SUPPLEMENTARY MATERIAL FOR

Testing Forecast Rationality for Measures of Central Tendency

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

This Version: May 15, 2023

This appendix contains additional details, discussion, and results. It is structured as follows: Section S.1 contains proofs of theorems in the main paper that are not included in the paper's appendix. Sections S.2 and S.3 discuss the impact of the choice of kernel and bandwidth on the results for mode forecast rationality testing in Section 2.3 of the main paper. Section S.4 presents a result showing that a convex combination of functionals is generally not elicitable, as stated in Remark 3.1 of the main paper. Section S.5 presents simulation results supplementing those presented in Section 4 of the main paper. Section S.6 presents tests of rationality using a cluster covariance estimator. Section S.7 presents the results of tests of rationality in the framework of Elliott et al. (2005) for the full set of stratifications considered in Section 5 of the main paper. Section S.8 presents two additional empirical analyses, the first to the "Greenbook" forecasts produced by the Board of Governors of the Federal Reserve, and the second to random walk forecasts of exchange rates.

S.1 Additional Proofs

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

$$\begin{aligned}
&\lambda^\top \widehat{\Omega}_{T,\text{Mode}} \lambda - \lambda^\top \Omega_{T,\text{Mode}} \lambda \\
&= \frac{1}{T}\sum_{t=1}^{T} \delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2 (\lambda^\top \boldsymbol{h}_t)^2 - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_t\left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2 (\lambda^\top \boldsymbol{h}_t)^2\right] \\
&+ \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_t\left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2 (\lambda^\top \boldsymbol{h}_t)^2\right] - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 f_t(0) \int K'(u)^2 \mathrm{d}u\right].
\end{aligned} \tag{S.1.1}$$

We start by showing that the last line in (S.1.1) is $o_P(1)$. It holds that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_t\left[\delta_T^{-1} K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2 (\lambda^\top \boldsymbol{h}_t)^2\right] - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 f_t(0) \int K'(u)^2 \mathrm{d}u\right] \tag{S.1.2}$$

$$= \frac{1}{T}\sum_{t=1}^{T} (\lambda^\top \boldsymbol{h}_t)^2 \int K'(u)^2 f_t(\delta_T u) \mathrm{d}u - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 \int K'(u)^2 f_t(0) \mathrm{d}u\right] \xrightarrow{P} 0 \tag{S.1.3}$$

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

We further show that the penultimate line in (S.1.1) 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 MDA, we get

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

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.1.28) 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 $\lambda^\top\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\phi}_{t,T}(\theta_0)\widehat{\phi}_{t,T}(\theta_0)^\top\right)\lambda$ and $\sigma_T^2 = \lambda^\top\Sigma_T(\theta_0)\lambda$, given in (A.17), for some arbitrary but fixed $\lambda \in \mathbb{R}^k$ such that $||\lambda||_2 = 1$. Then, we get that

$$\lambda^\top\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\phi}_{t,T}(\theta_0)\widehat{\phi}_{t,T}(\theta_0)^\top\right)\lambda = \frac{1}{T}\sum_{t=1}^{T}\left\{\theta_{10}^2\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2\right. \tag{S.1.4}$$
$$+\theta_{20}^2\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Med}}\lambda\right)^2\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)^2 + \theta_{30}^2\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mode}}\lambda\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2 \tag{S.1.5}$$
$$+2\theta_{10}\theta_{20}\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Med}}\lambda\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right) \tag{S.1.6}$$
$$+2\theta_{10}\theta_{30}\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mean}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mode}}\lambda\right)\varepsilon_t\ \delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right) \tag{S.1.7}$$
$$\left.+2\theta_{20}\theta_{30}\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Med}}\lambda\right)\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mode}}\lambda\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}}\right)\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right\}. \tag{S.1.8}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\theta_{10}^2\left(\boldsymbol{h}_t^\top\widehat{w}_{T,\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2 - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}^2\left(\boldsymbol{h}_t^\top w_{\mathrm{Mean}}\lambda\right)^2\varepsilon_t^2\right] \xrightarrow{P} 0 \tag{S.1.9}$$

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

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

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

$$= T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\int K\left(\frac{e}{\delta_T}\right)f_t'(e)\,\mathrm{d}e - T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\left[K\left(\frac{e}{\delta_T}\right)f_t(e)\right]_{e=-\infty}^{e=\infty}. \tag{S.1.11}$$

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 $t\leq T$. By transformation of variables, it further holds that

$$g_{t,T}^{e} = T^{-1/2}\delta_T^{1/2}\boldsymbol{h}_t\int K\left(\frac{e}{\delta_T}\right)f_t'(e)\,\mathrm{d}e = T^{-1/2}\delta_T^{3/2}\boldsymbol{h}_t\int K(u)\,f_t'(\delta_T u)\,\mathrm{d}u. \tag{S.1.12}$$

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

$$f_t'(\delta_T u) = f_t'(0) + (\delta_T u)f_t''(0) + \frac{(\delta_T u)^2}{2}f_t'''(\zeta\delta_T u), \tag{S.1.13}$$

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

$$\sum_{t=1}^{T} g_{t,T}^{e} = T^{-1/2}\delta_T^{5/2}\sum_{t=1}^{T}f_t''(0)\boldsymbol{h}_t\int uK(u)\,\mathrm{d}u \tag{S.1.14}$$

$$+\frac{1}{2}T^{-1/2}\delta_T^{7/2}\sum_{t=1}^{T}\boldsymbol{h}_t\int u^2K(u)\,f_t'''(\zeta\delta_T u)\,\mathrm{d}u. \tag{S.1.15}$$

As $\int uK(u)\,\mathrm{d}u = 0$ by assumption (A6), the first term is zero for all $T\in\mathbb{N}$. As $\sup_x f_t'''(x)\leq c$ by Assumption (A5) and $\int u^2K(u)\,\mathrm{d}u\leq c<\infty$ by Assumption (A6), we obtain

$$\frac{1}{2}T^{-1/2}\delta_T^{7/2}\sum_{t=1}^{T}\boldsymbol{h}_t\int u^2K(u)\,f_t'''(\zeta\delta_T u)\,\mathrm{d}u \leq \frac{1}{2}c^2\left(T\delta_T^7\right)^{1/2}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{h}_t \xrightarrow{P} 0, \tag{S.1.16}$$

as $T\delta_T^7\to 0$ for $T\to\infty$ by Assumption (A7) 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 α-mixing sequences (White, 2001, Corollary 3.48). The result of the lemma follows. □

Lemma S.1.2. *Given Assumption 2.5, it holds that* $\sum_{t=1}^{T}\mathrm{Var}\left(z_{t,T}\right)\to\omega^2=\lambda^\top\Omega_{\mathrm{Mode}}\lambda$.

