SUPPLEMENTARY MATERIAL FOR

Testing Forecast Rationality for Measures of Central Tendency

Timo Dimitriadis and Andrew J. Patton and Patrick W. Schmidt

This Version: September 28, 2020

S.1 Technical Proofs

Lemma S.1.1. *Given Assumption 2.5 and the null hypothesis in (2.8), it holds that* $\sum_{t=1}^{T} g_{t,T}^{e} \xrightarrow{P} 0$.

Proof. Applying integration by parts yields that

$$g_{t,T}^{e} = -\mathbb{E}_t\left[(T\delta_T)^{-1/2}K'\left(\frac{\varepsilon_t}{\delta_T}\right)\boldsymbol{h}_t\right] = -(T\delta_T)^{-1/2}\boldsymbol{h}_t\int K'\left(\frac{e}{\delta_T}\right)f_t(e)\,\mathrm{d}e \tag{S.1.1}$$

$$= T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\int K\left(\frac{e}{\delta_T}\right)f_t'(e)\,\mathrm{d}e - T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\left[K\left(\frac{e}{\delta_T}\right)f_t(e)\right]_{e=-\infty}^{e=\infty}. \tag{S.1.2}$$

As $\lim_{e\to\pm\infty}K(e)=0$ and f_t is bounded from above, the latter term is zero a.s. for all $T\in\mathbb{N}$. By transformation of variables, it further holds that

$$g_{t,T}^{e} = T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\int K\left(\frac{e}{\delta_T}\right)f_t'(e)\,\mathrm{d}e = T^{-1/2}\delta_T^{3/2}\boldsymbol{h}_t\int K(u)\,f_t'(\delta_T u)\,\mathrm{d}u. \tag{S.1.3}$$

A Taylor expansion of $f_t'(\delta_T u)$ around zero is given by

$$f_t'(\delta_T u) = f_t'(0) + (\delta_T u)f_t''(0) + \frac{(\delta_T u)^2}{2}f_t'''(\zeta\delta_T u), \tag{S.1.4}$$

for some $\zeta\in[0,1]$ and $f_t'(0)=0$ holds under the null hypothesis specified in (2.8). Consequently,

$$\sum_{t=1}^{T} g_{t,T}^{e} = T^{-1/2}\delta_T^{5/2}\sum_{t=1}^{T} f_t''(0)\boldsymbol{h}_t\int uK(u)\,\mathrm{d}u \tag{S.1.5}$$

$$+\frac{1}{2}T^{-1/2}\delta_T^{7/2}\sum_{t=1}^{T}\boldsymbol{h}_t\int u^2K(u)\,f_t'''(\zeta\delta_T u)\,\mathrm{d}u. \tag{S.1.6}$$

As $\int uK(u)\,\mathrm{d}u=0$ by assumption (A5), the first term is zero for all $T\in\mathbb{N}$. As $\sup_x f_t'''(x)\le c$ by

Assumption (A4) and $\int u^2 K(u)\,\mathrm{d}u \leq c < \infty$ by Assumption (A5), we obtain

$$\frac{1}{2}T^{-1/2}\delta_T^{7/2}\sum_{t=1}^{T}\boldsymbol{h}_t\int u^2 K(u)\, f_t'''(\zeta\delta_T u)\,\mathrm{d}u \leq \frac{1}{2}c^2\left(T\delta_T^7\right)^{1/2}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{h}_t \xrightarrow{P} 0, \tag{S.1.7}$$

as $T\delta_T^7 \to 0$ for $T \to \infty$ by Assumption (A6) and $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{h}_t \xrightarrow{P} \mathbb{E}[\boldsymbol{h}_t]$ by a law of large numbers for stationary and ergodic sequences. The result of the lemma follows. □

Lemma S.1.2. *Given Assumption 2.5 and under the null hypothesis in (2.8), it holds that* $\sum_{t=1}^{T}\mathrm{Var}(z_{t,T}) \to \bar{\omega}^2 = \lambda^\top\Omega_{\mathrm{Mode}}\lambda$.

Proof. We first observe that $\mathrm{Var}(z_{t,T}) = \mathbb{E}\left[\left(\lambda^\top\left(g_{t,T} - g_{t,T}^e\right)\right)^2\right]$ as $\mathbb{E}\left[\lambda^\top\left(g_{t,T} - g_{t,T}^e\right)\right] = 0$. Hence,

$$\mathrm{Var}[z_{t,T}] = \mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] - \mathbb{E}\left[\left(\lambda^\top g_{t,T}^e\right)^2\right], \tag{S.1.8}$$

as $\mathbb{E}\left[\left(\lambda^\top g_{t,T}^e\right)\cdot\left(\lambda^\top g_{t,T}\right)\right] = \mathbb{E}\left[\left(\lambda^\top g_{t,T}^e\right)\cdot\mathbb{E}_t\left[\lambda^\top g_{t,T}\right]\right] = \mathbb{E}\left[\left(\lambda^\top g_{t,T}^e\right)^2\right]$. For the first term in (S.1.8), we obtain

$$\mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] = \mathbb{E}\left[(T\delta_T)^{-1}(\lambda^\top\boldsymbol{h}_t)^2\mathbb{E}_t\left[K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2\right]\right] \tag{S.1.9}$$

$$= \mathbb{E}\left[(T\delta_T)^{-1}(\lambda^\top\boldsymbol{h}_t)^2\int K'\left(\frac{e}{\delta_T}\right)^2 f_t(e)\,\mathrm{d}e\right] \tag{S.1.10}$$

$$= \frac{1}{T}\,\mathbb{E}\left[(\lambda^\top\boldsymbol{h}_t)^2\int K'(u)^2\, f_t(\delta_T u)\,\mathrm{d}u\right]. \tag{S.1.11}$$

As $\delta_T \to 0$ when $T \to \infty$, we find

$$\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] \to \mathbb{E}\left[(\lambda^\top\boldsymbol{h}_t)^2 f_t(0)\right]\int K'(u)^2\,\mathrm{d}u = \lambda^\top\mathbb{E}\left[f_t(0)\boldsymbol{h}_t\boldsymbol{h}_t^\top\right]\lambda\int K'(u)^2\,\mathrm{d}u. \tag{S.1.12}$$

For the second term in (S.1.8), inserting the equality in (S.1.3) yields

$$\left(\lambda^\top g_{t,T}^e\right)^2 = \left(\delta_T^{3/2}T^{-1/2}(\lambda^\top\boldsymbol{h}_t)\int K'(u)\, f_t'(\delta_T u)\mathrm{d}u\right)^2 \leq \delta_T^3 T^{-1}||\lambda||^2||\boldsymbol{h}_t||^2\left|\int K'(u)\, f_t'(u\delta_T)\mathrm{d}u\right|^2.$$

As $\sup_x |f_t'(x)| \leq c$ and $\left|\int K'(u)\,\mathrm{d}u\right| \leq c$ by assumption, it holds that

$$\sum_{t=1}^{T} \mathbb{E}\left[\left(\lambda^\top g_{t,T}^e\right)^2\right] \leq \delta_T^3 c^2 ||\lambda||^2 \left(\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[||\boldsymbol{h}_t||^2\right]\right) \to 0, \tag{S.1.13}$$

as $\delta_T^3 \to 0$ as $T \to \infty$. The result of the lemma follows by combining (S.1.12) and (S.1.13). □

Lemma S.1.3. *Given Assumption 2.5 and under the null hypothesis in (2.8), it holds that* $\sum_{t=1}^{T} z_{t,T}^2 \xrightarrow{P} \bar{\omega}^2 = \lambda^\top \Omega_{\text{Mode}} \lambda$.

Proof. We define

$$h_{1,T} := \sum_{t=1}^{T} \left(z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right) \qquad \text{and} \qquad h_{2,T} := \sum_{t=1}^{T} \mathbb{E}_t\left[z_{t,T}^2\right] - \bar{\omega}^2, \tag{S.1.14}$$

such that $\sum_{t=1}^{T} z_{t,T}^2 - \bar{\omega}^2 = h_{1,T} + h_{2,T}$. We first show that $h_{1,T} \xrightarrow{L_p} 0$ for some $1 < p < 2$ sufficiently small enough and thus $h_{1,T} \xrightarrow{P} 0$. For this, first notice that $z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]$ is a $\mathcal{F}_{t+1}$-MDS by definition. Thus, we can apply the von Bahr and Esseen (1965)-inequality for some $p \in (1,2)$ for MDS (in the first line) in order to conclude that

$$\mathbb{E}\left[|h_{1,T}|^p\right] = \mathbb{E}\left[\left|\sum_{t=1}^{T} z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right] \leq 2\sum_{t=1}^{T} \mathbb{E}\left[\left|z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right] \tag{S.1.15}$$

$$\leq 2\sum_{t=1}^{T} 2^{p-1}\left(\mathbb{E}\left[\left|z_{t,T}^2\right|^p\right] + \mathbb{E}\left[\left|\mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right]\right) \leq 2^{p+1}\sum_{t=1}^{T} \mathbb{E}\left[\left|z_{t,T}\right|^{2p}\right], \tag{S.1.16}$$

where we use in the second inequality that $|a+b|^p \leq 2^{p-1}\left(|a|^p + |b|^p\right)$ for any $a, b \in \mathbb{R}$. Using the same inequality again yields

$$\mathbb{E}\left[\left|z_{t,T}\right|^{2p}\right] = \mathbb{E}\left[\left|\lambda^\top g_{t,T} - \lambda^\top g_{t,T}^e\right|^{2p}\right] \leq 2^{2p-1}\left(\mathbb{E}\left[\left|\lambda^\top g_{t,T}\right|^{2p}\right] + \mathbb{E}\left[\left|\lambda^\top g_{t,T}^e\right|^{2p}\right]\right). \tag{S.1.17}$$

