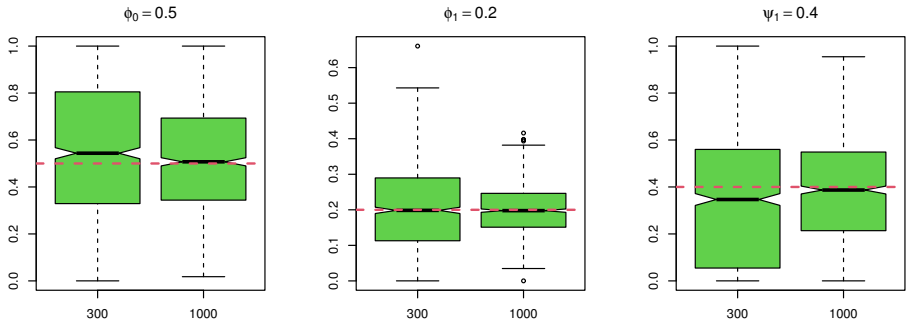


Figure F1: Boxplots of (B1) for the specified TDGV model.

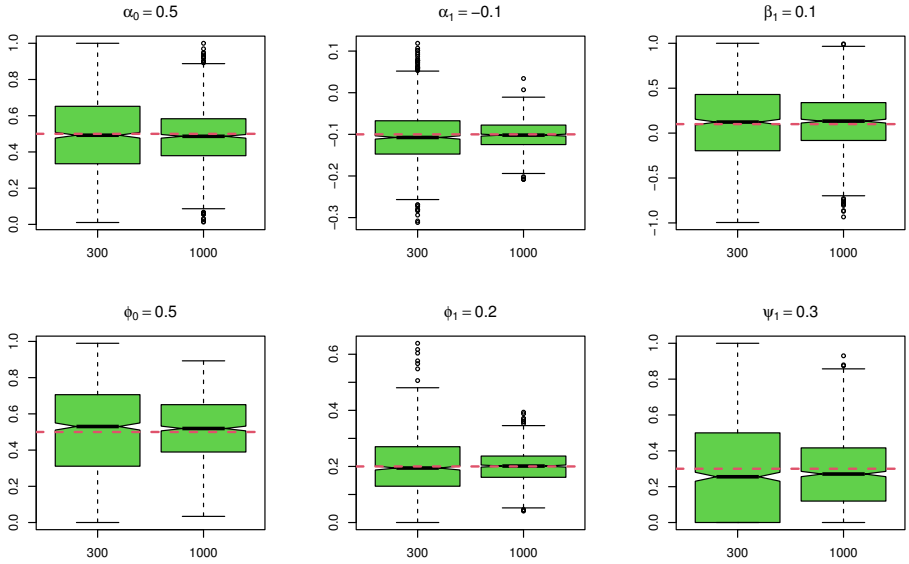


Figure F2: Boxplots of (C1) for the specifited TDGMV model.

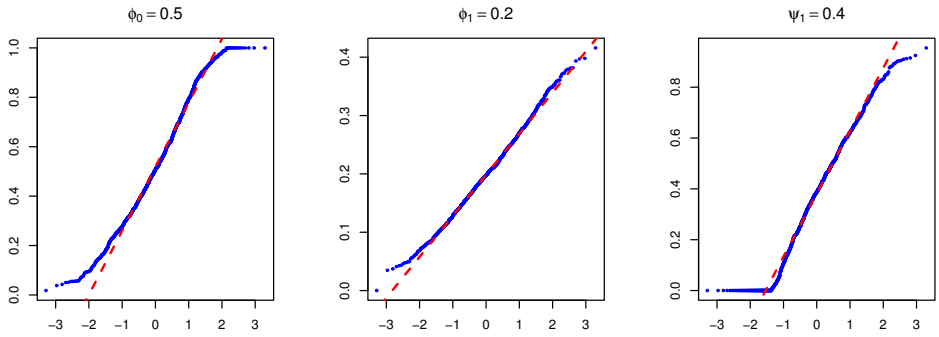


Figure F3: QQ-plots of (B1) for the specified TDGV model.