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

$$\text{Var}\left(z_{t,T}\right) = \mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] - \mathbb{E}\left[\left(\lambda^\top g^e_{t,T}\right)^2\right], \tag{S.1.17}$$

as $\mathbb{E}\left[\left(\lambda^\top g^e_{t,T}\right)\cdot\left(\lambda^\top g_{t,T}\right)\right] = \mathbb{E}\left[\left(\lambda^\top g^e_{t,T}\right)\cdot\mathbb{E}_t\left[\lambda^\top g_{t,T}\right]\right] = \mathbb{E}\left[\left(\lambda^\top g^e_{t,T}\right)^2\right]$. For the first term in (S.1.17), we obtain

$$\mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] = \mathbb{E}\left[(T\delta_T)^{-1}(\lambda^\top \boldsymbol{h}_t)^2\mathbb{E}_t\left[K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)^2\right]\right] \tag{S.1.18}$$

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

$$= \frac{1}{T}\,\mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2\int K'\left(u\right)^2 f_t(\delta_T u)\,\mathrm{d}u\right]. \tag{S.1.20}$$

As $\delta_T \to 0$ when $T \to \infty$ and as $\Omega_{T,\text{Mode}} \to \Omega_{\text{Mode}}$ from Assumption (A4), we get

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] - \lambda^\top\Omega_{\text{Mode}}\lambda \\
&= \sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top g_{t,T}\right)^2\right] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2 f_t(0)\right]\int K'\left(u\right)^2\mathrm{d}u \\
&\qquad + \lambda^\top\Omega_{T,\text{Mode}}\lambda - \lambda^\top\Omega_{\text{Mode}}\lambda \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[(\lambda^\top \boldsymbol{h}_t)^2\int K'\left(u\right)^2 f_t(\delta_T u)\,\mathrm{d}u\right] - \lambda^\top\mathbb{E}\left[f_t(0)\boldsymbol{h}_t\boldsymbol{h}_t^\top\right]\lambda\int K'\left(u\right)^2\mathrm{d}u\right) + o(1) \to 0.
\end{aligned} \tag{S.1.21}$$

For the second term in (S.1.17), inserting the equality in (S.1.12) yields

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

As $\sup_x |f_t'(x)| \leq c$ for all $t \in \mathbb{N}$ and $\left|\int K'\left(u\right)\mathrm{d}u\right| \leq c$ by assumption, it holds that

$$\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top g^e_{t,T}\right)^2\right] \leq \delta_T^3 c^2||\lambda||^2\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[||\boldsymbol{h}_t||^2\right]\right) \to 0, \tag{S.1.22}$$

as $\delta_T^3 \to 0$ as $T \to \infty$. The result of the lemma follows by combining (S.1.21) and (S.1.22). □

Lemma S.1.3. *Given Assumption 2.5, it holds that* $\sum_{t=1}^{T} z_{t,T}^2 \xrightarrow{P} \omega^2 = \lambda^\top\Omega_{\text{Mode}}\lambda$.

Proof. We define

$$h_{1,T} := \sum_{t=1}^{T} \left(z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right) \qquad \text{and} \qquad h_{2,T} := \sum_{t=1}^{T} \mathbb{E}_t\left[z_{t,T}^2\right] - \omega^2, \tag{S.1.23}$$

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 $\mathcal{F}_{t+1}$-MDA by definition. Thus, we can apply the von Bahr and Esseen (1965)-inequality for some $p \in (1,2)$ for MDA (in the first line) in order to conclude that

$$\mathbb{E}\left[|h_{1,T}|^p\right] = \mathbb{E}\left[\left|\sum_{t=1}^{T} z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right] \leq 2\sum_{t=1}^{T} \mathbb{E}\left[\left|z_{t,T}^2 - \mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right] \tag{S.1.24}$$

$$\leq 2\sum_{t=1}^{T} 2^{p-1}\left(\mathbb{E}\left[\left|z_{t,T}^2\right|^p\right] + \mathbb{E}\left[\left|\mathbb{E}_t\left[z_{t,T}^2\right]\right|^p\right]\right) \leq 2^{p+1}\sum_{t=1}^{T} \mathbb{E}\left[\left|z_{t,T}\right|^{2p}\right], \tag{S.1.25}$$

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

$$\mathbb{E}\left[\left|z_{t,T}\right|^{2p}\right] = \mathbb{E}\left[\left|\lambda^\top g_{t,T} - \lambda^\top g_{t,T}^e\right|^{2p}\right] \leq 2^{2p-1}\left(\mathbb{E}\left[\left|\lambda^\top g_{t,T}\right|^{2p}\right] + \mathbb{E}\left[\left|\lambda^\top g_{t,T}^e\right|^{2p}\right]\right). \tag{S.1.26}$$

Thus,

$$\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top g_{t,T}\right|^{2p}\right] = (T\delta_T)^{1-p}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p} \int \left|K'(u)\right|^{2p} f_t(\delta_T u)\mathrm{d}u\right] \to 0, \tag{S.1.27}$$

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 $\int \left|K'(u)\right|^{2p} f_t(\delta_T u)\mathrm{d}u \leq cc^{2p-1}\int \left|K'(u)\right|\mathrm{d}u < \infty$ as $\int \left|K'(u)\right|\mathrm{d}u < \infty$, $\sup_u \left|K'(u)\right| \leq c$ and $\sup_x f_t(x) \leq c$ a.s. by assumption. Similarly, we obtain

$$\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top g_{t,T}^e\right|^{2p}\right] = \delta_T^{2p-1}(T\delta_T)^{1-p}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left|\lambda^\top \boldsymbol{h}_t\right|^{2p}\left|\int K'(u)\, f_t(\delta_T u)\mathrm{d}u\right|^{2p}\right] \to 0, \tag{S.1.28}$$