Thus,

$$\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top g_{t,T}\right|^{2p}\right] = (T\delta_T)^{1-p}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p} \int \left|K'(u)\right|^{2p} f_t(\delta_T u)\mathrm{d}u\right] \to 0, \tag{S.1.18}$$

as $(T\delta_T)^{1-p} \to 0$ for all $p \in (1,2)$, $\mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p}\right] < \infty$ for $p > 1$ sufficiently small such that $2p \leq 2+\delta$ (for the $\delta > 0$ from Assumption (A2)), and $\int |K'(u)|^{2p} f_t(\delta_T u)\mathrm{d}u \leq cc^{2p-1}\int |K'(u)|\,\mathrm{d}u < \infty$ as $\int |K'(u)|\,\mathrm{d}u < \infty$, $\sup_u |K'(u)| \leq c$ and $\sup_x f_t(x) \leq c$ a.s. by assumption. Similarly, we obtain

$$\sum_{t=1}^{T}\mathbb{E}\left[\left|\lambda^\top g_{t,T}^e\right|^{2p}\right] = \delta_T^{2p-1}(T\delta_T)^{1-p}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p}\left|\int K'(u)\, f_t(\delta_T u)\mathrm{d}u\right|^{2p}\right] \to 0, \tag{S.1.19}$$

which shows that $h_{1,T} \xrightarrow{L_p} 0$ for some $p > 1$ sufficiently small which implies that $h_{1,T} \xrightarrow{P} 0$.

We continue by showing that $h_{2,T} \xrightarrow{P} 0$. Using the same argument as in (S.1.8), we split

$$h_{2,T} = \sum_{t=1}^{T}\mathbb{E}_t\left[z_{t,T}^2\right] - \bar{\omega}^2 = \sum_{t=1}^{T}\mathbb{E}_t\left[(\lambda^\top g_{t,T})^2\right] - \sum_{t=1}^{T}(\lambda^\top g_{t,T}^e)^2 - \bar{\omega}^2. \tag{S.1.20}$$

Applying a transformation of variables yields

$$\sum_{t=1}^{T}(\lambda^\top g_{t,T}^e)^2 = \delta_T\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\left(\int K'(u)f_t(\delta_T u)\mathrm{d}u\right)^2 \tag{S.1.21}$$

$$\leq \delta_T\left(\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\right)\left(c\int |K'(u)|\mathrm{d}u\right)^2 \xrightarrow{P} 0, \tag{S.1.22}$$

as $\delta_T \to 0$, $\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2 = \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2\right] + o_P(1)$ and $\left(\int |K'(u)|\mathrm{d}u\right)^2 \leq \int |K'(u)|^2\mathrm{d}u < \infty$ by assumption. Furthermore,

$$\sum_{t=1}^{T}\mathbb{E}_t\left[\left(\lambda^\top g_{t,T}\right)^2\right] = (T\delta_T)^{-1}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\mathbb{E}_t\left[K'\left(\frac{\varepsilon_t}{\delta_T}\right)^2\right] \tag{S.1.23}$$

$$= \left(\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\right)\int K'(u)^2 f_t(\delta_T u)\mathrm{d}u \xrightarrow{P} \mathbb{E}\left[f_t(0)(\lambda^\top \boldsymbol{h}_t)^2\right]\int K'(u)^2\mathrm{d}u = \bar{\omega}^2, \tag{S.1.24}$$

as for all $u \in \mathbb{R}$, $\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2 f_t(\delta_T u) \xrightarrow{P} \mathbb{E}\left[f_t(0)(\lambda^\top \boldsymbol{h}_t)^2\right]$. Thus, we find $h_{2,T} \xrightarrow{P} 0$ and consequently $\sum_{t=1}^{T} z_{t,T}^2 - \bar{\omega}^2 \xrightarrow{P} 0$, which concludes this proof. □

Lemma S.1.4. *Given Assumption 2.5 and the null hypothesis in (2.8), it holds that* $\max_{1\leq t\leq T}|h_{t,T}| \xrightarrow{P} 0$.

Proof. Let $\zeta > 0$ and $\delta > 0$ (sufficiently small such that $\mathbb{E}\left[||\boldsymbol{h}_t||^{2+\delta}\right] < \infty$). Then,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq t\leq T}|h_{t,T}| > \zeta\right) &= \mathbb{P}\left(\max_{1\leq t\leq T}|h_{t,T}|^{2+\delta} > \zeta^{2+\delta}\right) \leq \sum_{t=1}^{T}\mathbb{P}\left(|h_{t,T}|^{2+\delta} > \zeta^{2+\delta}\right) \\ &\leq \zeta^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[|h_{t,T}|^{2+\delta}\right] = \zeta^{-2-\delta}\bar{\omega}_T^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[|z_{t,T}|^{2+\delta}\right],\end{aligned} \tag{S.1.25}$$

by Markov's inequality. Employing the same steps as in the proof of Lemma S.1.3 following Equation (S.1.17) and replacing the exponent "$2p$" by "$2+\delta$" yields that $\sum_{t=1}^{T}\mathbb{E}\left[|z_{t,T}|^{2+\delta}\right] \to 0$. As $\bar{\omega}_T \to \bar{\omega}^2 > 0$, this directly implies that $\mathbb{P}\left(\max_{1\leq t\leq T}|h_{t,T}| > \zeta\right) \to 0$.

□

Lemma S.1.5. *Given Assumption 2.5, Assumption 3.1 and Assumption 3.2, for all $\lambda \in \mathbb{R}^k$ such that $||\lambda||_2 = 1$, it holds that $\sum_{t=1}^{T}\text{Var}\left(\phi_{t,T}^*(\theta_0)\lambda\right) \to \sigma^2$.*

Proof. As $\phi_{t,T}^*$ is a $\mathcal{F}_{t+1}$-MDS, it holds that $\mathbb{E}\left[\phi_{t,T}^*(\theta_0)\lambda\right] = 0$ and thus, $\text{Var}\left(\phi_{t,T}^*(\theta_0)\lambda\right) = \mathbb{E}\left[\left(\phi_{t,T}^*(\theta_0)\lambda\right)^2\right]$. We further find

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\left(\phi_{t,T}^*(\theta_0)\lambda\right)^2\right] &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}^2\left(\lambda^\top \boldsymbol{W}_{\text{Mean}}\boldsymbol{h}_t\right)^2\varepsilon_t^2\right] \\ &+ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{20}^2\left(\lambda^\top \boldsymbol{W}_{\text{Med}}\boldsymbol{h}_t\right)^2\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)^2\right] \\ &+ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{30}^2\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2\right] \\ &+ \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{20}\left(\lambda^\top \boldsymbol{W}_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top \boldsymbol{W}_{\text{Med}}\boldsymbol{h}_t\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)\right] \\ &+ \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top \boldsymbol{W}_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)\varepsilon_t\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right] \\ &+ \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{20}\theta_{30}\left(\lambda^\top \boldsymbol{W}_{\text{Med}}\boldsymbol{h}_t\right)\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right] \\ &+ \sum_{t=1}^{T}\mathbb{E}\left[(u_{t,T}(\theta_0)\lambda)^2\right] - 2\sum_{t=1}^{T}\mathbb{E}\left[\left(u_{t,T}(\theta_0)\lambda\right)\left(\tilde{\phi}_{t,T}(\theta_0)\lambda\right)\right].\end{aligned} \tag{S.1.26}$$

For the last two terms, we have $\sum_{t=1}^{T}\mathbb{E}\left[\left(u_{t,T}(\theta_0)\lambda\right)^2\right]\to 0$ and $\sum_{t=1}^{T}\mathbb{E}\left[\left(u_{t,T}(\theta_0)\lambda\right)\left(\tilde{\phi}_{t,T}(\theta_0)\lambda\right)\right]\to 0$ by assumption. For the fifth term,

$$\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top \boldsymbol{W}_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)\delta_T^{-1/2}\mathbb{E}_t\left[\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\right] \tag{S.1.27}$$

$$= -\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top \boldsymbol{W}_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)\delta_T^{3/2}\int uK'(u)\,f_t(\delta_T u)\mathrm{d}u\right]\to 0, \tag{S.1.28}$$

as $\delta_T^{3/2}\to 0$, $\int uK'(u)\,\mathrm{d}u<\infty$ and the respective moments are finite. The sixth term converges to zero by a similar argument by bounding $\left|\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right|\leq 1$. For the third term, it holds that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{30}^2\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2\right] &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{30}^2\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)^2\int K'(u)^2 f_t(\delta_T u)\mathrm{d}u\right]\\ &\to \mathbb{E}\left[\theta_{30}^2\left(\lambda^\top \boldsymbol{W}_{\text{Mode}}\boldsymbol{h}_t\right)^2 f_t(0)\int K'(u)^2\,\mathrm{d}u\right].\end{aligned}$$