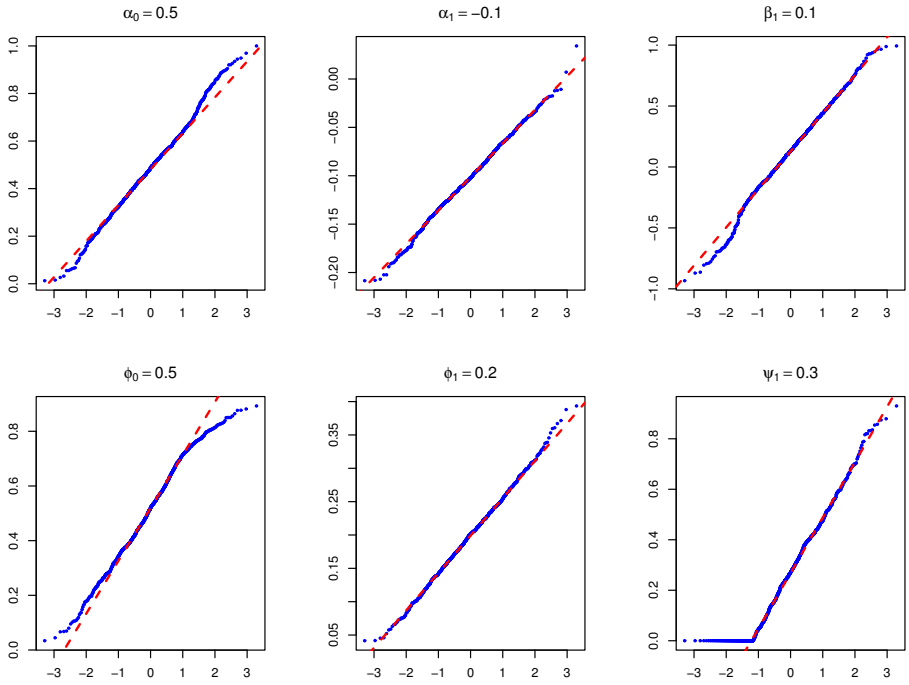


Figure F4: QQ-plots of (C1) for the specifited TDGMV model.

S3.2 Scenario II: Misspecified TDGV and TDGMV Models

In this part, we first generate samples with sizes 300 and 1000 from the truncated Skellam GARCH model ($n = 10$) using the same parameter combinations and replication times. We then compute $\hat{\boldsymbol{\theta}}$ by employing the `constrOptim` function to optimize the function in (3.12). Finally, we summarize the empirical bias, ESD and ASD, are summarized in Tables T6-T10, and some additional graphical information in Figures F5-F8.

Remark 1. Similar to the CML estimator of the classic INGARCH model (Fokianos et al., 2009), the numerical results, especially for α_0, are sensitive to the choice of initial value and sample size. This sensitivity leads to violations from normality when the sample size is small.

S3.3 Forecasting Performance Evaluation

To address the forecasting performance under distributional misspecification, we evaluate the calibration of the one-step-ahead predictive distributions using the probability integral transform (PIT). Here, we adopt the randomized PIT proposed by Czado et al. (2009), which is defined as

$$u_t = F(y_t - 1) + V \cdot [F(y_t) - F(y_t - 1)], \quad V \sim U(0, 1),$$

Table T6: Simulation results of (B1)-(B4) for the misspecified TDGV model.

	ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000
			(B1) $= (0.5, 0.2, 0.4)$			
Bias	0.0445	0.0236	−0.0004	−0.0046	−0.0529	−0.0330
ESD	0.2796	0.2303	0.1216	0.0707	0.2831	0.2349
ASD	0.5705	0.2700	0.1225	0.0702	0.5769	0.2748
			(B2) $= (0.5, 0.2, 0.5)$			
Bias	0.0626	0.0328	0.0071	0.0039	−0.0630	−0.0386
ESD	0.3185	0.2736	0.1390	0.0797	0.2788	0.2337
ASD	0.6839	0.2938	0.1351	0.0751	0.5661	0.2452
			(B3) $= (0.5, 0.3, 0.4)$			
Bias	0.0493	0.0279	0.0027	−0.0044	−0.0541	−0.0339
ESD	0.2738	0.2044	0.1460	0.0820	0.2633	0.1878
ASD	0.4334	0.2057	0.1395	0.0775	0.4008	0.1902
			(B4) $= (0.5, 0.3, 0.5)$			
Bias	0.0669	0.0259	0.0061	−0.0037	−0.0588	−0.0268
ESD	0.3043	0.2293	0.1598	0.0859	0.2417	0.1770
ASD	0.5247	0.2387	0.1531	0.0840	0.3913	0.1802

Table T7: Simulation results of (B5)-(B8) for the misspecified TDGV model.

	ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000
			(B5) $= (0.5, 0.4, 0.4)$			
Bias	0.0551	0.0087	−0.0067	−0.0049	−0.0468	−0.0153
ESD	0.2764	0.1808	0.1610	0.0946	0.2366	0.1540
ASD	0.3585	0.1676	0.1547	0.0839	0.2966	0.1425
			(B6) $= (0.5, 0.4, 0.3)$			
Bias	0.0153	0.0043	−0.0097	−0.0116	−0.0162	−0.0109
ESD	0.2257	0.1457	0.1559	0.0810	0.2259	0.1498
ASD	0.3176	0.1464	0.1446	0.0784	0.3235	0.1480
			(B7) $= (1, 0.3, 0.4)$			
Bias	−0.2193	−0.1703	−0.0166	−0.0143	0.0992	0.0752
ESD	0.3031	0.2395	0.1918	0.1052	0.1782	0.1322
ASD	1.0527	0.5561	0.2016	0.1179	0.5476	0.2878
			(B8) $= (1, 0.4, 0.3)$			
Bias	−0.2041	−0.1440	−0.0182	−0.0097	0.0969	0.0655
ESD	0.2722	0.1975	0.1959	0.1110	0.1840	0.1248
ASD	0.7563	0.3717	0.2033	0.1113	0.4425	0.2126

