In this supplementary appendix, we provide the following discussions and tables that are not shown in the paper for the sake of space.

- Derivation of Equation (29).
- An explicit formula of \( \hat{\Omega} \) for the \( G_\rho \) test and its derivation.
- Derivation of Condition (35).
- An example which shows that the \( G_{BP} \) test may not be asymptotically valid.
- A detailed discussion about the simulation results in Tables 1 and 2.
- A table of the estimators for GARCH-type models.
- A table of the sample correlations with various \((\varphi_{1t}, \varphi_{2t})'\)'s.

### 1 Derivation of (29)

By the definition of \( \hat{\rho}_k \) in (28), we can write that, for any fixed \( k \),

\[
\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1t}^c}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right) = \frac{1}{\sqrt{P}} \left( (P-k)\hat{\rho}_k + \sum_{t=R+1}^{R+k} \left( \frac{\hat{\varphi}_{1t}^c}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right) \right),
\]

\[
= \sqrt{P}\hat{\rho}_k + o_p(1),
\]

where the last equality holds under the assumption that the variance of \( \hat{\varphi}_{1t}^c \hat{\varphi}_{2,t-k/}^c \) is finite. Meanwhile, because \( \hat{\varphi}_{1t}^c = \hat{\varphi}_{1t} - \bar{\varphi}_1 \),

\[
\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1t}^c}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1t}^c}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right) - \left( \frac{\bar{\varphi}_1}{\sigma_1} \right) \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right)
\]

\[
= \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1t}^c}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-k}^c}{\sigma_2} \right) + o_p(1),
\]
where the last equality is due to the condition \( \bar{\varphi}_1 \xrightarrow{D} \mathbb{E} [\varphi_{1,ot}] = 0 \) and the assumption that 
\[ P^{-1/2} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{2,t-k}}{\sigma_2} \right) = \mathcal{O}_p(1). \]
From these two results, we can write that 
\[ \sqrt{P} \hat{\rho} = \left[ \begin{array}{c} \sqrt{P} \hat{\rho}_1 \\ \sqrt{P} \hat{\rho}_2 \\ \vdots \\ \sqrt{P} \hat{\rho}_n \end{array} \right] = \left[ \begin{array}{c} \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1,t}}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-1}}{\sigma_2} \right) \\ \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1,t}}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-2}}{\sigma_2} \right) \\ \vdots \\ \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T} \left( \frac{\hat{\varphi}_{1,t}}{\sigma_1} \right) \left( \frac{\hat{\varphi}_{2,t-n}}{\sigma_2} \right) \end{array} \right] + \mathcal{O}_p(1). \]
Accordingly, we can obtain (29) by using (30).

\[ \Box \]

## 2 Formula of \( \hat{\Omega} \) for the \( G_\rho \) Test

The \( \Omega \) for the \( G_\rho \) test is based on the \( x_t \) and \( \xi_t \) in (30). Correspondingly, for both (21) and (25), \( \Sigma \) is a \( n \times n \) matrix with the \( (i,j) \)-th element:

\[ \Sigma_{ij} := \mathbb{E} [\varphi_{1,ot}^2 \varphi_{2,ot-i}^2 \varphi_{2,ot-j}^2], \tag{A1} \]

for \( i, j = 1, \ldots, n \). In the case where \( \varphi_{1t} = \varphi_{1t}^{(1)} \) or \( \varphi_{1t}^{(2)} \), \( \Omega \) is in the form of (21) and comprises the matrices based on the Gaussian QML method and (13):

\[ A = \begin{cases} A_1 := (A_{11}, \ldots, A_{1n})^\top, & \text{if } \varphi_{1t} = \varphi_{1t}^{(1)}, \\
A_2 := (A_{21}, \ldots, A_{2n})^\top, & \text{if } \varphi_{1t} = \varphi_{1t}^{(2)}, \end{cases} \tag{A2} \]

with \( A_{1k} := -\mathbb{E} [\varphi_{2,ot-k}^2 w_{1,ot}^2] \) and \( A_{2k} := -\mathbb{E} [\varphi_{2,ot-k}^2 z_{1,ot}^2] \),

\[ B = -\left( \mathbb{E} [w_{1,ot} w_{1,ot}^\top] + \frac{1}{2} \mathbb{E} [z_{1,ot} z_{1,ot}^\top] \right), \]

\[ C = \begin{cases} -\left( A_1 + \frac{1}{2} \mathbb{E} [\varepsilon_{1,ot}^3] A_2 \right), & \text{if } \varphi_{1t} = \varphi_{1t}^{(1)}, \\
-\left( \mathbb{E} [\varepsilon_{1,ot}^3] A_1 + \frac{1}{2} \left( \mathbb{E} [\varepsilon_{1,ot}^4] - 1 \right) A_2 \right), & \text{if } \varphi_{1t} = \varphi_{1t}^{(2)}, \end{cases} \]