The remaining first, second and fourth terms obviously converge to the equivalent quantities of σ^2 by employing a standard CLT, which concludes the proof of this lemma. □

Lemma S.1.6. *Given Assumption 2.5, Assumption 3.1 and Assumption 3.2, for all $\lambda\in\mathbb{R}^k$ such that $||\lambda||_2=1$, it holds that $\sum_{t=1}^{T}\left(\phi_{t,T}^*(\theta_0)\lambda\right)^2\xrightarrow{P}\sigma^2$.*

Proof. We apply the same factorization as in (S.1.26) (however without the expectation operator). By applying a law of large numbers for stationary and ergodic sequences (Theorem 3.34 in White (2001)), we obtain that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left(\theta_{10}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right)\varepsilon_t\right)^2\xrightarrow{P}\mathbb{E}\left[\left(\theta_{10}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right)\varepsilon_t\right)^2\right],\\ &\frac{1}{T}\sum_{t=1}^{T}\left(\theta_{20}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\right)^2\xrightarrow{P}\mathbb{E}\left[\left(\theta_{20}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\right)^2\right],\\ &\frac{2}{T}\sum_{t=1}^{T}\left(\theta_{10}\theta_{20}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\right)^2\\ &\quad\xrightarrow{P}2\mathbb{E}\left[\left(\theta_{10}\theta_{20}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\right)^2\right].\end{aligned}$$

Furthermore, a slight modification of Lemma S.1.3 (multiplying with $\theta_{30}^2 \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right)^2$ instead of $\left(\boldsymbol{h}_t^\top \lambda\right)^2$) yields that

$$\frac{1}{T}\sum_{t=1}^{T} \theta_{30}^2 \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right)^2 \delta_T^{-1} K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2 \xrightarrow{P} \mathbb{E}\left[\theta_{30}^2 \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right)^2 f_t(0) \int K'(u)^2 \,\mathrm{d}u\right]. \quad \text{(S.1.29)}$$

We now show that the remaining four terms vanish asymptotically (in probability). For the mixed mean/mode term, we apply a similar addition of a zero (adding and subtracting $\mathbb{E}_t[\dots]$) as in the proof of Lemma S.1.3. For this, we first note that

$$\frac{2}{T}\sum_{t=1}^{T} \theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{-1/2} \mathbb{E}_t\left[\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right] \quad \text{(S.1.30)}$$

$$= -\frac{2}{T}\sum_{t=1}^{T} \theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{3/2} \int u K'(u)\, f_t(\delta_T u) \mathrm{d}u \xrightarrow{P} 0, \quad \text{(S.1.31)}$$

as $\delta_T^{3/2} \to 0$. In the following, we further show that

$$\frac{2}{T}\sum_{t=1}^{T} \theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{-1/2} \left\{\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right) - \mathbb{E}_t\left[\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\right\} \xrightarrow{L_p} 0,$$

for any $p \in (1,2)$ small enough. As in the proof of Lemma S.1.3, we apply the von Bahr and Esseen (1965) inequality and Minkowski's inequality in order to conclude that

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{2}{T}\sum_{t=1}^{T} \theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{-1/2} \left\{\varepsilon_t K'\left(\frac{\varepsilon_t}{\delta_T}\right) - \mathbb{E}_t\left[\varepsilon_t K'\left(\frac{\varepsilon_t}{\delta_T}\right)\right]\right\}\right|^p\right] \\
&\leq \frac{2^{p+2}}{T^p}\sum_{t=1}^{T} \left\{\mathbb{E}\left[\left|\theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{-1/2} \varepsilon_t K'\left(\frac{\varepsilon_t}{\delta_T}\right)\right|^p\right]\right. \\
&\qquad \left. + \mathbb{E}\left[\left|\theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right) \delta_T^{-1/2} \mathbb{E}_t\left[\varepsilon_t K'\left(\frac{\varepsilon_t}{\delta_T}\right)\right]\right|^p\right]\right\},
\end{aligned}$$

where the first term is bounded from above by

$$\leq \frac{2^{p+2}}{T^p}\sum_{t=1}^{T} \mathbb{E}\left[\left|\theta_{10}\theta_{30} \left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}} \lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda\right)\right|^p \delta_T^{-p/2} \int \left|e K'\left(\frac{e}{\delta_T}\right)\right|^p f_t(e) \mathrm{d}e\right]$$

$$= 2^{p+2}\delta_T^{1+p/2}T^{1-p}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\theta_{10}\theta_{30}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}}\lambda\right)\right|^p \int |uK'(u)|^p f_t(\delta_T u)\mathrm{d}u\right] \to 0,$$

as $\delta_T^{1+p/2} \to 0$, $T^{1-p} \to 0$ for any $p > 1$ and as the respective moments are bounded by assumption. Similar arguments also yield that the second term converges to zero (compare to (S.1.19)). Applying the same line of reasoning for the mixed median/mode terms shows that

$$\frac{2}{T}\sum_{t=1}^{T}\theta_{20}\theta_{30}\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right)\left(\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}}\lambda\right)\delta_T^{-1/2}\mathbb{E}_t\left[\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right] \xrightarrow{P} 0. \quad \text{(S.1.32)}$$

For the fourth and last term, $\sum_{t=1}^{T}\left(u_{t,T}(\theta_0)\lambda\right)^2 \xrightarrow{P} 0$ and $\sum_{t=1}^{T}\left(u_{t,T}(\theta_0)\lambda\right)\left(\tilde{\phi}_{t,T}(\theta_0)\lambda\right) \xrightarrow{P} 0$ by assumption, which concludes this proof. □

Lemma S.1.7. *Given Assumption 2.5, Assumption 3.1 and Assumption 3.2, for all $\lambda \in \mathbb{R}^k$ such that $||\lambda||_2 = 1$, it holds that $\max_{1\le t\le T}\left|\sigma^{-1}\phi^*_{t,T}(\theta_0)\lambda\right| \xrightarrow{P} 0$.*

Proof. Let $\zeta > 0$ and $\delta > 0$ (sufficiently small such that $\mathbb{E}\left[||\boldsymbol{h}_t||^{2+\delta}\right] < \infty$ holds). Then, as in (S.1.25) in the proof of Lemma S.1.4, we get that

$$\mathbb{P}\left(\max_{1\le t\le T}\left|\sigma_T^{-1}\phi^*_{t,T}(\theta_0)\lambda\right| > \zeta\right) \le \zeta^{-2-\delta}\sigma_T^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[\left|\phi^*_{t,T}(\theta_0)\lambda\right|^{2+\delta}\right], \quad \text{(S.1.33)}$$

by Markov's inequality. Furthermore, we get that

$$4^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[\left|\phi^*_{t,T}(\theta_0)\lambda\right|^{2+\delta}\right] \le \sum_{t=1}^{T}\mathbb{E}\left[\left|u_{t,T}(\theta_0)\right|^{2+\delta}\right] \quad \text{(S.1.34)}$$

$$+\theta_{10}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mean}}\lambda\right|^{2+\delta}|\varepsilon_t|^{2+\delta}\right] \quad \text{(S.1.35)}$$

$$+\theta_{20}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Med}}\lambda\right|^{2+\delta}\left|\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right|^{2+\delta}\right] \quad \text{(S.1.36)}$$

$$+\theta_{30}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}}\lambda\right|^{2+\delta}\delta_T^{-\frac{2+\delta}{2}}\left|K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right|^{2+\delta}\right]. \quad \text{(S.1.37)}$$

The first term converges to zero by Assumption 3.1. The second and third term converge to zero as $T^{-\frac{\delta}{2}} \to 0$ and the respective moments are bounded by assumption. For the last term, we obtain

convergence equivalently to the proof of Lemma S.1.4,

$$\theta_{30}^{2+\delta} T^{-\frac{\delta}{2}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\left| \boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda \right|^{2+\delta} \delta_T^{-\frac{2+\delta}{2}} \left| K'\left(\frac{-\varepsilon_t}{\delta_T} \right) \right|^{2+\delta} \right] \tag{S.1.38}$$

$$\leq \theta_{30}^{2+\delta} (T\delta_T)^{-\frac{\delta}{2}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\left| \boldsymbol{h}_t^\top \boldsymbol{W}_{\text{Mode}} \lambda \right|^{2+\delta} \int \left| K'(u) \right|^{2+\delta} f_t(\delta_T u) \mathrm{d}u \right], \tag{S.1.39}$$

which converges to zero as $(T\delta_T)^{-\frac{\delta}{2}} \to 0$ and the respective moments are bounded by assumption.