Table T8: Simulation results of (C1)-(C3) for the misspecified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
						(C1) $= (0.5, -0.1, 0.1, 0.5, 0.2, 0.3)$						
Bias	−0.0000	0.0010	−0.0077	−0.0051	0.0054	−0.0001	0.0023	−0.0081	0.0014	−0.0022	−0.0387	−0.0209
ESD	0.1363	0.1190	0.0640	0.0332	0.2720	0.2359	0.1563	0.1297	0.0949	0.0518	0.1893	0.1560
ASD	0.2971	0.1771	0.0586	0.0332	0.5766	0.3459	0.4324	0.1993	0.1025	0.0539	0.5276	0.2406
						(C2) $= (-0.5, 0.1, 0.2, 0.5, 0.2, 0.3)$						
Bias	−0.0346	−0.0196	0.0041	0.0020	−0.0571	−0.0317	−0.0087	−0.0170	−0.0083	−0.0126	−0.0345	−0.0166
ESD	0.1438	0.1283	0.0616	0.0318	0.2213	0.1868	0.1537	0.1277	0.0816	0.0440	0.1859	0.1545
ASD	0.4130	0.2385	0.0618	0.0334	0.5886	0.3424	0.3063	0.1621	0.0824	0.0444	0.3719	0.1966
						(C3) $= (-0.5, -0.1, 0.1, 0.5, 0.2, 0.3)$						
Bias	0.0121	−0.0006	−0.0092	−0.0030	0.0245	0.0019	0.0074	−0.0103	−0.0012	−0.0003	−0.0467	−0.0195
ESD	0.1397	0.1219	0.0641	0.0325	0.2744	0.2411	0.1541	0.1294	0.0904	0.0548	0.1875	0.1599
ASD	0.2796	0.1706	0.0588	0.0331	0.5493	0.3312	0.4204	0.1877	0.0996	0.0539	0.5136	0.2273

Table T9: Simulation results of (C4)-(C6) for the misspecified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
						(C4) $= (-0.5, -0.1, 0.2, 0.5, 0.2, 0.3)$						
Bias	−0.0156	−0.0054	−0.0144	−0.0031	−0.0182	−0.0089	−0.0056	−0.0115	0.0041	0.0002	−0.0360	−0.0225
ESD	0.1283	0.1114	0.0612	0.0315	0.2300	0.1975	0.1556	0.1257	0.0881	0.0476	0.1870	0.1519
ASD	0.2816	0.1926	0.0582	0.0330	0.4893	0.3381	0.3433	0.1636	0.0948	0.0503	0.4180	0.1990
						(C5) $= (-0.5, 0.1, 0.1, 0.5, 0.2, 0.3)$						
Bias	−0.0173	−0.0141	0.0046	0.0015	−0.0379	−0.0262	−0.0068	−0.0165	−0.0114	−0.0113	−0.0297	−0.0136
ESD	0.1553	0.1359	0.0658	0.0322	0.2589	0.2243	0.1536	0.1284	0.0863	0.0485	0.1884	0.1567
ASD	0.3634	0.2060	0.0598	0.0330	0.5874	0.3336	0.3756	0.1791	0.0889	0.0485	0.4590	0.2180
						(C6) $= (-0.5, -0.1, -0.1, 0.5, 0.2, 0.3)$						
Bias	0.0259	0.0125	−0.0107	−0.0032	0.0660	0.0331	0.0058	0.0032	−0.0017	−0.0046	−0.0390	−0.0280
ESD	0.1370	0.1106	0.0637	0.0334	0.3039	0.2548	0.1542	0.1342	0.0973	0.0560	0.1903	0.1643
ASD	0.2489	0.1398	0.0595	0.0333	0.5763	0.3244	0.4414	0.2066	0.1063	0.0574	0.5388	0.2495

Table T10: Simulation results of (C7)-(C9) for the misspecified TDGMV model.

	α_0		α_1		β_1		ϕ_0		ϕ_1		ψ_1	
	300	1000	300	1000	300	1000	300	1000	300	1000	300	1000
						(C7) $= (0.5, -0.1, -0.2, 0.5, 0.2, 0.3)$						
Bias	−0.0446	−0.0224	−0.0126	−0.0047	0.1172	0.0604	0.0236	0.0062	−0.0045	−0.0034	−0.0553	−0.0289
ESD	0.1304	0.0969	0.0628	0.0323	0.2974	0.2385	0.1515	0.1383	0.1041	0.0562	0.1874	0.1654
ASD	0.2239	0.1341	0.0597	0.0340	0.5636	0.3363	0.5023	0.2273	0.1078	0.0599	0.6207	0.2738
						(C8) $= (0.5, -0.1, -0.1, 0.5, 0.2, 0.3)$						
Bias	−0.0287	−0.0187	−0.0111	−0.0034	0.0748	0.0448	0.0092	0.0055	−0.0047	−0.0016	−0.0369	−0.0346
ESD	0.1355	0.1085	0.0623	0.0335	0.3065	0.2520	0.1545	0.1376	0.1013	0.0601	0.1882	0.1684
ASD	0.2550	0.1390	0.0597	0.0335	0.5866	0.3240	0.4729	0.2118	0.1080	0.0588	0.5787	0.2581
						(C9) $= (-0.5, -0.1, -0.2, 0.5, 0.2, 0.3)$						
Bias	0.0372	0.0231	−0.0092	−0.0031	0.1180	0.0658	0.0108	0.0010	−0.0063	−0.0060	−0.0394	−0.0237
ESD	0.1340	0.1019	0.0619	0.0331	0.3022	0.2458	0.1549	0.1370	0.0980	0.0582	0.1886	0.1654
ASD	0.2342	0.1396	0.0594	0.0340	0.5786	0.3478	0.4915	0.2284	0.1100	0.0592	0.6019	0.2760