and

\[ D = \mathbb{E} [w_{1,ot} w_{1,ot}^\top] + \frac{1}{2} \mathbb{E} [\varepsilon_{1,ot}^3] \left( \mathbb{E} [w_{1,ot} z_{1,ot}^\top] + \mathbb{E} [z_{1,ot} w_{1,ot}^\top] \right) + \frac{1}{4} \left( \mathbb{E} [\varepsilon_{1,ot}^4] - 1 \right) \mathbb{E} [z_{1,ot} z_{1,ot}^\top], \]
where \( w_{1,ot} := w_{1t}|_{\alpha_1 = \alpha_{1o}} \) and \( z_{1,ot} := z_{1t}|_{\alpha_1 = \alpha_{1o}} \). In the case where \( \varphi_{1t} = \varphi_{1t}^{(q)} \), \( \Omega \) is in the form of (25), and comprises the matrices based on the ML method and (15):

\[
C = \begin{bmatrix}
E \left[ \varphi_{2,ot-1} \varphi_{1,ot} \ell_{\theta_1,ot}^T \right] \\
\vdots \\
E \left[ \varphi_{2,ot-n} \varphi_{1,ot} \ell_{\theta_1,ot}^T \right]
\end{bmatrix}
\]

and the \( D \) in (24), where \( \ell_{\theta_1,ot} \) follows (16) with \( \theta_1 = \theta_{1o} \).

The following \( \hat{\Omega} \) is given by the sample analogue of this \( \Omega \). For \( \varphi_{1t} = \varphi_{1t}^{(1)} \) or \( \varphi_{1t}^{(2)} \), \( \hat{\Omega} \) is of the form:

\[
\hat{\Omega} = \hat{\Sigma} - \hat{\eta}_1 (\hat{A} \hat{B}^{-1} \hat{C}^T + \hat{C} (\hat{A} \hat{B}^{-1})^T) + \hat{\eta}_2 \hat{A} \hat{B}^{-1} \hat{D} (\hat{A} \hat{B}^{-1})^T
\]

(A3)

and based on the Gaussian QMLE \( \hat{\theta}_{1t} = \hat{\alpha}_{1t} \) for \( \theta_{1o} = \alpha_{1o} \) in (7). For \( \varphi_{1t} = \varphi_{1t}^{(q)} \) or any other \( \varphi_{1t} \) that cannot be expressed as a linear combination of \( \varphi_{1t}^{(1)} \) and \( \varphi_{1t}^{(2)} \), \( \hat{\Omega} \) is of the form:

\[
\hat{\Omega} = \hat{\Sigma} + (\hat{\eta}_2 - 2\hat{\eta}_1) \hat{C} \hat{D}^{-1} \hat{C}^T
\]

(A4)

and based on the MLE \( \hat{\theta}_{1t} = (\hat{\alpha}_{1t}^T, \hat{\beta}_{1t}^T)^T \) for \( \theta_{1o} = (\alpha_{1o}^T, \beta_{1o}^T)^T \) in (8). The components of (A3) and (A4) are defined below.

For both (A3) and (A4), \( (\hat{\eta}_1, \hat{\eta}_2) := (\eta_1, \eta_2)|_{\lambda = \hat{\lambda}} \) where \( \hat{\lambda} := P/R \), and

\[
\hat{\Sigma} = \frac{1}{P - n} Y^T Y
\]

with the \((P - n) \times n\) matrix:

\[
Y := [Y_1, Y_2, \ldots, Y_n], \quad \text{where} \quad Y_k := \begin{bmatrix}
\hat{\varphi}_{1,R+1+n} \\
\vdots \\
\hat{\varphi}_{1,T} \\
\hat{\varphi}_{2,R+1+n-k} \\
\vdots \\
\hat{\varphi}_{2,T-k}
\end{bmatrix},
\]

For (A3),

\[
\hat{A} = \begin{cases}
\frac{1}{P - n} \Phi^T W, & \text{if } \varphi_{1t} = \varphi_{1t}^{(1)}, \\
\frac{1}{P - n} \Phi^T Z, & \text{if } \varphi_{1t} = \varphi_{1t}^{(2)},
\end{cases}
\]

\[
\hat{B} = -\left( \frac{1}{P - n} W^T W + \frac{1}{2(P - n)} Z^T Z \right),
\]

with

\[
\Phi := \begin{bmatrix}
\varphi_{1,ot} \\
\varphi_{2,ot-1} \\
\vdots \\
\varphi_{2,ot-n}
\end{bmatrix},
\]

\[
\Omega := E \left[ \Phi \Phi^T \right],
\]

\[
\Phi^T := \begin{bmatrix}
\varphi_{1,0} \\
\varphi_{2,0-1} \\
\vdots \\
\varphi_{2,0-n}
\end{bmatrix},
\]

\[
\Omega^T := E \left[ \Phi^T \Phi \right],
\]

and

\[
\varphi_{1,0} := \begin{bmatrix}
\varphi_{1,0} \\
\varphi_{2,0-1} \\
\vdots \\
\varphi_{2,0-n}
\end{bmatrix},
\]

\[
\varphi_{1,ot} := \begin{bmatrix}
\varphi_{1,ot} \\
\varphi_{2,ot-1} \\
\vdots \\
\varphi_{2,ot-n}
\end{bmatrix},
\]

\[
E \left[ \Phi \Phi^T \right] := \begin{bmatrix}
E \left[ \varphi_{1,ot} \varphi_{1,ot}^T \right] \\
\vdots \\
E \left[ \varphi_{2,ot-n} \varphi_{1,ot}^T \right]
\end{bmatrix}
\]