□

Proof of Theorem 2.7. Let $\lambda \in \mathbb{R}^k$, $||\lambda||_2 = 1$ be a fixed and deterministic vector. Then,

$$\begin{aligned}
&\lambda^\top \widehat{\Omega}_{T,\text{Mode}} \lambda - \lambda^\top \Omega_{\text{Mode}} \lambda \\
&= \frac{1}{T} \sum_{t=1}^{T} \delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_t \left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right] \\
&+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_t \left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right] - \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 f_t(0) \int K'(u)^2 \mathrm{d}u \right].
\end{aligned} \tag{S.1.40}$$

We start by showing that the last line in (S.1.40) is $o_P(1)$,

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_t \left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right] = \frac{1}{T} \sum_{t=1}^{T} (\lambda^\top \boldsymbol{h}_t)^2 \delta_T^{-1} \int K'\left(\frac{e}{\delta_T} \right)^2 f_t(e) \mathrm{d}e \tag{S.1.41}$$

$$= \frac{1}{T} \sum_{t=1}^{T} (\lambda^\top \boldsymbol{h}_t)^2 \int K'(u)^2 f_t(\delta_T u) \mathrm{d}u \xrightarrow{P} \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 f_t(0) \int K'(u)^2 \mathrm{d}u \right], \tag{S.1.42}$$

as $f_t(\delta_T u) \to f_t(0) \leq c$, and by further applying a weak law of large numbers for stationary and ergodic data as $\mathbb{E}\left[||\boldsymbol{h}_t||^{2+\delta} \right] < \infty$.

We further show that the penultimate line in (S.1.40) converges to zero in L_p (p-th mean) for some $p > 1$ small enough. By applying the von Bahr and Esseen (1965) inequality for MDS, we get

$$\begin{aligned}
&\mathbb{E}\left[\left| \frac{1}{T} \sum_{t=1}^{T} \delta_T^{-1} K'\left(\frac{\varepsilon_t}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_t \left[\delta_T^{-1} K'\left(\frac{\varepsilon_t}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right] \right|^p \right] \\
&\leq 2T^{-p} \sum_{t=1}^{T} \mathbb{E}\left[\left| \delta_T^{-1} K'\left(\frac{\varepsilon_t}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right|^p \right] + 2T^{-p} \sum_{t=1}^{T} \mathbb{E}\left[\left| \mathbb{E}_t \left[\delta_T^{-1} K'\left(\frac{\varepsilon_t}{\delta_T} \right)^2 (\lambda^\top \boldsymbol{h}_t)^2 \right] \right|^p \right].
\end{aligned}$$

For the first term, we get that

$$T^{-p}\sum_{t=1}^{T}\mathbb{E}\left[\left|\delta_T^{-1}K'\left(\frac{\varepsilon_t}{\delta_T}\right)^2(\lambda^\top \boldsymbol{h}_t)^2\right|^p\right] = (T\delta_T)^{-p}\sum_{t=1}^{T}\mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p}\int\left|K'\left(\frac{e}{\delta_T}\right)\right|^{2p} f_t(e)\mathrm{d}e\right]$$
$$= (T\delta_T)^{1-p}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p}\int\left|K'(u)\right|^2 f_t(\delta_T u)\mathrm{d}u\right] \to 0,$$

as $(T\delta_T)^{1-p}\to 0$ for any $p>1$, $\mathbb{E}\left[||\boldsymbol{h}_t||^{2p}\right]<\infty$ for $p>1$ small enough, the density f_t is bounded from above, and $\int|K'(u)|^2\mathrm{d}u<\infty$ by assumption. The second term converges by a similar argument as further detailed in (S.1.19) in the proof of Lemma S.1.3. As L_p convergence for any $p>1$ implies convergence in probability, the result of the theorem follows. □

Proof of Theorem 3.4. For notational simplicity, we show consistency of the covariance estimator by considering the bilinear forms $\lambda^\top\left(\frac{1}{T}\sum_{t=1}^{T}\phi_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top\right)\lambda$ and $\sigma^2 := \lambda^\top\Sigma(\theta_0)\lambda$, given in (A.20), for some arbitrary but fixed $\lambda\in\mathbb{R}^k$ such that $||\lambda||_2=1$. Then, we get that

$$\lambda^\top\left(\frac{1}{T}\sum_{t=1}^{T}\phi_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top\right)\lambda = \frac{1}{T}\sum_{t=1}^{T}\theta_{10}^2\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2 \tag{S.1.43}$$
$$+\theta_{20}^2\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Med}}\lambda\right)^2\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)^2+\theta_{30}^2\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mode}}\lambda\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2 \tag{S.1.44}$$
$$+2\theta_{10}\theta_{20}\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Med}}\lambda\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right) \tag{S.1.45}$$
$$+2\theta_{10}\theta_{30}\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mode}}\lambda\right)\varepsilon_t\ \delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right) \tag{S.1.46}$$
$$+2\theta_{20}\theta_{30}\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Med}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mode}}\lambda\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right). \tag{S.1.47}$$

We show convergence in probability for the individual matrix components for the first term,

$$\frac{1}{T}\sum_{t=1}^{T}\theta_{10}^2\left(\boldsymbol{h}_t^\top\widehat{\boldsymbol{W}}_{T,\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2 = \sum_{i,j,\iota,l}\widehat{\boldsymbol{W}}_{T,\mathrm{Mean},ij}\widehat{\boldsymbol{W}}_{T,\mathrm{Mean},\iota l}\frac{1}{T}\sum_{t=1}^{T}\theta_{10}^2\boldsymbol{h}_{t,i}\lambda_j\boldsymbol{h}_{t,\iota}\lambda_l\varepsilon_t^2 \tag{S.1.48}$$
$$\xrightarrow{P}\sum_{i,j,\iota,l}\boldsymbol{W}_{\mathrm{Mean},ij}\boldsymbol{W}_{\mathrm{Mean},\iota l}\,\mathbb{E}\left[\theta_{10}^2\boldsymbol{h}_{t,i}\lambda_j\boldsymbol{h}_{t,\iota}\lambda_l\varepsilon_t^2\right] = \mathbb{E}\left[\theta_{10}^2\left(\boldsymbol{h}_t^\top\boldsymbol{W}_{\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2\right]. \tag{S.1.49}$$

Convergence of the remaining terms follows analogously by considering the terms component-wisely and by applying similar arguments as in Lemma S.1.6. □

S.2 Kernel Choice

The asymptotic results presented in Section 2.3 rely on the chosen kernel K satisfying Assumption (A5). Besides the normalization $\int K(u)\mathrm{d}u = 1$ and boundedness assumptions, we impose the *first-order kernel* condition $\int uK(u)\mathrm{d}u = 0$ (and $\int u^2K(u)\mathrm{d}u > 0$ follows from the non-negativity of K). As discussed in Li and Racine (2006), higher-order kernels allow one to apply a Taylor expansion of higher order and can thereby obtain a faster rate of convergence, which could in theory be made arbitrarily close to $\sqrt{T}$, at the cost of stronger smoothness assumptions on the underlying density function. However, in our application of kernel functions to the generalized modal midpoint in Definition 2.3, we need to ensure that the limit of this quantity is well-defined and unique, and that the identification is *strict*. For this, we assume in Theorem 2.4 that the kernel function is log-concave which is automatically violated for higher-order kernels. Consequently, we do not consider higher-order kernels in this work.

It is well-known in the literature on nonparametric statistics that kernels with bounded support can be more efficient. However, *strict* identifiability of the generalized modal midpoint only holds for kernel functions with unbounded support, which motivates our usage of unbounded kernel functions such as the Gaussian kernel. Figure S.3 further illustrates that the test power does not increase by employing a biweight kernel, which has bounded support.

S.3 Bandwidth Choice

We follow the rule-of-thumb proposed by Kemp and Silva (2012) and Kemp et al. (2020) in setting the bandwidth parameter, with one modification to deal with skewness. Specifically, as discussed in Section 2.3, in order to obtain an optimal convergence for our nonparametric test (for first-order kernels), we choose $\delta_T \approx T^{-1/7}$. Following Kemp and Silva (2012), we choose δ_T proportional to $T^{-0.143}$, which is almost $T^{-1/7}$:

$$\delta_T = k_1 \cdot k_2 \cdot T^{-0.143}. \tag{S.3.1}$$

As in Kemp and Silva (2012) and Kemp et al. (2020), we choose k_1 proportional to the median

absolute deviation of the forecast error, a robust measure for the variation in the data,

$$k_1 = 2.4 \times \widehat{\text{Med}}_t\big(\big|(X_t - Y_{t+1}) - \widehat{\text{Med}}_s(X_s - Y_{s+1})\big|\big). \tag{S.3.2}$$

The choice of the bandwidth parameter should be proportional to the scale of the underlying data such that test results are robust to linear re-scaling. Using preliminary simulations, we found better finite-sample results when this measure is scaled by 2.4.