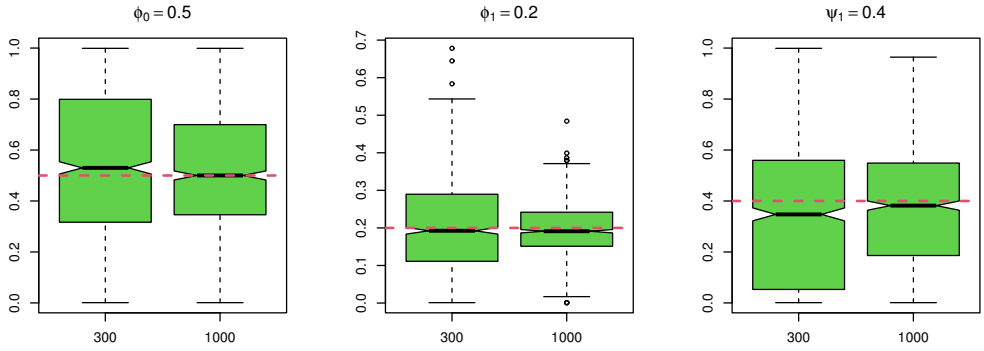


Figure F5: Boxplots of (B1) for the misspecified TDGV model.

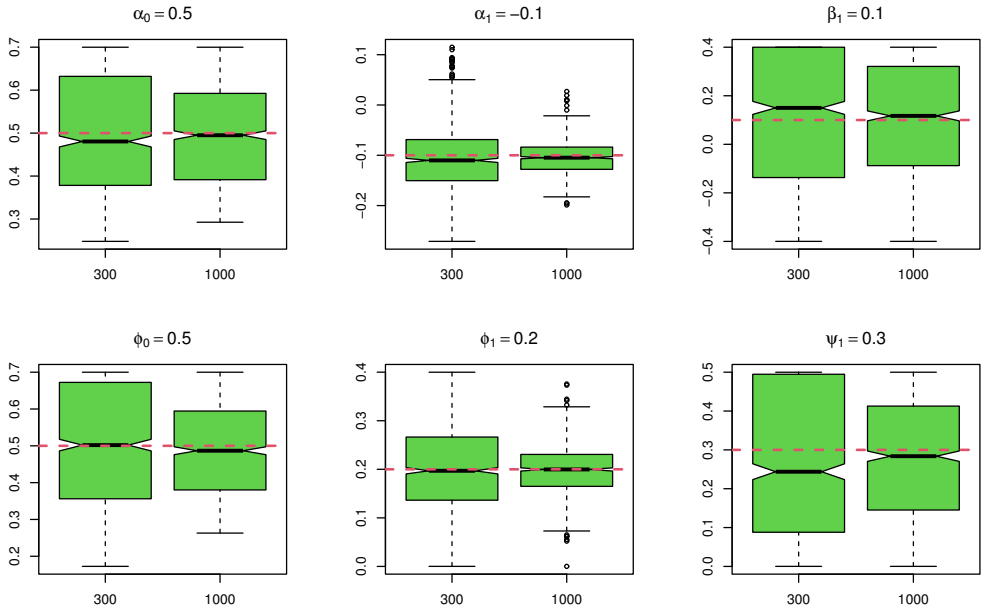


Figure F6: Boxplots of (C1) for the misspecified TDGMV model.

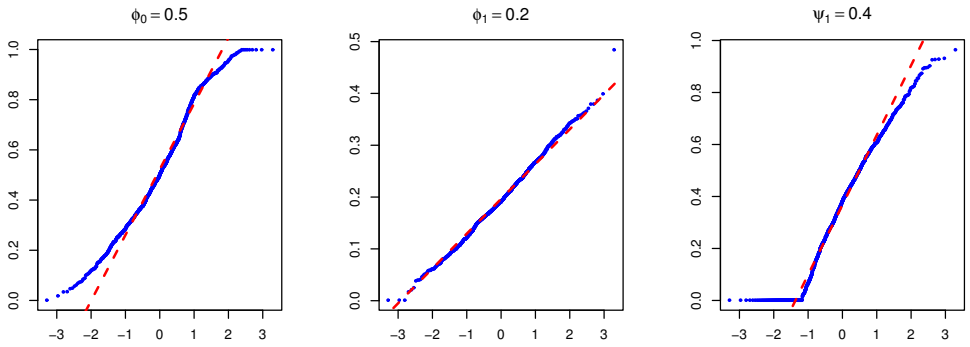


Figure F7: QQ-plots of (B1) for the misspecified TDGV model.