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.1.17), we split

$$h_{2,T} = \sum_{t=1}^{T} \mathbb{E}_t\left[z_{t,T}^2\right] - \omega^2 = \sum_{t=1}^{T} \mathbb{E}_t\left[(\lambda^\top g_{t,T})^2\right] - \sum_{t=1}^{T} (\lambda^\top g_{t,T}^e)^2 - \omega^2. \tag{S.1.29}$$

Applying a transformation of variables yields

$$\sum_{t=1}^{T}(\lambda^\top g^e_{t,T})^2 = \delta_T \frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2 \left(\int K'(u) f_t(\delta_T u)\mathrm{d}u\right)^2 \tag{S.1.30}$$

$$\leq \delta_T \left(\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\right)\left(c\int |K'(u)|\mathrm{d}u\right)^2 \xrightarrow{P} 0, \tag{S.1.31}$$

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,

$$\sum_{t=1}^{T}\mathbb{E}_t\left[\left(\lambda^\top g_{t,T}\right)^2\right] - \omega^2 \tag{S.1.32}$$

$$= \left(\frac{1}{T}\sum_{t=1}^{T}(\lambda^\top \boldsymbol{h}_t)^2\right)\int K'(u)^2 f_t(\delta_T u)\mathrm{d}u - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[f_t(0)(\lambda^\top \boldsymbol{h}_t)^2\right]\int K'(u)^2\mathrm{d}u + \bar{\omega}_T^2 - \omega^2 \xrightarrow{P} 0, \tag{S.1.33}$$

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

Lemma S.1.4. *Given Assumption 2.5, 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,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq t\leq T}|h_{t,T}| > \zeta\right) &= \mathbb{P}\left(\max_{1\leq t\leq T}|h_{t,T}|^{2+\delta} > \zeta^{2+\delta}\right) \leq \sum_{t=1}^{T}\mathbb{P}\left(|h_{t,T}|^{2+\delta} > \zeta^{2+\delta}\right) \\ &\leq \zeta^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[|h_{t,T}|^{2+\delta}\right] = \zeta^{-2-\delta}\bar{\omega}_T^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[|z_{t,T}|^{2+\delta}\right],\end{aligned} \tag{S.1.34}$$

by Markov's inequality. Employing the same steps as in the proof of Lemma S.1.3 following Equation (S.1.26) 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 $\bar{\omega}_T \to \omega^2 > 0$, 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 and Assumption 3.2, for all* $\lambda \in \mathbb{R}^k$ *such that* $||\lambda||_2 = 1$, *it holds that* $\sum_{t=1}^{T}\mathrm{Var}\left(T^{-1/2}\phi^*_{t,T}(\theta_0)^\top\lambda\right) - \sigma_T^2 \to 0$.

Proof. As $T^{-1/2}\phi^*_{t,T}(\theta_0)$ is a $\mathcal{F}_{t+1}$-MDA, it holds that $\mathbb{E}\left[T^{-1/2}\phi^*_{t,T}(\theta_0)^\top\lambda\right] = 0$ and thus, $\mathrm{Var}\left(T^{-1/2}\phi^*_{t,T}(\theta_0)^\top\lambda\right) =$

$\mathbb{E}\left[\left(T^{-1/2}\phi^*_{t,T}(\theta_0)^\top\lambda\right)^2\right]$. We further have

$$
\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\left[\left(T^{-1/2}\phi^*_{t,T}(\theta_0)^\top\lambda\right)^2\right] \qquad (S.1.35)\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}^2\left(\lambda^\top w_{\text{Mean}}\boldsymbol{h}_t\right)^2\varepsilon_t^2\right]+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{20}^2\left(\lambda^\top w_{\text{Med}}\boldsymbol{h}_t\right)^2\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)^2\right]\\
&+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{30}^2\left(\lambda^\top w_{\text{Mode}}\boldsymbol{h}_t\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2\right]\\
&+\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{20}\left(\lambda^\top w_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top w_{\text{Med}}\boldsymbol{h}_t\right)\varepsilon_t\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\right]\\
&+\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top w_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top w_{\text{Mode}}\boldsymbol{h}_t\right)\varepsilon_t\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\\
&+\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{20}\theta_{30}\left(\lambda^\top w_{\text{Med}}\boldsymbol{h}_t\right)\left(\lambda^\top w_{\text{Mode}}\boldsymbol{h}_t\right)\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)\delta_T^{-1/2}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\\
&+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[(\lambda^\top u_{t,T}(\theta_0))^2\right]-\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top u_{t,T}(\theta_0)\right)\left(\lambda^\top\phi_{t,T}(\theta_0)\right)\right].
\end{aligned}
$$

For the last two terms, we have $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top u_{t,T}(\theta_0)\right)^2\right]\to 0$ and $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\left(\lambda^\top u_{t,T}(\theta_0)\right)\left(\lambda^\top\phi_{t,T}(\theta_0)\right)\right]\to 0$ by Assumption 3.2 (D). For the fifth term,

$$
\begin{aligned}
&\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top w_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top w_{\text{Mode}}\boldsymbol{h}_t\right)\delta_T^{-1/2}\mathbb{E}_t\left[\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\right] \qquad (S.1.36)\\
&=-\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{10}\theta_{30}\left(\lambda^\top w_{\text{Mean}}\boldsymbol{h}_t\right)\left(\lambda^\top w_{\text{Mode}}\boldsymbol{h}_t\right)\delta_T^{3/2}\int uK'(u)\,f_t(\delta_T u)\mathrm{d}u\right]\to 0, \qquad (S.1.37)
\end{aligned}
$$

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 $\left|\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right|\le 1$. For the third term, it holds that

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

The remaining first, second and fourth terms in (S.1.35) appear equivalently in σ_T^2, which concludes the proof of this lemma. □

Lemma S.1.6. *Given Assumption 2.5 and Assumption 3.2, for all $\lambda\in\mathbb{R}^k$ such that $||\lambda||_2=1$, it holds that*