Following early simulation analyses, we introduce a second constant, k_2, to adjust the bandwidth for the skewness of the forecast error, measured by the absolute value of Pearson's second skewness coefficient, $\hat{\gamma}$.

$$k_2 = \exp(-3\,|\hat{\gamma}|), \qquad \text{where } \hat{\gamma} = \frac{3\left(\frac{1}{T}\sum_t (X_t - Y_{t+1}) - \widehat{\text{Med}}_t[X_t - Y_{t+1}]\right)}{\hat{\sigma}(X_t - Y_{t+1})}. \tag{S.3.3}$$

For symmetric distributions, $k_2 = 1$ and this term vanishes from the bandwidth formula. For such distributions, and assuming a symmetric kernel as in our empirical work, the generalized modal midpoint equals the mode and employing a larger bandwidth increases efficiency. As skewness increases in magnitude the distance between the mode and the generalized modal midpoint increases for a fixed bandwidth, and to ensure satisfactory finite-sample properties a smaller bandwidth is needed. Our simple expression for k_2 achieves this.

S.4 Power study for the mode forecast rationality test

To analyze the power of the mode forecast rationality test introduced in Section 2.3 of the main paper we use the same DGPs as in Section 4.1 and consider two forms of sub-optimal forecasts:

(a) **Bias**: $\tilde{X}_t = X_t + \kappa\sigma_X$, where $\sigma_X = \sqrt{\text{Var}(X_t)}$ for $\kappa \in (-1, 1)$, and

(b) **Noise**: $\tilde{X}_t = X_t + \mathcal{N}(0, \kappa\sigma_X^2)$, where $\sigma_X = \sqrt{\text{Var}(X_t)}$, for $\kappa \in (0, 1)$.

The first type of misspecification introduces deterministic bias, where the degree of misspecification depends on the misspecification parameter κ. We standardize the bias using the unconditional standard deviation of the optimal forecasts, $\sqrt{\text{Var}(X_t)}$. The second type of misspecification introduces

independent noise, and the magnitude of the noise is regulated through the parameter κ: for $\kappa = 1$, the signal-to-noise ratio is one, as the standard deviation of the signal equals the standard deviation of the independent noise, and as κ shrinks to zero the noise vanishes.

Figure S.1 presents power plots for the "biased" forecasts and Figure S.2 presents power plots for the "noisy" forecasts. In each of the plots, we plot the rejection rate against the degree of misspecification κ. For all plots, we use the instrument choice $(1, X_t)$, a Gaussian kernel, and a nominal level of 5%. Notice that for $\kappa = 0$ the figures reveal the empirical test size.

Figure S.1 and Figure S.2 reveal that the proposed mode rationality test exhibits, as expected, increasing power for an increasing degree of misspecification. Also as expected, larger sample sizes lead to tests with greater power, although even the two smaller sample sizes exhibit reasonable power, particular in the case of biased forecasts. The figures also reveal that increasing degree of skewness yields to a slight loss of power. This is driven through the bandwidth choice, where larger values of the (empirical) skewness result in a smaller bandwidth, and consequently a lower test power (analogous to the bias-variance trade-off in the nonparametric estimation literature).

Results corresponding to the "biased" forecast case when using a biweight kernel are presented in Figure S.3. That figure reveals that the finite-sample size and power are very similar to those for the Gaussian kernel.

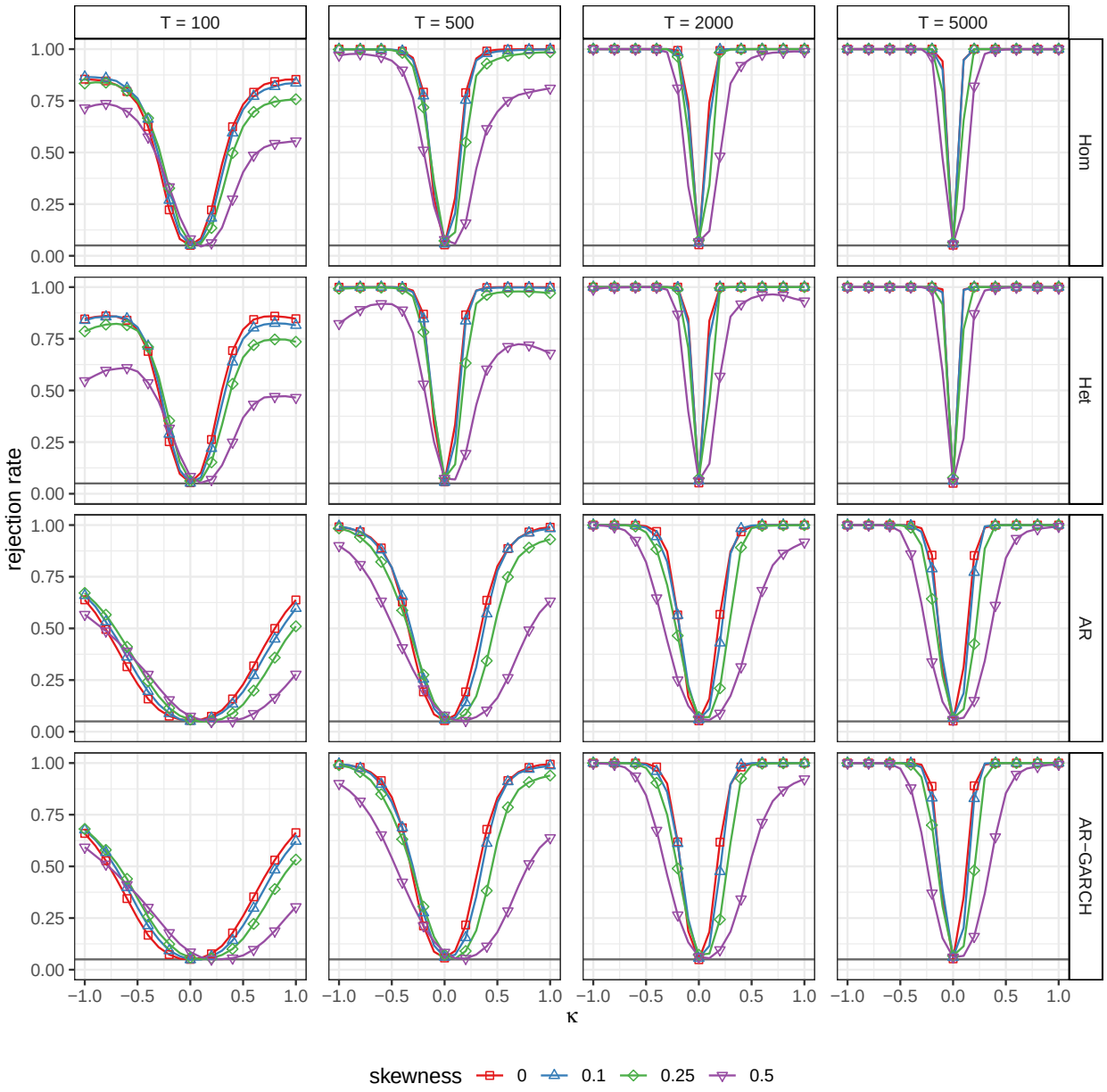


Figure S.1: Test power for the bias simulation setup. This figure plots the empirical rejection frequencies against the degrees of misspecification κ for different sample sizes in the vertical panels and for the four DGPs in the horizontal panels. The misspecification follows the "bias" setup and we utilize the instrument vector $(1, X_t)$ and a nominal significance level of 5%.