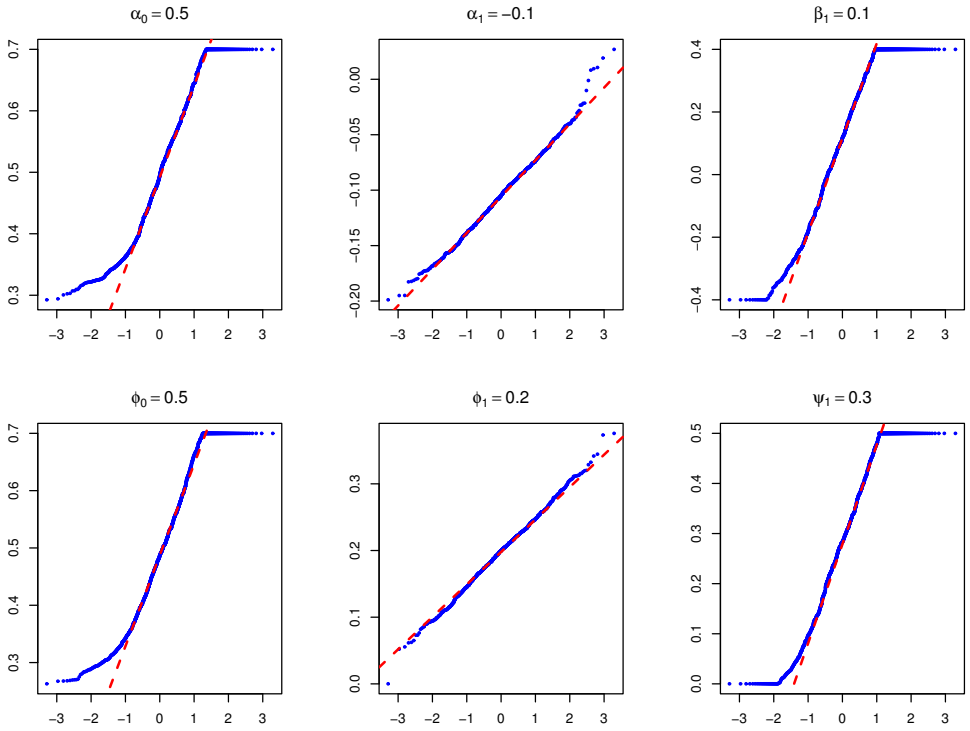


Figure F8: QQ-plots of (C1) for the misspecified TDGMV model.

where $F(\cdot)$ denotes the cumulative distribution function of the one-step-ahead predictive distribution and y_t is the realized observation. If the predictive distribution is correctly specified, The sequence $\{u_t\}$ follows a uniform distribution on $[0, 1]$.

In what follows, we take $\phi_0 = 0.5, \phi_1 = 0.1, \psi_1 = 0.3$ for the TDGV model and $\alpha_0 = -0.5, \alpha_1 = -0.1, \beta_1 = 0.1,\ \phi_0 = 0.5, \phi_1 = 0.2, \psi_1 = 0.3$ for the TDGMV model for example, and then employ PIT to examine whether the predictive cumulative distribution functions are uniformly distributed. Figures F9 and F10 display their PIT histograms for the one-step-ahead predictive distributions under scenarios I and II.

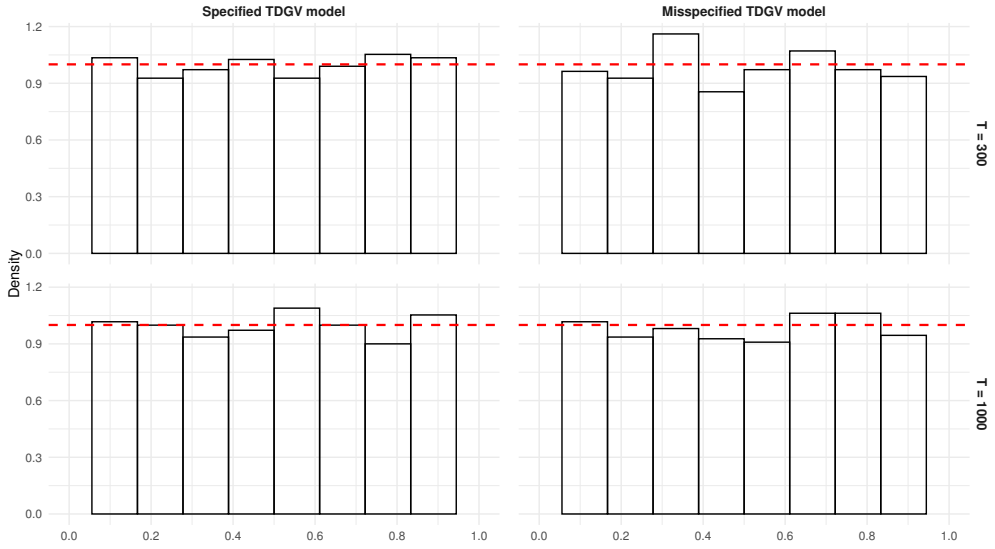


Figure F9: PIT histograms for the TDGV model under scenarios I and II.