\[
E \left[ \Phi^T \Phi \right] := \begin{bmatrix}
E \left[ \varphi_{1,0} \varphi_{1,0}^T \right] \\
\vdots \\
E \left[ \varphi_{2,0-n} \varphi_{2,0-n}^T \right]
\end{bmatrix}
\]
\[
\hat{C} = \begin{cases} 
\frac{1}{p-n} \Phi^T W + \frac{1}{2} \left( \frac{1}{p-n} \sum_{t=R+n+1}^{T} \varepsilon_{lt}^{3} \right) \frac{1}{p-n} \Phi^T Z, & \text{if } \varphi_{1t} = \varphi_{1t}^{(1)}, \\
\quad \left( \frac{1}{p-n} \sum_{t=R+n+1}^{T} \varepsilon_{lt}^{3} \right) \frac{1}{p-n} \Phi^T W \quad & \text{if } \varphi_{1t} = \varphi_{1t}^{(2)}, \\
\quad \frac{1}{4} \left( \frac{1}{p-n} \sum_{t=R+n+1}^{T} \varepsilon_{lt}^{4} - 1 \right) \frac{1}{p-n} \Phi^T Z, & \end{cases}
\]

and

\[
\hat{D} = \frac{1}{p-n} W^T W + \frac{1}{2} \left( \frac{1}{p-n} \sum_{t=R+n+1}^{T} \varepsilon_{lt}^{3} \right) \frac{1}{p-n} (W^T Z + Z^T W) \\
+ \frac{1}{4} \left( \frac{1}{p-n} \sum_{t=R+n+1}^{T} \varepsilon_{lt}^{4} - 1 \right) \frac{1}{p-n} Z^T Z,
\]

with the \((p-n) \times n\) matrix:

\[
\Phi := [\Phi_1, \Phi_2, \ldots, \Phi_n], \quad \text{where } \Phi_k := \begin{bmatrix} \hat{\varphi}_{2,R+1+n-k}^c \\
\hat{\varphi}_{2,R+2+n-k}^c \\
\vdots \\
\hat{\varphi}_{2,T-k}^c \end{bmatrix},
\]

and the \((p-n) \times \dim(\alpha_1)\) matrices:

\[
W := \begin{bmatrix} \hat{w}_{1,R+1+n}^T \\
\hat{w}_{1,R+2+n}^T \\
\vdots \\
\hat{w}_{1,T}^T \end{bmatrix} \quad \text{and} \quad Z := \begin{bmatrix} \hat{z}_{1,R+1+n}^T \\
\hat{z}_{1,R+2+n}^T \\
\vdots \\
\hat{z}_{1,T}^T \end{bmatrix},
\]

in which \(\varepsilon_{lt} := \varepsilon_{lt}|_{\alpha_1 = \hat{\alpha}_1}\), \(w_{lt} := w_{lt}|_{\alpha_1 = \hat{\alpha}_1}\) and \(z_{lt} := z_{lt}|_{\alpha_1 = \hat{\alpha}_1}\) are based on the Gaussian QMLE \(\hat{\alpha}_1\); recall that \(w_{lt}\) and \(z_{lt}\) are defined following (13). For \(\mathbf{A4}\),

\[
\hat{C} = \frac{1}{p-n} Y^T X \quad \text{and} \quad \hat{D} = \frac{1}{p-n} X^T X,
\]

with the \((p-n) \times \dim(\theta_1)\) matrix:

\[
X := \begin{bmatrix} \hat{\ell}_{\theta_1,R+1+n}^T \\
\hat{\ell}_{\theta_1,R+2+n}^T \\
\vdots \\
\hat{\ell}_{\theta_1,T}^T \end{bmatrix}, \quad \text{where } \hat{\ell}_{\theta_1,t} := \begin{bmatrix} -\left( \hat{w}_{lt} \hat{\ell}_{\varepsilon_1,t} + \frac{1}{2} \hat{\varepsilon}_{lt} \left( \hat{\ell}_{\varepsilon_1,t} \hat{\varepsilon}_{lt} + 1 \right) \right) \\
\hat{\ell}_{\beta_1,t} \end{bmatrix},
\]

in which \(\hat{w}_{lt} := w_{lt}|_{\alpha_1 = \hat{\alpha}_1}\), \(\hat{z}_{lt} := z_{lt}|_{\alpha_1 = \hat{\alpha}_1}\), \(\hat{\varepsilon}_{lt} := \varepsilon_{lt}|_{\alpha_1 = \hat{\alpha}_1}\), \(\hat{\ell}_{\varepsilon_1,t} := \ell_{\varepsilon_1,t}|_{\beta_1 = \hat{\beta}_1}\) and \(\hat{\ell}_{\beta_1,t} := \ell_{\beta_1,t}|_{\beta_1 = \hat{\beta}_1}\) are based on the MLE \(\hat{\theta}_1\); see (16) for \(\ell_{\varepsilon_1,t}\) and \(\ell_{\beta_1,t}\). \(\square\)
3 Derivation of Condition (35)