$T^{-1}\sum_{t=1}^{T}\left(\phi_{t,T}^{*}(\theta_0)^{\top}\lambda\right)^2-\sigma_T^2\xrightarrow{P}0.$

Proof. We apply the same factorization as in (S.1.35) (however without the expectation operator). By applying a law of large numbers for mixing sequences (Corollary 3.48 in White (2001)), we obtain that

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

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

$$\frac{1}{T}\sum_{t=1}^{T}\theta_{30}^2\left(\boldsymbol{h}_t^{\top}w_{\text{Mode}}\lambda\right)^2\delta_T^{-1}K'\left(\frac{-\varepsilon_t}{\delta_T}\right)^2-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\theta_{30}^2\left(\boldsymbol{h}_t^{\top}w_{\text{Mode}}\lambda\right)^2 f_t(0)\int K'(u)^2\,\mathrm{d}u\right]\xrightarrow{P}0. \tag{S.1.38}$$

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

$$\begin{aligned}
&\frac{2}{T}\sum_{t=1}^{T}\theta_{10}\theta_{30}\left(\boldsymbol{h}_t^{\top}w_{\text{Mean}}\lambda\right)\left(\boldsymbol{h}_t^{\top}w_{\text{Mode}}\lambda\right)\delta_T^{-1/2}\mathbb{E}_t\left[\varepsilon_t K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right] &\text{(S.1.39)}\\
=&-\frac{2}{T}\sum_{t=1}^{T}\theta_{10}\theta_{30}\left(\boldsymbol{h}_t^{\top}w_{\text{Mean}}\lambda\right)\left(\boldsymbol{h}_t^{\top}w_{\text{Mode}}\lambda\right)\delta_T^{3/2}\int uK'(u)\,f_t(\delta_T u)\mathrm{d}u\xrightarrow{P}0, &\text{(S.1.40)}
\end{aligned}$$

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

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

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

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

where the first term is bounded from above by

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

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.1.28)). Applying the same line of reasoning for the mixed median/mode terms shows that

$$\frac{2}{T}\sum_{t=1}^{T}\theta_{20}\theta_{30}\left(\boldsymbol{h}_t^\top w_{\text{Med}}\lambda\right)\left(\boldsymbol{h}_t^\top w_{\text{Mode}}\lambda\right)\delta_T^{-1/2}\mathbb{E}_t\left[\left(\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right)K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\xrightarrow{P}0. \tag{S.1.41}$$

For the fourth and last term, $T^{-1}\sum_{t=1}^{T}\left(u_{t,T}(\theta_0)^\top\lambda\right)^2\xrightarrow{P}0$ and $T^{-1}\sum_{t=1}^{T}\left(u_{t,T}(\theta_0)^\top\lambda\right)\left(\phi_{t,T}(\theta_0)^\top\lambda\right)\xrightarrow{P}0$ by Assumption 3.2 (D), which concludes this proof. □

Lemma S.1.7. *Given Assumption 2.5 and Assumption 3.2, for all $\lambda\in\mathbb{R}^k$ such that $||\lambda||_2=1$, it holds that* $\max_{1\leq t\leq T}\left|\sigma_T^{-1}T^{-1/2}\phi_{t,T}^*(\theta_0)^\top\lambda\right|\xrightarrow{P}0$.

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.1.34) in the proof of Lemma S.1.4, we get that

$$\mathbb{P}\left(\max_{1\leq t\leq T}\left|\sigma_T^{-1}T^{-1/2}\phi_{t,T}^*(\theta_0)^\top\lambda\right|>\zeta\right)\leq\zeta^{-2-\delta}\sigma_T^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[\left|T^{-1/2}\phi_{t,T}^*(\theta_0)^\top\lambda\right|^{2+\delta}\right], \tag{S.1.42}$$

by Markov's inequality. Furthermore, we get that

$$4^{-2-\delta}\sum_{t=1}^{T}\mathbb{E}\left[\left|T^{-1/2}\phi_{t,T}^*(\theta_0)^\top\lambda\right|^{2+\delta}\right]\leq\sum_{t=1}^{T}\mathbb{E}\left[\left|T^{-1/2}u_{t,T}(\theta_0)^\top\lambda\right|^{2+\delta}\right] \tag{S.1.43}$$

$$+\theta_{10}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top w_{\text{Mean}}\lambda\right|^{2+\delta}|\varepsilon_t|^{2+\delta}\right] \tag{S.1.44}$$

$$+\theta_{20}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top w_{\text{Med}}\lambda\right|^{2+\delta}\left|\mathbb{1}_{\{\varepsilon_t>0\}}-\mathbb{1}_{\{\varepsilon_t<0\}}\right|^{2+\delta}\right] \tag{S.1.45}$$

$$+\theta_{30}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top w_{\text{Mode}}\lambda\right|^{2+\delta}\delta_T^{-\frac{2+\delta}{2}}\left|K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right|^{2+\delta}\right]. \tag{S.1.46}$$

The first term converges to zero by Assumption 3.2. 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,

$$\theta_{30}^{2+\delta}T^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top w_{\text{Mode}}\lambda\right|^{2+\delta}\delta_T^{-\frac{2+\delta}{2}}\left|K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right|^{2+\delta}\right] \tag{S.1.47}$$

$$\leq\theta_{30}^{2+\delta}(T\delta_T)^{-\frac{\delta}{2}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{h}_t^\top w_{\text{Mode}}\lambda\right|^{2+\delta}\int\left|K'(u)\right|^{2+\delta}f_t(\delta_T u)\mathrm{d}u\right], \tag{S.1.48}$$

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

Lemma S.1.8. *If X_t is the mode of F_t for all $t\in\mathbb{N}$, the choice of $u_{t,T}(\theta_0)$ in (3.6) satisfies Assumption 3.2 (D).*

Proof. If X_t is the mode of F_t, we set $\theta_0=(0,0,1)$ and thus, $\phi_{t,T}(\theta_0)=-\omega_{\text{Mode}}\delta_T^{-1/2}K'(\varepsilon_t/\delta_T)\boldsymbol{h}_t$ and we get $\phi_{t,T}(\theta_0)=\phi_{t,T}^*(\theta_0)+u_{t,T}(\theta_0)$ by setting

$$T^{-1/2}\phi_{t,T}^*(\theta_0)=\omega_{\text{Mode}}\,g_{t,T}^* \qquad\text{and}\qquad T^{-1/2}u_{t,T}(\theta_0)=\omega_{\text{Mode}}\,g_{t,T}^e,$$

as in the proof of Theorem 2.6. Thus, $T^{-1/2}\phi_{t,T}^*(\theta_0)=\omega_{\text{Mode}}\,g_{t,T}^*$ is a MDA satisfying condition (a).