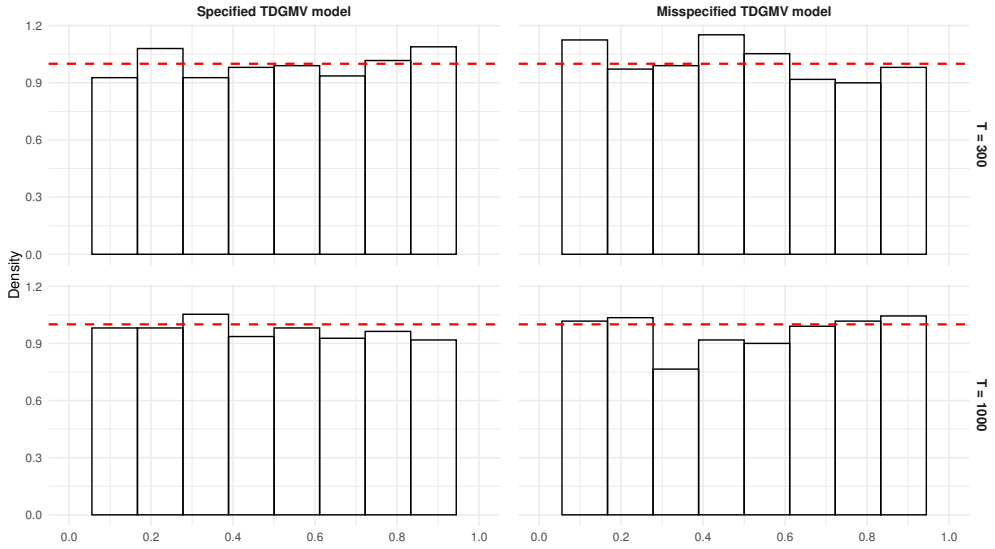


Figure F10: PIT histograms for the TDGMV model under scenarios I and II.

From Figures F9 and F10, in general, all the histograms are approximately uniform, i.e., the proposed models possess a certain degree of robustness to distributional misspecification in terms of predictive calibration.

In addition, we employ the kolmogorov-smirnov (KS) test for uniformity and report the KS test statistic and its corresponding p-value. Furthermore, we also report the sample mean and standard deviation (sd) of the Pearson residuals to present diagnostic information, which are summarized in Table T11.

From Table T11, the KS tests fail to reject the null hypothesis of uniformity in most cases, and the means and standard deviations of the Pearson

Table T11: Results of KS test and the Pearson residuals' mean and sd.

Size	Models	KS statistic	p-value	mean	sd
300	Specified TDGV	0.0225	0.6931	−0.0023	0.9993
	Misspecified TDGV	0.0157	0.9669	−0.0037	1.0110
	Specified TDGMV	0.0193	0.8513	−0.0100	1.0340
	Misspecified TDGMV	0.0212	0.7576	−0.0010	1.0240
1000	Specified TDGV	0.0157	0.9666	0.0007	0.9993
	Misspecified TDGV	0.0267	0.4752	0.0005	1.0110
	Specified TDGMV	0.0240	0.6107	−0.0003	0.9991
	Misspecified TDGMV	0.0379	0.1137	−0.0006	1.0220

residuals are close to 0 and 1, respectively, providing additional evidence that the predictive distributions are well calibrated.

S4 Some Auxiliary Results

In this part, we give some auxiliary results.

S4.1 Auxiliary Lemma

Lemma 1. *If* $Z \sim \text{TD1}(n, r_1, r_2)$ *with its pmf taking form (2.2), then there exist two positive constants* ϖ_1 *and* ϖ_2 *such that* $\varpi_1 \leq \mathbb{P}_n(k) \leq \varpi_2$, $\forall k = 0, \pm 1, \cdots, \pm n$.

Proof. As discussed in Chen et al.(2025a), the pmf of the trinomial difference distribution is unimodal and attains its maximum at $[n(r_1 - r_2)]$ or $[n(r_1 - r_2) + 1]$, where $[x]$ stands for the integer part of x. Hence,

$$\min\{\mathbb{P}_n(n), \mathbb{P}_n(-n)\} \leq \mathbb{P}_n(k) \leq \max\{\mathbb{P}_n([n(r_1 - r_2)]), \mathbb{P}_n([n(r_1 - r_2) + 1])\},$$

where $k = 0, \pm 1, \cdots, \pm n$. Specially, let $\varpi_1 = \min\{\mathbb{P}_n(n), \mathbb{P}_n(-n)\}$ and $\varpi_2 = \max\{\mathbb{P}_n([n(r_1 - r_2)]), \mathbb{P}_n([n(r_1 - r_2) + 1])\}$. □

Lemma 2. *If $Z \sim \mathrm{TD1}(n, r_1, r_2)$ with its pmf taking form (2.2), then*

$$\mathbb{P}_n(k) = r_1\mathbb{P}_{n-1}(k-1) + (1 - r_1 - r_2)\mathbb{P}_{n-1}(k) + r_2\mathbb{P}_{n-1}(k+1),$$

$$\partial\mathbb{P}_n(k)/\partial r_1 = n\left(\mathbb{P}_{n-1}(k-1) - \mathbb{P}_{n-1}(k)\right),$$

$$\partial\mathbb{P}_n(k)/\partial r_2 = n\left(\mathbb{P}_{n-1}(k+1) - \mathbb{P}_{n-1}(k)\right),$$

where $n \in \mathbb{Z}^+$, $0 < r_i < 1$ and $r_1 + r_2 < 1, i = 1, 2$, $\mathbb{P}_n(j) = 0$, if $|j| > n$.