The independence between \( \{\varepsilon_{1,ot}\} \) and \( \{\varepsilon_{2,ot}\} \) implies that (A2) degenerates to

\[
A = 0; \tag{A5}
\]

that is, the effect of parameter estimation uncertainty on \( \Omega \) is asymptotically negligible in the sense that

\[
\Omega = \Sigma. \tag{A6}
\]

Meanwhile, this independence also allows us to rewrite (A1) as:

\[
\Sigma_{ij} = \mathbb{E}[\varphi_{1,ot}^{2}]\mathbb{E}[\varphi_{2,ot-i}^{2}],
\]

Given the implication of (17): \( \mathbb{E}[\varphi_{1,ot}] = 0 \) and the restriction: \( \mathbb{E}[\varphi_{2,ot}^{2}] = 0 \), we can further use the IIDness of \( \{\varepsilon_{2,ot}\} \) to show that

\[
\Sigma_{ij} = \begin{cases} 
\sigma_{1}^{2}\sigma_{2}^{2}, & i = j, \\
0, & i \neq j;
\end{cases}
\]

that is,

\[
\Sigma = (\sigma_{1}^{2}\sigma_{2}^{2}) \times I_{n} = V^{2}. \tag{A7}
\]

Thus, given (A6) and (A7), the independence assumption implies condition (35). \( \square \)

Example: AR(1)

Let \( F(\cdot|\mathcal{Y}_{t-1}) \) be the conditional distribution of \((y_{1t}, y_{2t})|\mathcal{Y}_{t-1}\). Suppose that \( F(\cdot|\mathcal{Y}_{t-1}) \) is a bivariate conditional normal distribution driven by a stationary AR(1) process:

\[
y_{it} = \alpha_{i1} + \alpha_{i2}y_{i,t-1} + \alpha_{i3}\varepsilon_{it},
\]

where \( |\alpha_{i2}| < 1 \), for both \( i = 1, 2 \), and

\[
(\varepsilon_{1t}, \varepsilon_{2t})|\mathcal{Y}_{t-1} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{A8}
\]

with the correlation coefficient \( \rho \), for all \( t \)'s. In this example, we focus on the \( G_{\rho} \) tests with \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(1)}, \varphi_{2t}^{(1)}) \) and \((\varphi_{1t}^{(2)}, \varphi_{2t}^{(2)}) \) for simplicity.
The AR(1) model implies that $\mu_i(\alpha_i) = \alpha_{i1} + \alpha_{i2} y_{i,t-1}$, $h_i(\alpha_i) = \alpha_{i3}^2$ and $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})^\top$, and has the MA($\infty$) representation:

$$y_{i,t} = \frac{\alpha_{i1}}{1 - \alpha_{i2}} + \alpha_{i3} \sum_{j=0}^{\infty} \alpha_{i2}^j \varepsilon_{i,t-j}.$$ 

Accordingly, we can use (A8) and the formulae of $w_{1t}$ and $z_{1t}$ in this example:

$$w_{1t} = \frac{1}{\alpha_{i3}} \begin{bmatrix} 1 \\ y_{1,t-1} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_{i3}}(1 - \alpha_{i2})^{-1} + \sum_{j=0}^{\infty} \alpha_{i2}^j \varepsilon_{1,t-1-j} \\ 0 \end{bmatrix}$$

and $z_{1t} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

to show that (A2) has the following elements:

$$A_{1k}^\top = - \begin{bmatrix} \sum_{j=0}^{\infty} \alpha_{i2}^j \mathbb{E}[\varepsilon_{1,ot-1-j}] \\ 0 \end{bmatrix} = - \begin{bmatrix} 0 \\ \rho \alpha_{12}^{k-1} \end{bmatrix}$$

and $A_{2k}^\top = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Thus, (A2) has the components:

$$A_1 = \begin{bmatrix} 0 & -\rho & 0 \\ 0 & -\rho \alpha_{12} & 0 \\ \vdots & \vdots & \vdots \\ 0 & -\rho \alpha_{12}^{n-1} & 0 \end{bmatrix}$$

and $A_2$ is the $n \times 3$ zero matrix.

Note that (A8) implies $\mathbb{E}[\varepsilon_{1,ot}^2 | \mathcal{Y}_{t-1}] = 1$ and $\mathbb{E}[\varepsilon_{1,ot}^4 | \mathcal{Y}_{t-1}] = 3$. Accordingly, we can further use the law of iterated expectations to rewrite (A1) as:

$$\Sigma_{ij} = \mathbb{E}[\varepsilon_{1,ot}^2 \varepsilon_{2,ot-i} \varepsilon_{2,ot-j}] = \mathbb{E}[\mathbb{E}[\varepsilon_{1,ot}^2 | \mathcal{Y}_{t-1}] \varepsilon_{2,ot-i} \varepsilon_{2,ot-j}] = \mathbb{E}[\varepsilon_{2,ot-i} \varepsilon_{2,ot-j}]$$

when $(\varphi_{1t}, \varphi_{2t}) = \left( \varphi_{1t}^{(1)}, \varphi_{2t}^{(1)} \right)$. And

$$\Sigma_{ij} = \mathbb{E}[(\varepsilon_{1,ot}^2 - 1)^2 (\varepsilon_{2,ot-i}^2 - 1)(\varepsilon_{2,ot-j}^2 - 1)]$$