For the remaining conditions (b) and (c) of Assumption 3.2 (D), as in the proof of Lemma S.1.1, we get

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

As $\delta_T\to 0$ and $\int u^2K(u)\,f_t'''(\zeta\delta_T u)$ is bounded as argued in the proof of Lemma S.1.1, we get

$$T^{-1}\sum_{t=1}^{T}||u_{t,T}(\theta_0)||^2=\sum_{t=1}^{T}||\omega_{\text{Mode}}\,g_{t,T}^e||^2=\delta_T^7\frac{\omega_{\text{Mode}}^2}{4}\,\frac{1}{T}\sum_{t=1}^{T}\left|\left|\boldsymbol{h}_t\int u^2K(u)\,f_t'''(\zeta\delta_T u)\,\mathrm{d}u\right|\right|^2\xrightarrow{P}0.$$

Furthermore, by using similar arguments as above, we get that

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} u_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top &= -\omega_{\text{Mode}}^2 \sum_{t=1}^{T} g_{t,T}^e \cdot (T\delta_T)^{-1/2} K'(\varepsilon_t/\delta_T)\boldsymbol{h}_t^\top \\
&= -\omega_{\text{Mode}}^2 \sum_{t=1}^{T} \left(\frac{1}{2}T^{-1/2}\delta_T^{7/2}\boldsymbol{h}_t \int u^2 K(u)\, f_t'''(\zeta\delta_T u)\,\mathrm{d}u\right)(T\delta_T)^{-1/2} K'(\varepsilon_t/\delta_T)\boldsymbol{h}_t^\top \\
&= \delta_T^3 \frac{\omega_{\text{Mode}}^2}{2}\frac{1}{T}\sum_{t=1}^{T} K'(\varepsilon_t/\delta_T)\int u^2 K(u)\, f_t'''(\zeta\delta_T u)\,\mathrm{d}u\, \boldsymbol{h}_t\boldsymbol{h}_t^\top \xrightarrow{P} 0.
\end{aligned}
$$

The condition $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[u_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top\right] \to 0$ follows by exactly the same arguments, but working under expectation. Eventually, we get that

$$
\sum_{t=1}^{T}\mathbb{E}\left[||T^{-1/2}u_{t,T}(\theta_0)||^{2+\delta}\right] = T^{-\delta/2}\delta_T^{(14+7\delta)/2}\,\frac{\omega_{\text{Mode}}^{2+\delta}}{2^{2+\delta}}\,\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left|\left|\boldsymbol{h}_t\int u^2 K(u)\, f_t'''(\zeta\delta_T u)\,\mathrm{d}u\right|\right|^{2+\delta}\right] \to 0,
$$

as $T^{-\delta/2}\delta_T^{(14+7\delta)/2} \to 0$ and the remaining term is bounded, which concludes the proof. □

S.2 Kernel Choice

The asymptotic results presented in Section 2.3 rely on the chosen kernel K satisfying Assumption (A6). 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 also well-known in the literature on nonparametric statistics that kernels with bounded support can be more efficient (Li and Racine, 2006). However, note that the kernel choice enters the asymptotic variance of nonparametric density estimation through the quantity $\int K(u)^2\mathrm{d}u$, while the covariance Ω_{Mode} in Theorem 2.6 depends upon $\int K'(u)^2\mathrm{d}u$, revealing that the efficiency of our mode rationality tests depends upon a different quantity. Figure S.1 illustrates that the test power does not increase by employing a biweight kernel, which has bounded support, and is usually found to be relatively efficient in nonparametric estimation.

Strict identifiability of the generalized modal midpoint—and hence, asymptotic identifiability of the mode—only holds for kernel functions with unbounded support, which is satisfied by the Gaussian kernel, but not so for the biweight kernel.

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 Convex combination of functional values

Here, we illustrate that a convex combination of functionals is generally neither elicitable nor identifiable. This result shows that—as stated in Remark 3.1—testing forecast rationality directly for convex functional combinations is impossible. For this, we adapt the simplified notation of Section 2.2.

Proposition S.4.1. *Let $\mathcal{P}$ be a convex class of distributions and let $\Gamma_\beta(P) = \beta\Gamma^l(P)+(1-\beta)\Gamma^n(P)$, $\beta \in [0,1]$ be the convex combination of a* linear *functional $\Gamma^l : \mathcal{P} \to \mathbb{R}$ and a non-linear functional $\Gamma^n : \mathcal{P} \to \mathbb{R}$, which are both continuous (in the distribution P) and translation equivariant, i.e., if $P \in \mathcal{P}$, then for $c \in \mathbb{R}$ the shifted $P + c \in \mathcal{P}$ and $\Gamma^l(P + c) = \Gamma^l(P) + c$ and $\Gamma^n(P + c) = \Gamma^n(P) + c$. Then, the functional Γ_β is neither elicitable nor identifiable.*

Proof of Proposition S.4.1. Theorem 6 of Gneiting (2011) and Proposition 3.11 of Fissler and Hoga (2023) show that convex level sets, i.e.,

$$\text{for } P_1, P_2 \in \mathcal{P} \text{ with } \Gamma_\beta(P_1) = \Gamma_\beta(P_2) \quad \Longrightarrow \quad \Gamma_\beta(\alpha P_1 + (1-\alpha)P_2) = \Gamma_\beta(P_1), \tag{S.4.1}$$

for $\alpha \in (0,1)$ are a necessary condition for elicitability and identifiability of a functional. We show that the functional Γ_β does not have convex level sets.