Proof. By Proposition 1, there exist an independent and identically distributed random variable sequence $\{Z_i\}$ such that $Z = \sum_{i=1}^{n} Z_i$. Therefore,

$$\begin{aligned}\mathbb{P}_n(k) = \mathbb{P}_n(Z = k) &= \mathbb{P}_n(\sum\nolimits_{i=1}^{n-1} Z_i + Z_n = k) \\ &= r_1\mathbb{P}_{n-1}(k-1) + (1 - r_1 - r_2)\mathbb{P}_{n-1}(k) + r_2\mathbb{P}_{n-1}(k+1).\end{aligned}$$

Denote $r = r_1 + r_2$. By (2.2), $\partial\mathbb{P}_n(k)/\partial r_1 = \sum \binom{n}{k+j,j}\big((k+j)r_1^{k+j-1}r_2^j(1-r)^{n-k-2j} - (n-k-2j)r_1^{k+j-1}r_2^j(1-r)^{n-k-2j-1}\big) = n\sum \binom{n-1}{k-1+j,j}r_1^{k-1+j}r_2^j(1-r)^{n-k-2j} - n\sum \binom{n-1}{k+j,j}r_1^{k+j}r_2^j(1-r)^{n-1-k-2j} = n\left(\mathbb{P}_{n-1}(k-1) - \mathbb{P}_{n-1}(k)\right)$ and

$\partial \mathbb{P}_n(k)/\partial r_2 = \sum \binom{n}{k+j,j} \big(j r_1^{k+j} r_2^{j-1} (1-r)^{n-k-2j} - (n-k-2j) r_1^{k+j-1} r_2^j (1-r)^{n-k-2j-1} \big) = n \sum \binom{n-1}{k+j,j-1} r_1^{k+j} r_2^{j-1} (1-r)^{n-1-k-2j} - n \sum_j \binom{n-1}{k+j,j} r_1^{k+j} r_2^j (1-r)^{n-1-k-2j} = n \left(\mathbb{P}_{n-1}(k+1) - \mathbb{P}_{n-1}(k) \right).$ □

Lemma 3. *If Assumption 1 holds, then*

$$\mathrm{E}\left(\sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} |l_t(\boldsymbol{\theta})| \right) < \infty \text{ and } \sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} \left| \sum\nolimits_{t=2}^{T} l_t(\boldsymbol{\theta})/T - \mathrm{E} l_t(\boldsymbol{\theta}) \right| = o_p(1),$$

where $o_p(1)$ converges to 0 as $T \to +\infty$. Furthermore, $\mathrm{E} l_t(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0$.

Proof. By Lemma 1, $\log \varpi_1 \leq \log \mathbb{P}_n(x_t(\boldsymbol{\theta})) \leq \varpi_2, \forall \boldsymbol{\theta} \in \Theta$. Hence, there exist some positive constant $M = \max\{|\log \varpi_1|, |\log \varpi_2|\}$ such that

$$\mathrm{E}\left(\sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} |l_t(\boldsymbol{\theta})| \right) = \mathrm{E}\left(\sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} |\log \mathbb{P}_n(x_t(\boldsymbol{\theta}))| \right) < M.$$

By using Theorem 3.1 in Ling and McAleer (2003), $\frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) \to \mathrm{E} l_t(\boldsymbol{\theta})$ in probability as $T \to \infty$. By Jensen's inequality, we have

$$\begin{aligned} \mathrm{E}(l_t(\boldsymbol{\theta})) - \mathrm{E}(l_t(\boldsymbol{\theta}_0)) &= \mathrm{E} \log \frac{\mathbb{P}_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}}}{\mathbb{P}_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}} \\ &\leq \log \mathrm{E}\left(\frac{\mathbb{P}_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}}}{\mathbb{P}_n(X_t|\mathcal{F}_{t-1})_{\boldsymbol{\theta}_0}} \right) = 0. \end{aligned}$$

Thus, $\mathrm{E} l_t(\boldsymbol{\theta})$ attains a strict local maximum at $\boldsymbol{\theta}_0$. □

Lemma 4. *If Assumption 1 holds, then* $\mathrm{E}\left(\sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} \left\| \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| \right) < \infty$ *and* $\sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} \left\| \sum_{t=2}^{T} \frac{1}{T} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - \mathrm{E} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| = o_p(1).$

Proof. By (S1.3), then $\mathrm{E}\left(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left\|\partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\right\|\right)<\infty$. Similarly, we obtain that

$$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left\|\sum_{t=2}^{T}\frac{1}{T}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}-\mathrm{E}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right\|=o_p(1)$$

by Theorem 3.1 in Ling and McAleer (2003). □

Lemma 5. *If Assumption 1 holds, then* $\mathrm{E}\left(\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^\top\right)=\boldsymbol{V}(\boldsymbol{\theta}_0)>0$, *and* $T^{-1/2}\sum_{t=2}^{T}\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{V}(\boldsymbol{\theta}_0))$.

Proof. Denote $\boldsymbol{\eta}_{t-1}(\boldsymbol{\theta}_0)=\left(\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\mu_t},\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\mu_t}X_{t-1},\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\mu_t}\mu_{t-1},\right.$ $\left.\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\sigma_t^2},\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\sigma_t^2}|X_{t-1}|,\frac{\partial\mathbb{P}_n(x_t(\boldsymbol{\theta}_0))}{\partial\sigma_t^2}\sigma_{t-1}^2\right)^\top$. By (S1.2), we obtain

$$\boldsymbol{V}(\boldsymbol{\theta}_0)=\mathrm{E}\left(\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^\top\right)=\mathrm{E}\left(\mathbb{P}_n^{-2}(x_t(\boldsymbol{\theta}_0))\boldsymbol{\eta}_{t-1}(\boldsymbol{\theta}_0)\boldsymbol{\eta}_{t-1}^\top(\boldsymbol{\theta}_0)\right).$$