$$= \mathbb{E}[\mathbb{E}[(\varepsilon_{1,ot}^2 - 1)^2 | \mathcal{Y}_{t-1}] (\varepsilon_{2,ot-i}^2 - 1)(\varepsilon_{2,ot-j}^2 - 1)]$$

when $(\varphi_{1t}, \varphi_{2t}) = \left( \varphi_{1t}^{(2)}, \varphi_{2t}^{(2)} \right)$. Because (A8) also implies that $\mathbb{E}[\varepsilon_{2,ot}^2 | \mathcal{Y}_{t-1}] = 0$ and $\mathbb{E}[\varepsilon_{2,ot}^2 | \mathcal{Y}_{t-1}] = 1$, we can also use the law of iterated expectations to show that $\Sigma_{ij} = 0$ when $i \neq j$. Thus, (A7), that is $\Sigma = V^2$, holds for this example. Meanwhile, the Gaussian
QML method also degenerates to the Gaussian ML method under (A8). Thus, the Ω in (21) can be simplified as (25). Accordingly, we can write that

\[ \Omega - V^2 = (\eta_2 - 2\eta_1)CD^{-1}C^T. \]

By the generalized information matrix equality: \( C = -A = -A_1 \) if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(1)}, \varphi_{2t}^{(1)})\), \( C = -A = -A_2 = 0 \) if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(2)}, \varphi_{2t}^{(2)})\) and

\[ D = -B = \begin{bmatrix}
\frac{1}{\alpha_{13}^2} & \frac{\alpha_{11}}{(1-\alpha_{12})\alpha_{13}^2} & \frac{\alpha_{12}^2 - 1}{(1-\alpha_{12})^2\alpha_{13}^2} & 0 \\
\frac{\alpha_{11}}{(1-\alpha_{12})\alpha_{13}^2} & \frac{\alpha_{11}}{(1-\alpha_{12})\alpha_{13}^2} & 0 & 0 \\
0 & 0 & 0 & \frac{2}{\alpha_{13}^2} \\
0 & 0 & 0 & 0
\end{bmatrix}, \]

we can further write that

\[ \Omega - V^2 = \rho^2(\eta_2 - 2\eta_1)d \begin{bmatrix}
1 & \alpha_{12} & \ldots & \alpha_{12}^{n-1} \\
\alpha_{12} & \alpha_{12}^2 & \ldots & \alpha_{12}^n \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{12}^{n-1} & \alpha_{12}^n & \ldots & \alpha_{12}^{2(n-1)}
\end{bmatrix}, \]

where

\[ d := \frac{1}{\alpha_{13}^2} \left( \frac{\alpha_{11}^2}{(1-\alpha_{12})^2\alpha_{13}^2} + \frac{1}{1-\alpha_{12}} \right) - \frac{\alpha_{12}^2}{(1-\alpha_{12})^2\alpha_{13}^2} = 1 - \alpha_{12}^2 \]

if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(1)}, \varphi_{2t}^{(1)})\) and \(\Omega = V^2\) if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(2)}, \varphi_{2t}^{(2)})\) because \(C = 0\). Thus,

\[ \Omega - V^2 = \rho^2(\eta_2 - 2\eta_1)(1 - \alpha_{12}^2) \begin{bmatrix}
1 & \alpha_{12} & \ldots & \alpha_{12}^{n-1} \\
\alpha_{12} & \alpha_{12}^2 & \ldots & \alpha_{12}^n \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{12}^{n-1} & \alpha_{12}^n & \ldots & \alpha_{12}^{2(n-1)}
\end{bmatrix} \quad (A9) \]

if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(1)}, \varphi_{2t}^{(1)})\), and \(\Omega = V^2\) if \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(2)}, \varphi_{2t}^{(2)})\).

This result means that, in this example, the \(G_{BP}\) test is asymptotically equivalent to the \(G_\rho\) test in checking Granger causality in variance, but is not ensured to be asymptotically valid in checking Granger causality in mean because condition (35) may fail in certain cases. This condition holds for \((\varphi_{1t}, \varphi_{2t}) = (\varphi_{1t}^{(1)}, \varphi_{2t}^{(1)})\) when \(\rho = 0\) or \(\eta_2 = 2\eta_1\), but may fail in other cases. From (A9) and the stationarity condition: \(|\alpha_{12}| < 1\) which implies that \(|\alpha_{12}^k|\) decays to zero as \(k\) increases, we can also observe that the discrepancy
between $\Omega$ and $V^2$, if it appears, is mainly generated by the $k \times k$ upper-left submatrix of $\Omega - V^2$ for some small $k$. Thus, fixing (A9), we may expect that the approximation error of the $G_{BP}$ test would be milder when $S\hat{\rho} = \hat{\rho}_k$ with a large $k$ or when $S = I_n$ with a larger $n$. This reminds us that, without the independence assumption, the $G_{BP}$ test should be used with caution because its asymptotic validity would be dependent on the choice of $\varphi_{it}$, the setting of $S$, the unknown $F(\cdot|Y_{t-1})$ and the full-sample or out-of-sample context being used.