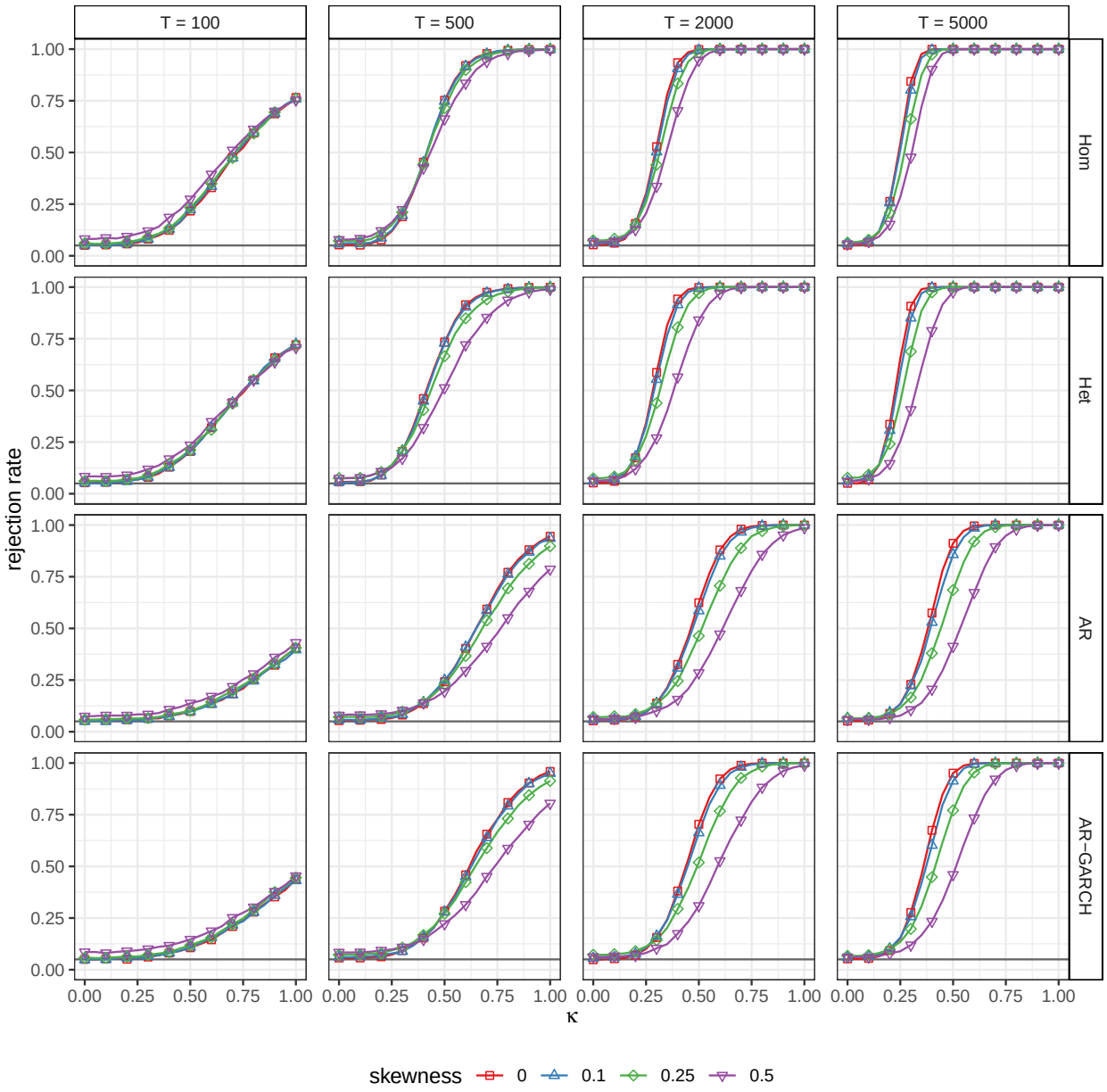


Figure S.2: Test power for the noise simulation setup. This figure plots the empirical rejection frequencies against the degrees of misspecification κ for different sample sizes in the vertical panels and for the four DGPs in the horizontal panels. The misspecification follows the "noise" setup and we utilize the instrument vector $(1, X_t)$ and a nominal significance level of 5%.

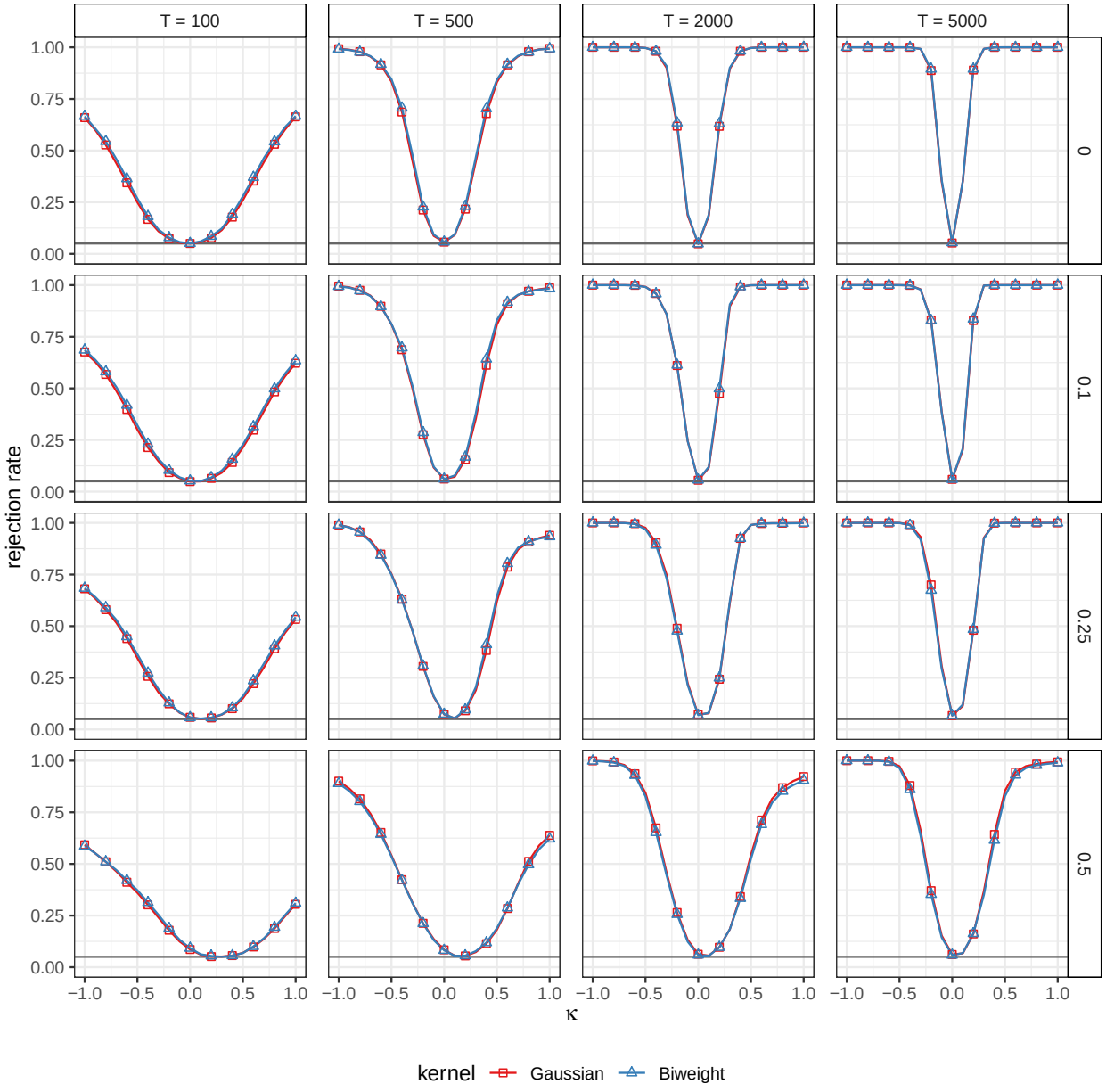


Figure S.3: Test power for different kernel functions. This figure plots the empirical rejection frequencies for the Gaussian and the biweight kernels against the degrees of misspecification κ for different sample sizes in the vertical panels and for four skewness levels in the horizontal panels. We simulate data from the AR-GARCH process, the misspecification follows the *bias* setup and we utilize the instrument vector $(1, X_t)$ and a nominal significance level of 5%.

S.5 Additional Plots and Tables

Table S.1: Empirical size of the mode rationality test: 1% significance level

	Instrument set 1				Instrument set 2				Instrument set 3			
Skewness	0	0.1	0.25	0.5	0	0.1	0.25	0.5	0	0.1	0.25	0.5
Sample size												
	Panel A: Homoskedastic iid data											
100	0.9	1.1	1.2	2.4	1.0	1.2	1.3	1.9	1.2	1.2	1.2	1.8
500	1.0	1.2	2.0	2.8	1.0	1.2	1.6	2.2	1.0	1.2	1.7	1.8
2000	1.2	1.5	2.2	1.9	1.0	1.3	1.6	1.5	0.9	1.2	1.6	1.3
5000	0.9	1.6	1.8	1.4	1.0	1.2	1.7	1.3	1.1	1.0	1.5	1.2
	Panel B: Heteroskedastic data											
100	1.0	1.1	1.5	2.6	1.2	1.0	1.4	2.3	1.3	1.1	1.2	2.0
500	1.4	1.2	2.3	2.2	1.1	1.1	1.9	1.8	1.0	1.2	1.6	1.7
2000	1.0	1.6	2.5	1.8	1.1	1.4	1.9	1.6	1.0	1.3	1.8	1.4
5000	1.0	1.6	2.7	1.4	0.9	1.3	2.0	1.2	1.0	1.1	1.7	1.2
	Panel C: Autoregressive data											
100	0.9	0.8	1.3	2.6	1.2	0.8	1.1	1.7	1.4	1.1	1.1	1.6
500	1.1	1.2	2.0	3.1	1.1	1.2	1.5	2.4	1.1	1.1	1.4	2.1
2000	1.1	1.2	2.2	1.8	1.2	1.1	1.8	1.5	1.0	1.1	1.6	1.4
5000	1.0	1.5	1.7	1.6	1.0	1.5	1.6	1.2	1.1	1.4	1.6	1.4
	Panel D: AR-GARCH data											
100	0.8	0.8	1.1	2.6	0.9	1.1	1.1	2.0	1.0	1.1	1.2	1.9
500	1.0	1.4	2.1	2.8	1.2	1.3	1.6	2.4	1.1	1.2	1.5	2.2
2000	1.0	1.4	2.3	1.8	1.0	1.1	1.8	1.4	0.9	1.2	1.7	1.3
5000	1.1	1.6	2.0	1.5	1.0	1.4	1.7	1.3	0.9	1.2	1.5	1.1

Notes: This table presents the empirical size of the mode rationality test for a Gaussian kernel, varying sample sizes, varying levels of skewness in the residual distribution and different instrument choices for a nominal significance level of 1%.