If $\boldsymbol{V}(\boldsymbol{\theta}_0)$ is not positive definite, then there exists a constant vector $\boldsymbol{c}=(c_1,\cdots,c_6)^\top$ such that $\boldsymbol{c}^\top\boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{c}=0$, and then

$$\boldsymbol{c}^\top\mathrm{E}\left(\mathbb{P}_n^{-2}(x_t(\boldsymbol{\theta}_0))\boldsymbol{\eta}_{t-1}(\boldsymbol{\theta}_0)\boldsymbol{\eta}_{t-1}^\top(\boldsymbol{\theta}_0)\right)\boldsymbol{c}=\mathrm{E}\left(\left(\mathbb{P}_n^{-2}(x_t(\boldsymbol{\theta}_0))\left(\boldsymbol{\eta}_{t-1}^\top c\right)^2\right)=0.\right.$$

This is impossible. Therefore, $\boldsymbol{V}(\boldsymbol{\theta}_0)$ is positive definite. Then, by using the martingale central limit theorem and Cramér-Wold device, it is direct to obtain $T^{-1/2}\sum_{t=2}^{T}\partial l_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{V}(\boldsymbol{\theta}_0))$. □

Lemma 6. *If Assumption 1 holds, then* $\mathrm{E}\left(\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left\|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\|\right)<\infty$, *and* $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left\|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}-\mathrm{E}\left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right)\right\|=o_p(1)$.

Proof. By Lemma 1 given in supplementary materials S4 and (S1.1), there exist two constants $\underline{c}$ and $\bar{c}$ such that $\underline{c} \leq f_i(x_t) \leq \bar{c}, i = 1, 2, 3$. Therefore, $\forall i = 1, 2, \cdots, 6$, $\forall j = 1, 2, \cdots, 6$, $\left|\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right| \leq n^2 \max\{|\underline{c}|, |\bar{c}|\}$. Thus, $\left\| \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \mathrm{E}\left(\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right) \right\| = o_p(1)$ by Theorem 3.1 in Ling and McAleer (2003). □

S4.2 Auxiliary Discussion with regard to TD2 Distribution

To account for the characteristics of the $Z \sim \mathrm{TD2}(n, \alpha, \nu)$, we present the histogram of Z in Figure F11 for $\mu \in \{-5, 0, 2\}$ and $\sigma^2 \in \{4.5, 5.6, 6\}$ with $n = 10$, when k varies from -10 to 10 with increment one.

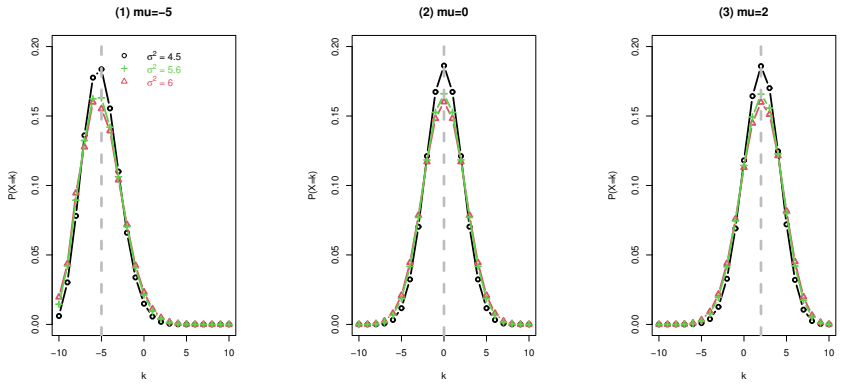


Figure F11: PMF of $\mathrm{TD2}(n, \alpha, \nu)$.

By Figure F11, the sample from $\mathrm{TD2}(n, \alpha, \nu)$ distribution seems unimodal. For fixed σ^2, when $\mu = 0$, the TD distribution is symmetric about

the $x = 0$ and its peak is located at zero. As μ changes, the graph shifts horizontally. If μ increases (decreases), the graph shifts to the right (left). The entire curve moves along the x-axis while maintaining its shape with the peak now located at μ. For fixed μ, if σ^2 decreases (increases), the graph becomes narrower and taller (wider and flatter), the data are more concentrated around the mean (spread out).

Bibliography

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data. *Biometrics* **65**, 1254–1261.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica* **26**, 1673–1707.

Doukhan, P. (1994). *Mixing: properties and examples.* Springer-Verlag.

Doukhan, P., Naushad, M. and Michael, H. (2021). Mixing properties of integer-valued GARCH processes. *ALEA* **18**, 401–420.

Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with

infinite memory, *Stochastic Processes and their Applications* **118**, 1997–2013,

Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104**, 1430–1439.

Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review* **70**, 419–435.

Jensen, S. T. and Rahbek, A. (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203–1226.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.

Liu, H. (2012). *Some models for time series of counts*. Dissertations, Columbia University. https://core.ac.uk/download/pdf/27293903.pdf.

Meyn, S.P. and Tweedie, R.L. (2009). *Markov chains and stochastic stability (second edition)*. Cambridge University Press, New York.

Omair, A., Alzaid, A. and Odhah, O. (2016). A trinomial difference distribution. *Revista Colombiana de Estadística* **39**, 1–15.