4 Results of Tables 1 and 2

Regarding the size performance of the $G_{\rho}$ test under DGP1, Table 1 shows that the empirical sizes of the $G_{\rho}$ test are reasonably close to the 5% level for most cases being considered, especially when $T = 2000$. This evidence supports the validity of the asymptotic distribution theory of the $G_{\rho}$ test. However, Table 1 also indicates that, given $T = 1000$, the full-sample $G_{\rho}$ test is of the empirical size: 11.6% (18.4%) when $\varphi_{it} = \varphi_{it}^{(2)}$, $S\hat{\rho} = \hat{\rho}_1$ ($S = I_n$) and $\gamma_c = 0.9$ and the empirical sizes: 11.7%, 10.1% and 10.0% when $\varphi_{it} = \varphi_{it}^{(q_b)}$, $S = I_n$ and $\gamma_c = 0, 0.5$ and 0.9, respectively. In these exceptional cases, the full-sample $G_{\rho}$ test is obviously over-sized. Nonetheless, because the associated empirical sizes of this test decay to the values: 8.9% (12.1%), 7.7%, 6.6% and 7.0%, respectively, when $T$ increases to 2000, we may regard this size-distortion as a finite-sample problem.

The above result suggests that the $G_{\rho}$ test is suitable for applications with large sample sizes, such as empirical studies with daily or high-frequency financial data. This is consistent with the fact that the $G_{\rho}$ test is designed for financial time series analysis. Nonetheless, the result also reminds us that the $G_{\rho}$ test may not behave uniformly well for all possible $\varphi_{it}$’s, if the sample size is not large enough. To further discuss this finite-sample issue, we make the following comparison. Firstly, this issue appears when the $G_{\rho}$ test is applied to testing Granger causality in lower quantiles (with $\varphi_{it} = \varphi_{it}^{(q_b)}$) or in variance (with $\varphi_{it} = \varphi_{it}^{(2)}$) under high correlation ($\gamma_c = 0.9$). In comparison, the $G_{\rho}$ test performs reasonably well in testing Granger causality in mean (with $\varphi_{it} = \varphi_{it}^{(1)}$) or testing Granger causality in the left part of distribution (with $\varphi_{it} = \varphi_{it}^{(q_a)}$), even when $T = 1000$. Thus, this finite-sample issue is likely to be influenced by the choice of $\varphi_{it}$. Secondly, as explained before, DGP2 is more complicated than DGP1 for the size experiments of the $G_{\rho}$ tests with $\varphi_{it} = \varphi_{it}^{(1)}$ and $\varphi_{it}^{(2)}$. Nonetheless, by comparing Table 1 with Table 2, we can see that the empirical sizes of these two $G_{\rho}$ tests are close to the 5% level under
both DGP1 and DGP2 when \( T = 1000 \) and \( \gamma_c \leq 0.5 \). Thus, the change from DGP1 to DGP2 itself does not cause additional finite-sample size distortions to the \( G_\rho \) tests with \( \varphi_{it} = \varphi_{it}^{(1)} \) and \( \varphi_{it}^{(2)} \). This indeed reflects the fact that these two tests are established by the Gaussian QML method and robust to the conditional non-normality. It may be interesting to further explore whether the finite-sample issue could be remedied when the \( G_\rho \) test is applied to a simpler model, such as VAR, but such a comparison is beyond our simulation context.

From Table 1, we can also observe that, like the \( G_\rho \) test, the \( G_{BP} \) test also has proper size performance under DGP1 when \( \gamma_c = 0 \). This is consistent with the fact that the \( G_\rho \) test and the \( G_{BP} \) test are asymptotically equivalent in this case. By contrast, the \( G_{BP} \) tests could be substantially size-distorted in the presence of contemporaneous dependence. Given \( \gamma_c = 0.5 \) (\( \gamma_c = 0.9 \)), the full-sample \( G_{BP} \) test with \( \varphi_{it} = \varphi_{it}^{(1)} \) and \( \hat{S}\hat{\rho} = \hat{\rho}_1 \) has the empirical size: 3.8% (0.5%) even when \( T = 2000 \), and the out-of-sample \( G_{BP} \) tests with \( \lambda = 1, 2 \) and 5 are, respectively, of the empirical sizes: 7.7%, 9.1% and 13.2% (11.5%, 18.3% and 31.5%) when \( R = 2000 \). In comparison, the \( G_\rho \) test has the empirical sizes: 6.7%, 5.2%, 4.2% and 4.6% (6.3%, 4.9%, 4.7% and 4.2%) in the associated cases. Clearly, unlike the \( G_\rho \) test, the full-sample (out-of-sample) \( G_{BP} \) test tends to be under-sized (over-sized especially when \( \lambda \) and hence \( \eta_2 - 2\eta_1 \) become large) under contemporaneous dependence.

Similar but milder size-distortions can also be found in Table 1 for the \( G_{BP} \) tests with \( \varphi_{it} = \varphi_{it}^{(2)} \) and the \( \varphi_{it}^{(q)} \)'s. This comparison shows that the \( G_{BP} \) test is not robust to the failure of the independence assumption, and confirms the theoretical advantage of the \( G_\rho \) test over the \( G_{BP} \) test in this scenario. However, as explained in Section 3.2 and shown by Table 1, this problem becomes milder for the \( G_{BP} \) test when \( \hat{S}\hat{\rho} = \hat{\rho}_k \) with a larger \( k \).