As Γ^n is not linear, there exists $\alpha \in (0,1)$, and P_1, $P_2 \in \mathcal{P}$ such that

$$\Gamma^n(\alpha P_1 + (1-\alpha)P_2) \neq \alpha\Gamma^n(P_1) + (1-\alpha)\Gamma^n(P_2).$$

Define $P_2' = P_2 - \Gamma_\beta(P_2) + \Gamma_\beta(P_1)$ such that $\Gamma_\beta(P_2') = \Gamma_\beta(P_1)$ and

$$\Gamma^n(\alpha P_1 + (1-\alpha)P_2') \neq \alpha\Gamma^n(P_1) + (1-\alpha)\Gamma^n(P_2')$$

as $\Gamma^n(P + c) = \Gamma^n(P) + c$. It follows that

$$\begin{aligned}
\Gamma_\beta(\alpha P_1 + (1-\alpha)P_2') &= \beta\Gamma^l(\alpha P_1 + (1-\alpha)P_2') + (1-\beta)\Gamma^n(\alpha P_1 + (1-\alpha)P_2') \\
&\neq \beta\alpha\Gamma^l(P_1) + \beta(1-\alpha)\Gamma^l(P_2') + (1-\beta)\alpha\Gamma^n(P_1) + (1-\beta)(1-\alpha)\Gamma^n(P_2') \\
&= \alpha\Gamma_\beta(P_1) + (1-\alpha)\Gamma_\beta(P_2') \\
&= \Gamma_\beta(P_1)
\end{aligned}$$

and hence Γ_β does not have convex level sets. □

As functionals are linear if and only if they are expectations (Abernethy and Frongillo, 2012), the mean is linear and the median is non-linear for classes $\mathcal{P}$ sufficiently rich enough such that it contains distributions for which the median does not equal the mean (i.e., asymmetric distributions). Hence, Proposition S.4.1 shows that a convex combination of the mean and median is generally neither elicitable nor identifiable. Eliciting a convex combination that may further include the mode (which is itself only asymptotically elicitable) is only possible in the unusual case where a convex (sub-)combination of the nonlinear median and mode functionals becomes linear, an outcome that does not generally hold.

Fortunately, testing rationality for functionals elicited through a convex combination of loss functions is feasible and its interpretability is supported by the following result that convexity of the combination weights is preserved when moving from a functional elicited by a convex combination of loss (identification) functions to a convex combination of functional values.

The loss functions L_{Mean}, L_{Med} and $L_{\text{Mode},\delta} = \delta^{3/2} L_\delta^K$ are defined just before equation (3.1) and the identification functions V_{Mean}, V_{Med} and $V_{\text{Mode},\delta} = \delta^{3/2} V_\delta^K$ at equations (2.2), (2.4) and (2.7). We use the scaling by $\delta^{3/2}$ for the asymptotic mode loss and identification functions to be consistent with Section 3.

Proposition S.4.2. *Let $\mathcal{P}$ be some class of distributions such that the mean, μ, the median, m, and the generalized modal midpoint with parameter δ, Γ_δ^K, exist and are elicited by their loss functions L_{Mean}, L_{Med}, and $L_{\text{Mode},\delta}$. Let x be the functional defined by*

$$x(P) = \underset{\tilde{x}\in\mathbb{R}}{\arg\min}\, \mathbb{E}_{Y\sim P}\left[\theta_0^\top \left(L_{\text{Mean}}(\tilde{x}, Y), L_{\text{Med}}(\tilde{x}, Y), L_{\text{Mode},\delta}(\tilde{x}, Y)\right)^\top\right] \tag{S.4.2}$$

for some $\theta_0 \in \Theta$. Then, for every $P \in \mathcal{P}$ there exists some $\beta_0 \in \Theta$, such that $x(P) = \beta_0^\top \left(\mu(P), m(P), \Gamma_\delta^K(P)\right)^\top$.

Proof of Proposition S.4.2. Let $P \in \mathcal{P}$, where we assume without loss of generality that the three functionals are not all equal. For notational convenience we drop P when denoting functional values, e.g., we write μ instead of $\mu(P)$. For the elicited forecast x, it holds that

$$\overline{V}(x) := \theta_0^\top \left(\overline{V}_{\text{Mean}}(x, P), \overline{V}_{\text{Med}}(x, P), \overline{V}_{\text{Mode},\delta}(x, P)\right)^\top = 0. \tag{S.4.3}$$

We define $L := \min(\mu, m, \Gamma_\delta^K)$ and $U := \max(\mu, m, \Gamma_\delta^K)$ as the lower and upper functional values where it holds that $L < U$. Further let $\bar{V}_L(x, P)$ and $\bar{V}_U(x, P)$ denote the corresponding expected identification functions for the distribution P. Suppose that $x < L$. Then, it must hold that $\overline{V}(x) > 0$ as all three expected identification functions have the same sign as they are oriented in the sense of Steinwart et al. (2014). Similarly, if $x > U$, it must hold that $\overline{V}(x) < 0$. Hence, we can conclude that $x \in [L, U]$, which

implies that there exists $\zeta \in [0,1]$ such that $x = \zeta L + (1-\zeta)U$. Thus, x can be constructed as a convex combination of the functional values, i.e., there exists a $\beta_0 \in \Theta$ such that $x = \beta_0^\top \left(\mu, m, \Gamma_\delta^K\right)^\top$. □

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.