Table S.2: Empirical size of the mode rationality test: 10% significance level

	Instrument set 1				Instrument set 2				Instrument set 3			
Skewness	0	0.1	0.25	0.5	0	0.1	0.25	0.5	0	0.1	0.25	0.5
Sample size												
	Panel A: Homoskedastic iid data											
100	9.5	9.8	11.3	15.3	10.4	11.0	11.4	14.7	10.5	10.9	12.2	14.1
500	10.9	12.0	14.3	14.7	10.4	10.9	13.3	13.8	10.8	10.9	13.2	13.2
2000	11.1	12.6	14.7	12.7	10.6	11.8	13.2	12.5	10.3	11.7	12.8	12.0
5000	10.2	11.4	12.8	11.6	10.3	11.0	11.9	11.2	10.0	10.8	11.9	10.6
	Panel B: Heteroskedastic data											
100	10.0	10.4	12.1	15.7	10.6	10.6	11.9	14.8	11.1	10.9	11.9	14.2
500	11.4	11.6	14.4	14.8	11.5	11.1	13.6	13.1	10.9	11.1	13.4	12.5
2000	10.1	12.4	15.1	13.0	10.3	12.1	13.7	11.6	10.2	11.7	12.9	11.5
5000	10.3	13.0	15.4	12.0	10.1	11.9	13.7	11.8	10.1	11.4	12.8	11.1
	Panel C: Autoregressive data											
100	9.6	10.2	11.4	14.7	10.8	10.6	11.5	13.6	11.3	10.9	11.6	13.2
500	11.2	12.0	14.2	15.1	10.7	11.2	13.1	14.1	10.8	10.9	12.5	13.7
2000	10.5	12.3	13.9	12.2	10.4	11.1	12.4	11.7	10.3	11.3	12.2	11.6
5000	10.2	12.1	13.3	11.9	10.6	11.8	12.0	11.6	10.4	11.2	11.9	11.1
	Panel D: AR-GARCH data											
100	9.6	9.7	11.1	16.6	10.2	10.5	11.4	15.2	10.7	10.6	11.8	14.8
500	11.2	12.2	14.6	15.3	11.1	11.6	13.7	14.5	11.2	10.9	13.0	13.9
2000	10.5	12.2	14.0	12.3	10.3	11.1	13.0	11.8	10.0	10.7	12.4	11.4
5000	10.0	12.2	13.9	12.0	10.5	11.4	12.5	11.5	10.3	11.3	12.0	11.2

Notes: This table presents the empirical size of the mode rationality test for a Gaussian kernel, varying sample sizes, varying levels of skewness in the residual distribution and different instrument choices for a nominal significance level of 10%.

Table S.3: Empirical coverage of the confidence sets for central tendency: Cross-sectional data

Centrality measure	θ_{Mean}	θ_{Med}	θ_{Mode}	Symmetric data 100	500	2000	5000	Skewed data 100	500	2000	5000
				Panel A: Homoskedastic iid data							
Mean	1.00	0.00	0.00	89.3	90.0	89.8	89.2	89.4	90.2	89.6	90.2
Mode	0.00	0.00	1.00	90.1	89.1	89.8	90.1	85.7	86.0	87.6	88.8
Median	0.28	0.00	0.72	90.0	88.9	89.4	89.7	91.3	93.0	93.3	92.2
Median	0.15	0.50	0.35	89.5	89.0	89.6	89.6	90.3	91.9	91.4	91.2
Median	0.00	1.00	0.00	89.5	89.5	89.7	89.6	89.5	90.2	89.8	90.4
Mean-Mode	0.15	0.00	0.85	90.1	88.9	89.3	89.9	91.1	92.2	92.2	92.1
Mean-Mode	0.08	0.18	0.74	90.1	89.0	89.4	90.0	90.7	92.2	92.0	91.9
Mean-Mode	0.00	0.37	0.63	89.9	88.7	89.6	89.8	90.6	91.9	91.8	91.9
Mean-Median	0.50	0.00	0.50	89.8	89.2	89.2	89.6	91.0	92.0	92.3	90.4
Mean-Median	0.49	0.29	0.22	89.8	89.6	89.6	89.5	90.4	91.1	90.3	90.4
Mean-Median	0.41	0.59	0.00	89.5	89.8	89.8	89.2	89.7	90.3	89.4	89.7
Median-Mode	0.08	0.00	0.92	90.1	88.9	89.6	89.9	90.2	91.0	90.8	91.3
Median-Mode	0.04	0.08	0.88	89.9	89.0	89.6	89.8	90.1	90.9	90.7	91.4
Median-Mode	0.00	0.17	0.83	89.9	88.9	89.6	89.9	90.0	90.9	90.7	91.4
Mean-Median-Mode	0.18	0.00	0.82	90.2	88.9	89.2	89.9	91.3	92.5	92.7	92.6
Mean-Median-Mode	0.10	0.24	0.66	90.1	88.9	89.5	89.8	90.9	92.3	92.3	91.8
Mean-Median-Mode	0.00	0.51	0.49	89.7	88.8	89.6	89.6	90.5	91.8	91.8	91.4
				Panel B: Heteroskedastic data							
Mean	1.00	0.00	0.00	89.2	89.9	89.6	90.3	88.7	89.6	89.4	89.5
Mode	0.00	0.00	1.00	89.5	89.5	89.9	89.8	85.8	86.4	87.9	88.7
Median	0.28	0.00	0.72	89.3	89.2	89.9	90.1	90.4	91.1	91.1	91.8
Median	0.15	0.50	0.35	89.0	89.8	90.1	89.9	90.0	90.9	90.8	90.7
Median	0.00	1.00	0.00	89.3	89.9	90.1	90.2	89.3	89.9	89.7	89.5
Mean-Mode	0.15	0.00	0.85	89.3	89.4	89.8	89.9	90.2	91.3	91.1	91.4
Mean-Mode	0.08	0.18	0.74	89.2	89.5	89.8	90.0	90.0	91.1	91.0	90.9
Mean-Mode	0.00	0.37	0.63	89.3	89.7	89.9	90.0	89.9	90.8	90.7	91.0
Mean-Median	0.50	0.00	0.50	89.1	89.4	90.1	90.3	89.5	90.0	89.6	90.5
Mean-Median	0.49	0.29	0.22	89.2	89.6	90.0	90.5	89.3	90.2	89.7	90.0
Mean-Median	0.41	0.59	0.00	89.2	89.8	89.9	90.6	89.0	89.8	89.2	88.1
Median-Mode	0.08	0.00	0.92	89.3	89.3	89.8	89.9	89.2	90.0	90.3	90.8
Median-Mode	0.04	0.08	0.88	89.3	89.4	89.8	90.0	89.0	90.0	90.3	90.8
Median-Mode	0.00	0.17	0.83	89.4	89.5	89.9	90.0	89.1	90.1	90.1	90.4
Mean-Median-Mode	0.18	0.00	0.82	89.2	89.4	89.7	90.0	90.6	91.3	91.1	91.5
Mean-Median-Mode	0.10	0.24	0.66	89.2	89.7	89.9	90.1	90.3	91.2	91.1	91.3
Mean-Median-Mode	0.00	0.51	0.49	89.2	89.6	89.8	90.0	90.1	90.9	90.6	90.9

Notes: This tables presents the empirical coverage rates of the confidence sets for the forecasts of central tendency with a nominal coverage rate of 90%. We report the results for symmetric ($\gamma = 0$) and skewed data ($\gamma = 0.5$), for four sample sizes ($T = 100, 500, 2000, 5000$) and the two cross-sectional DGPs. We fix the instruments $\boldsymbol{h}_t = (1, X_t)$ and use a Gaussian kernel. In this application the set of identification function weights (θ) corresponding to a particular forecast combination weight vector is either a singleton (for the mean and mode) or a line. For the cases where the set is a line we present results for the end-points and the mid-point of this line.