Table 2 shows that the \( G_\rho \) test and the \( G_{BP} \) test share the following features in their power performance. Firstly, the empirical powers of the full-sample (out-of-sample) tests systematically increase with \( T \) (\( R \) and \( P \)) under DGP3 or DGP4 for all the \( \varphi_{it} \)'s being considered and for \( \gamma_c = 0, 0.5 \). Secondly, the tests with different \( \varphi_{it} \)'s may perform in different ways under the same DGP. In particular, the tests with \( \varphi_{it} = \varphi_{it}^{(1)} \) or \( \varphi_{it}^{(2)} \) are more powerful than the tests with the \( \varphi_{it}^{(q)} \)'s under DGP3. This is sensible because, under DGP3, \( y_{2t} \) directly Granger-causes \( y_{1t} \) in mean and variance but indirectly Granger-causes \( y_{1t} \) in quantiles. In addition, the empirical powers of the tests with \( \varphi_{it} = \varphi_{it}^{(1)} \) or \( \varphi_{it}^{(2)} \) under DGP3 are quite close to their counterparts under DGP4. This is because, by construction, these tests are robust to Granger causality in higher moments. Thirdly, the tests with
When \( \varphi_{it} = \varphi_{it}^{(qa)} \) under DGP4, it is more powerful than their counterparts under DGP3. This reasonably reflects that \( y_{2t} \) also directly Granger-causes \( y_{1t} \) in quantiles under DGP4. However, the empirical powers of the tests with \( \varphi_{it} = \varphi_{it}^{(qb)} \) under DGP3 are quite close to their counterparts under DGP4. Fourthly, the tests with \( \varphi_{it} = \varphi_{it}^{(qa)} \) or \( \varphi_{it}^{(qb)} \) are essential of no power under DGP2. This suggests that it is difficult to detect Granger causality in quantiles if there is no Granger causality in mean or variance.

Table 2 also indicates that the empirical powers of the \( G_{BP} \) test tend to be higher than those of the \( G_{\rho} \) test in several cases, especially when \( T = 1000 \) and \( \varphi_{it} = \varphi_{it}^{(2)} \) or \( \varphi_{it}^{(qa)} \). However, the explanation about this result should be dependent on whether \( \gamma_c = 0 \) or \( \gamma_c = 0.5 \). In the case where \( \gamma_c = 0 \) and the \( G_{BP} \) test is asymptotically valid, the power advantage of the \( G_{BP} \) test over the \( G_{\rho} \) test is likely due to the fact that these two tests deal with the asymptotic covariance matrix \( \Omega \) in a different way. As mentioned previously, the \( G_{BP} \) test is established by using condition (35), and hence is free of the sampling variations caused by estimating \( V \) and \( \Omega \). However, in the case where \( \gamma_c = 0.5 \), the above restriction is not ensured to be valid and the empirical powers of the \( G_{BP} \) test may be contaminated by the size-distortions of the \( G_{BP} \) test.

Generally speaking, this simulation suggests that the \( G_{BP} \) test tends to have better finite-sample power performance than the \( G_{\rho} \) test when both tests are asymptotically valid. It is more attractive than the \( G_{\rho} \) test in applications because of its simplicity and power advantage in this scenario. However, the \( G_{BP} \) test is not robust to the contemporaneous dependence of standardized errors. To be conservative, the \( G_{\rho} \) test may be recommended for applications when this data structure is empirically relevant.

## 5 Empirical Tables

In this section, we present two tables mentioned in the empirical part of the paper.
Table A.1: The estimators for GARCH-type models