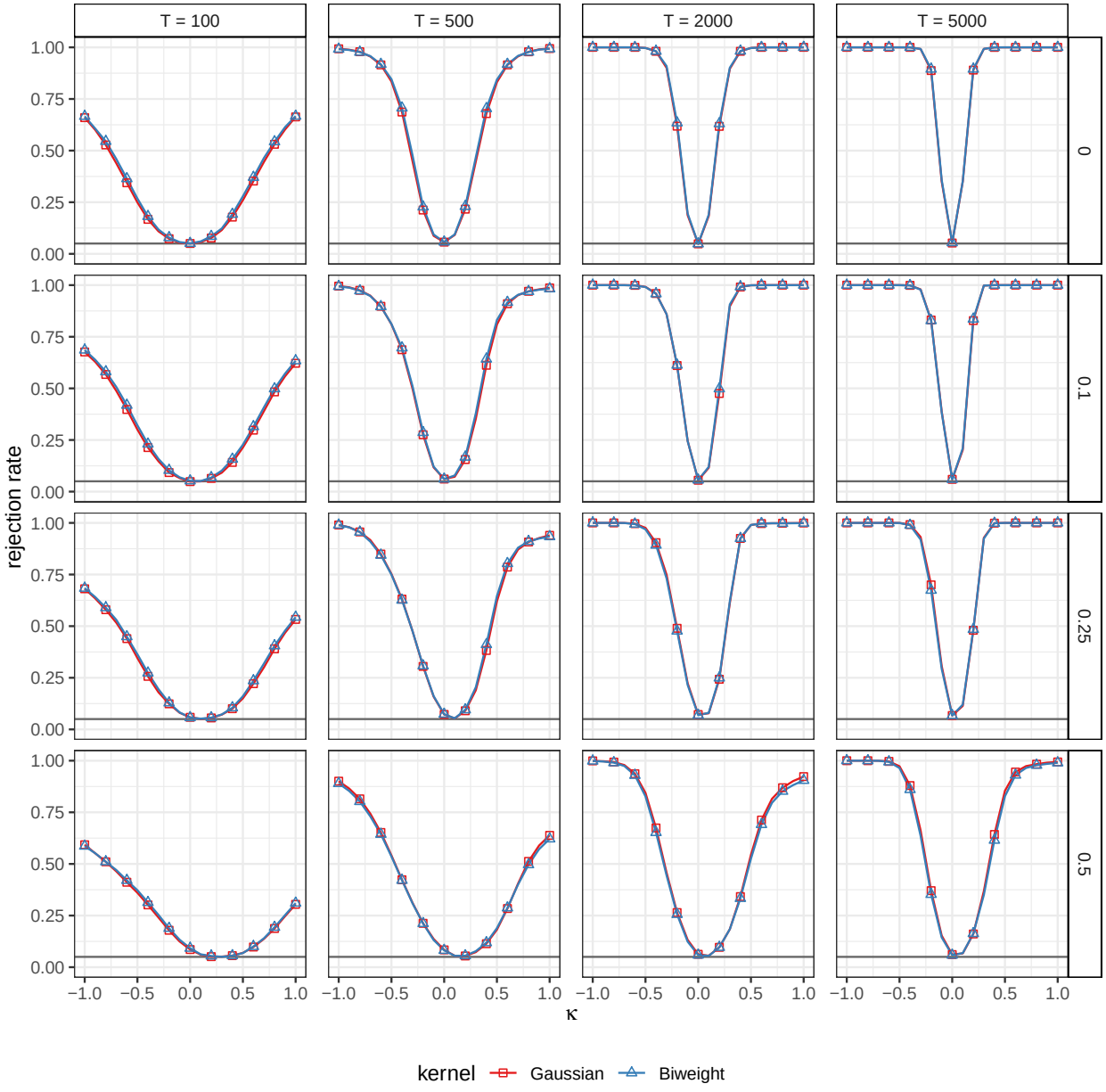
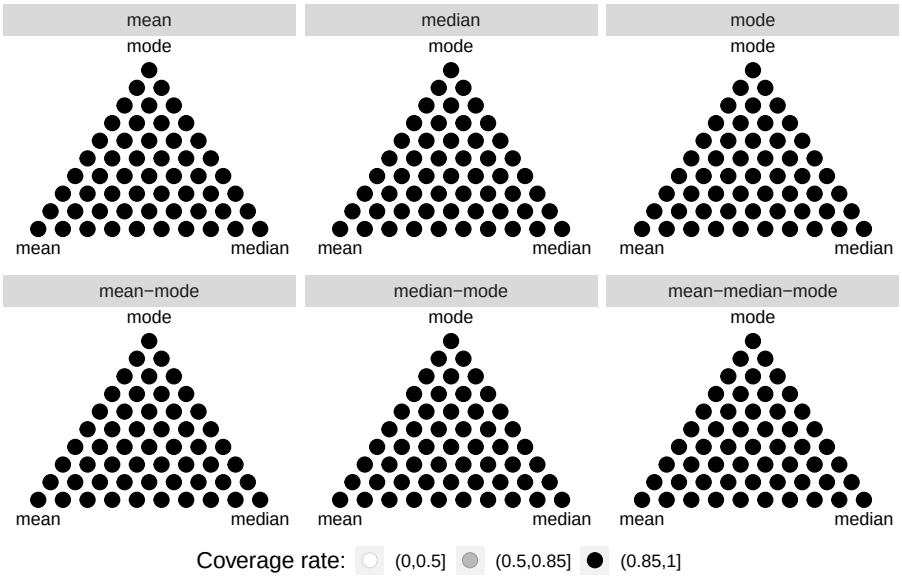


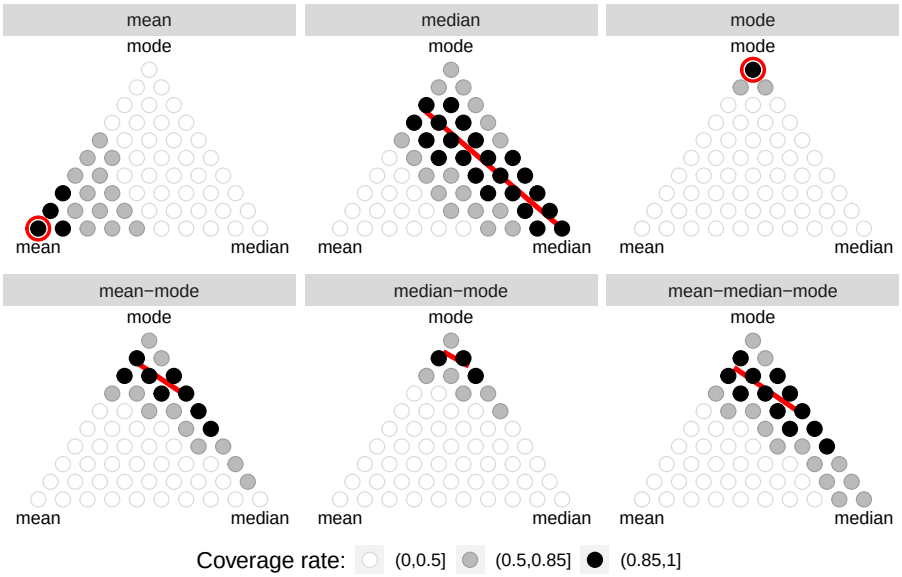
Figure S.1: Test power for different kernel functions. This figure plots the empirical rejection frequencies for the Gaussian and the biweight kernels against the degrees of misspecification κ for different sample sizes in the vertical panels and for four skewness levels in the horizontal panels. We 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%.