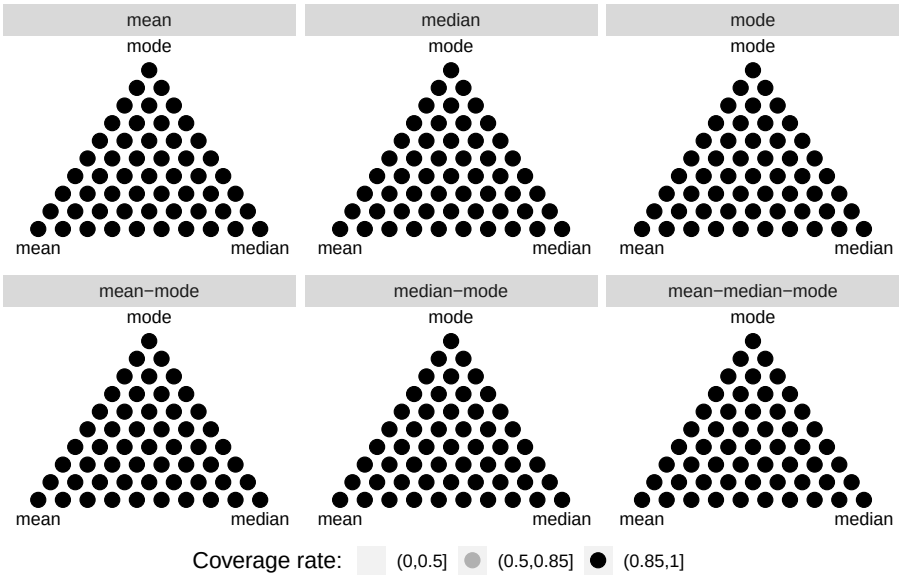
Table S.4: Empirical coverage of the confidence sets for central tendency: Time series data

Centrality measure	θ_{Mean}	θ_{Med}	θ_{Mode}	Symmetric data				Skewed data			
				100	500	2000	5000	100	500	2000	5000
				Panel A: Autoregressive data							
Mean	1.00	0.00	0.00	89.2	90.1	89.9	90.3	89.4	89.4	89.8	90.3
Mode	0.00	0.00	1.00	89.2	89.8	89.3	89.9	85.4	86.2	88.0	88.7
Median	0.28	0.00	0.72	89.1	89.7	89.0	89.6	91.5	92.3	93.5	91.7
Median	0.15	0.50	0.35	89.0	89.8	89.1	89.7	90.4	90.7	92.3	90.7
Median	0.00	1.00	0.00	88.9	90.0	89.2	89.9	89.0	89.2	90.7	89.9
Mean-Mode	0.15	0.00	0.85	89.1	89.6	89.1	89.8	90.5	92.0	92.7	91.7
Mean-Mode	0.08	0.18	0.74	88.9	89.9	89.2	89.7	90.5	91.6	92.5	91.3
Mean-Mode	0.00	0.37	0.63	89.0	89.9	89.4	89.6	90.2	91.4	92.3	91.5
Mean-Median	0.50	0.00	0.50	88.9	89.8	89.3	89.9	90.8	91.0	92.8	90.1
Mean-Median	0.49	0.29	0.22	88.8	90.2	89.3	89.9	90.0	89.8	91.2	89.7
Mean-Median	0.41	0.59	0.00	88.8	90.5	89.6	90.0	89.2	88.8	90.0	89.4
Median-Mode	0.08	0.00	0.92	89.2	89.7	89.1	89.8	89.4	90.8	91.4	91.1
Median-Mode	0.04	0.08	0.88	89.1	89.9	89.1	89.9	89.4	90.5	91.5	91.1
Median-Mode	0.00	0.17	0.83	89.1	90.0	89.1	89.8	89.3	90.6	91.5	91.0
Mean-Median-Mode	0.18	0.00	0.82	89.0	89.7	89.1	89.7	91.2	92.2	93.2	92.2
Mean-Median-Mode	0.10	0.24	0.66	89.0	89.7	89.2	89.6	90.7	91.7	92.7	91.7
Mean-Median-Mode	0.00	0.51	0.49	89.1	89.7	89.2	89.6	90.2	91.1	92.0	91.2
				Panel B: AR-GARCH data							
Mean	1.00	0.00	0.00	89.2	90.4	89.9	90.0	89.5	89.9	89.9	89.8
Mode	0.00	0.00	1.00	90.0	89.0	89.6	89.9	85.9	86.1	88.4	88.9
Median	0.28	0.00	0.72	89.5	89.1	89.1	89.4	91.1	92.5	93.1	92.0
Median	0.15	0.50	0.35	88.8	89.4	89.3	89.5	90.1	91.2	91.3	90.8
Median	0.00	1.00	0.00	88.6	89.5	89.8	89.6	89.1	89.7	89.8	90.1
Mean-Mode	0.15	0.00	0.85	89.9	88.9	89.4	89.7	90.4	91.5	92.7	91.5
Mean-Mode	0.08	0.18	0.74	89.6	89.1	89.6	89.6	90.3	91.3	92.3	91.3
Mean-Mode	0.00	0.37	0.63	89.5	89.1	89.6	89.6	90.2	91.2	92.1	91.1
Mean-Median	0.50	0.00	0.50	89.0	89.3	89.1	89.3	90.9	91.6	92.1	90.5
Mean-Median	0.49	0.29	0.22	88.7	89.7	89.3	89.4	90.1	90.7	90.8	89.9
Mean-Median	0.41	0.59	0.00	88.8	89.8	89.6	89.6	89.2	89.7	89.9	89.3
Median-Mode	0.08	0.00	0.92	90.2	89.0	89.5	89.7	89.4	90.5	92.1	91.0
Median-Mode	0.04	0.08	0.88	90.1	89.0	89.5	89.7	89.4	90.6	92.0	91.1
Median-Mode	0.00	0.17	0.83	90.0	89.2	89.6	89.7	89.3	90.5	91.8	90.8
Mean-Median-Mode	0.18	0.00	0.82	89.9	88.9	89.3	89.7	91.0	92.0	92.8	92.4
Mean-Median-Mode	0.10	0.24	0.66	89.6	89.2	89.3	89.7	90.4	91.7	92.3	91.2
Mean-Median-Mode	0.00	0.51	0.49	89.4	89.2	89.3	89.6	89.9	91.4	91.9	91.1

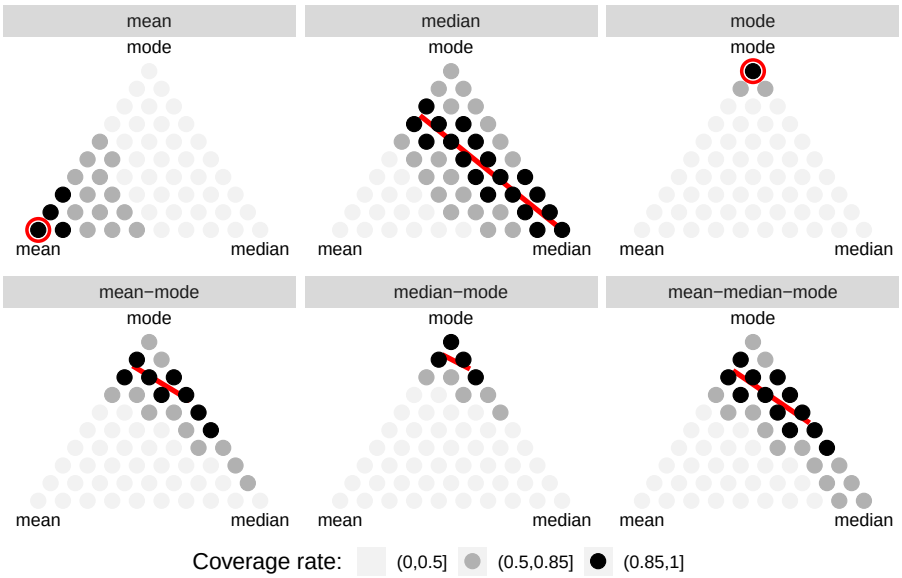
Notes: This tables presents the empirical coverage rates of the confidence sets for the forecasts of central tendency with a nominal coverage rate of 90%. We report the results for symmetric ($\gamma = 0$) and skewed data ($\gamma = 0.5$), for four sample sizes ($T = 100, 500, 2000, 5000$) and the two time-series DGPs. We fix the instruments $\boldsymbol{h}_t = (1, X_t)$ and use a Gaussian kernel. In this application the set of identification function weights (θ) corresponding to a particular forecast combination weight vector is either a singleton (for the mean and mode) or a line. For the cases where the set is a line we present results for the end-points and the mid-point of this line.

Figure S.4: Coverage rates of the confidence regions for central tendency measures for the homoskedastic DGP.

(a) Cross Sectional Homoskedastic DGP with skewness $\gamma = 0$



(b) Cross Sectional Homoskedastic DGP with skewness $\gamma = 0.5$



This figure shows coverage rates of 90% confidence regions for the measures of central tendency for the homoskedastic DGP. The true forecasted functional is given in the text above the triangle. The points inside the triangle correspond to convex combinations of the vertices, which are the mean, median and mode functionals. The color of the points indicates how often a specific point is contained in the 90% confidence regions. The upper panel shows results for the symmetric DGP, where all central tendency measures are equal. The lower panel uses a skewed DGP, with $\gamma = 0.5$. We use a red circle or a red line to indicate the (set of) central tendency measure(s) that correspond(s) to the forecast. We consider sample size $T = 2000$, instruments $\boldsymbol{h}_t = (1, X_t)$ and use a Gaussian kernel.