<table>
<thead>
<tr>
<th>Models</th>
<th>$y_{it} = y_{s}^t$</th>
<th></th>
<th>$y_{it} = y_{b}^t$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{i2}$</td>
<td>0.01088</td>
<td>0.00816</td>
<td>0.00097</td>
<td>-0.00030</td>
</tr>
<tr>
<td>(1.40361)</td>
<td>(0.25894)</td>
<td>(0.00461)</td>
<td>(0.00461)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{i3}$</td>
<td>-0.06831</td>
<td>-0.07232</td>
<td>0.00270</td>
<td>0.00048</td>
</tr>
<tr>
<td>(0.01887)</td>
<td>(0.01197)</td>
<td>(0.02021)</td>
<td>(0.01984)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{i4}$</td>
<td>-0.06607</td>
<td>-0.06544</td>
<td>0.00041</td>
<td>0.00044</td>
</tr>
<tr>
<td>(0.01437)</td>
<td>(0.00595)</td>
<td>(0.00015)</td>
<td>(0.00009)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{i5}$</td>
<td>0.98662</td>
<td>0.98794</td>
<td>0.96170</td>
<td>0.95966</td>
</tr>
<tr>
<td>(0.00332)</td>
<td>(0.00103)</td>
<td>(0.00379)</td>
<td>(0.00290)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{i6}$</td>
<td>-0.10985</td>
<td>-0.11535</td>
<td>0.03225</td>
<td>0.03382</td>
</tr>
<tr>
<td>(0.01298)</td>
<td>(0.00572)</td>
<td>(0.00015)</td>
<td>(0.00009)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{i1}$</td>
<td>0.08591</td>
<td>0.08521</td>
<td>0.08591</td>
<td>0.08521</td>
</tr>
<tr>
<td>(0.01903)</td>
<td>(0.00791)</td>
<td>(0.00791)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{i2}$</td>
<td>1.49378</td>
<td>1.18161</td>
<td>1.18161</td>
<td>1.18161</td>
</tr>
<tr>
<td>(0.16106)</td>
<td>(0.15505)</td>
<td>(0.15505)</td>
<td>(0.15505)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The estimators for (7) and (8) are, respectively, the Gaussian MLEs and the MLEs for the parameters of the AR(1)-EGARCH(1,1) model and the AR(1)-EGARCH(1,1)-APD model of \{y_{s}^t\} and for the parameters of the AR(1)-GARCH(1,1) model and the AR(1)-GARCH(1,1)-APD model of \{y_{b}^t\}. The specifications of AR(1), GARCH(1,1) and standardized APD are the same as those considered in the simulation with the restriction: $\gamma_i = 0$. The EGARCH(1,1) specification is of the form: $h_{it} = \exp (\alpha_{i3} + \alpha_{i4} \ln h_{i,t-1} + \alpha_{i5}s_{i,t} + \alpha_{i6}|s_{i,t}|)$. The entries in the parentheses are the asymptotic standard deviation estimates.
Table A.2: Sample correlations with various \((\varphi_{1t}, \varphi_{2t})\)'s.

<table>
<thead>
<tr>
<th>(\varphi_{2t} )</th>
<th>(\hat{\rho}_k)</th>
<th>(\varphi_{1t} = (\varphi_{1t}^{(1)}, \varphi_{1t}^{(2)}, \varphi_{1t}^{(q_1)}, \varphi_{1t}^{(q_2)}, \varphi_{1t}^{(q_3)}, \varphi_{1t}^{(q_4)}, \varphi_{1t}^{(q_5)}))</th>
<th>(\varphi_{1t} = (\varphi_{1t}^{(1)}, \varphi_{1t}^{(2)}, \varphi_{1t}^{(q_1)}, \varphi_{1t}^{(q_2)}, \varphi_{1t}^{(q_3)}, \varphi_{1t}^{(q_4)}, \varphi_{1t}^{(q_5)}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varphi_{2t}^{(1)})</td>
<td>(\hat{\rho}_0)</td>
<td>(0.219)</td>
<td>(-0.219)</td>
</tr>
<tr>
<td>min</td>
<td>0.073</td>
<td>0.167</td>
<td>0.011</td>
</tr>
<tr>
<td>max</td>
<td>0.035</td>
<td>0.019</td>
<td>-0.005</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(2)})</td>
<td>(\hat{\rho}_0)</td>
<td>(-0.035)</td>
<td>(-0.023)</td>
</tr>
<tr>
<td>min</td>
<td>0.173</td>
<td>0.082</td>
<td>0.025</td>
</tr>
<tr>
<td>max</td>
<td>0.020</td>
<td>0.019</td>
<td>0.026</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(q_1)})</td>
<td>(\hat{\rho}_0)</td>
<td>(0.141)</td>
<td>(-0.33)</td>
</tr>
<tr>
<td>min</td>
<td>0.028</td>
<td>-0.073</td>
<td>-0.021</td>
</tr>
<tr>
<td>max</td>
<td>0.017</td>
<td>0.016</td>
<td>0.019</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(q_2)})</td>
<td>(\hat{\rho}_0)</td>
<td>(-0.002)</td>
<td>(-0.032)</td>
</tr>
<tr>
<td>min</td>
<td>0.012</td>
<td>-0.053</td>
<td>-0.061</td>
</tr>
<tr>
<td>max</td>
<td>0.000</td>
<td>0.047</td>
<td>0.023</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(q_3)})</td>
<td>(\hat{\rho}_0)</td>
<td>(-0.043)</td>
<td>(-0.024)</td>
</tr>
<tr>
<td>min</td>
<td>0.022</td>
<td>-0.053</td>
<td>-0.061</td>
</tr>
<tr>
<td>max</td>
<td>0.000</td>
<td>0.030</td>
<td>0.024</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(q_4)})</td>
<td>(\hat{\rho}_0)</td>
<td>(-0.020)</td>
<td>(-0.002)</td>
</tr>
<tr>
<td>min</td>
<td>0.014</td>
<td>-0.019</td>
<td>-0.038</td>
</tr>
<tr>
<td>max</td>
<td>0.000</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td>(\varphi_{2t}^{(q_5)})</td>
<td>(\hat{\rho}_0)</td>
<td>(-0.090)</td>
<td>(-0.018)</td>
</tr>
<tr>
<td>min</td>
<td>0.022</td>
<td>-0.040</td>
<td>-0.088</td>
</tr>
<tr>
<td>max</td>
<td>0.000</td>
<td>0.065</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Note: The entries in “min” and “max” are, respectively, the minimum and maximum values of the \(\hat{\rho}_k\)’s with \(k = 1, 2, \ldots, 10\).