Figure S.2: 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 each triangle. The dots that comprise each triangle correspond to specific convex combinations of the vertices, which are the mean, median and mode functionals. The color of the dots 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 the sample size $T = 2000$, the instruments $\boldsymbol{h}_t = (1, X_t)$ and use a Gaussian kernel.

S.6 Clustered Covariance Estimator

Figures S.3 to S.6 below are equivalent to Figures 4 to 7 after substituting equation (3.9) with a clustered covariance estimator $\widehat{\Sigma}_T^{CL}$. Let $\phi_{i,t,T}(\theta)$ denote the moment function of individual i at time t. $\mathcal{T}$ denotes the number of waves and n_t the number of observations within wave such that $T = \sum_{t=1}^{\mathcal{T}} n_t$.

$$\widehat{\Sigma}_T^{CL}(\theta) = \frac{1}{T} \sum_{t=1}^{\mathcal{T}} \left(\sum_{i=1}^{n_t} \widehat{\phi}_{i,t,T}(\theta) \right) \left(\sum_{i=1}^{n_t} \widehat{\phi}_{i,t,T}(\theta) \right)^{\top} \tag{S.6.1}$$

Overall, the results are robust to clustering at the time level. While the mean rejection is less pronounced in Figure S.3, the confidence sets are sharper for the subpopulations. In Figure S.4 to S.6 mean rationality is consistently rejected for lower income individuals at the 5% level.

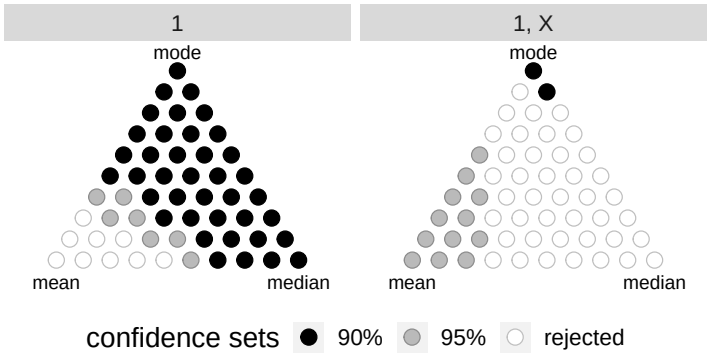


Figure S.3: Confidence sets for income survey forecasts. This figure shows the measures of centrality that rationalize the New York Federal Reserve income survey forecasts with clustered covariance estimator.

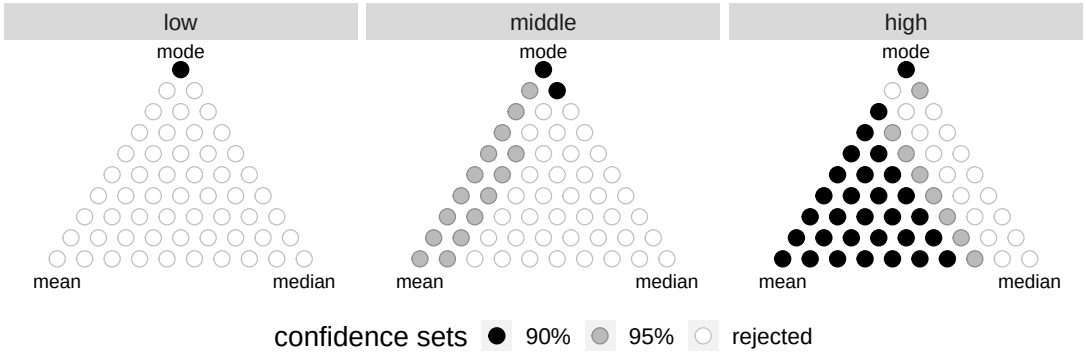


Figure S.4: Confidence sets for income survey forecasts, stratified by income. This figure shows the measures of centrality for low-, middle- and high-income respondents with clustered covariance estimator.

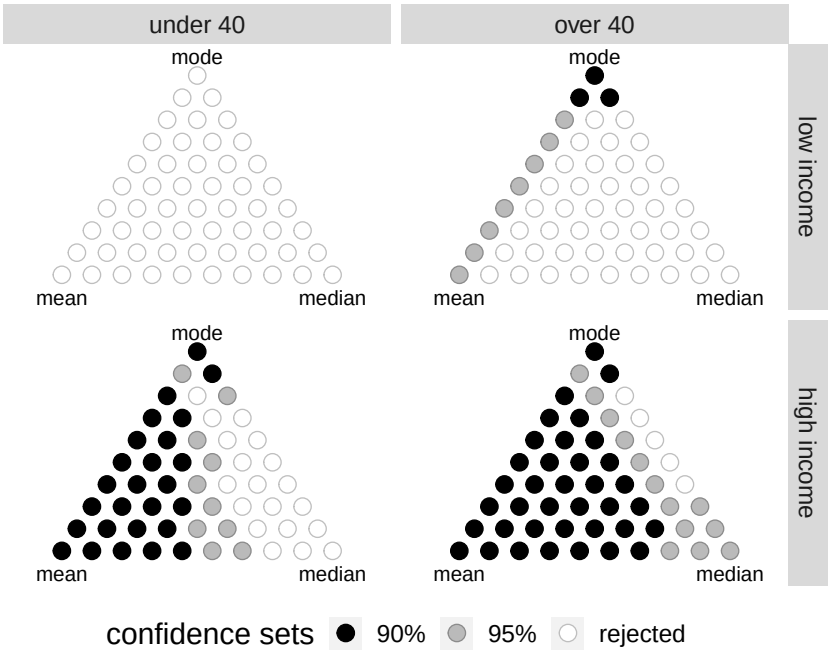


Figure S.5: Confidence sets for income survey forecasts, stratified by income and age. This figure shows the measures of centrality that rationalize the New York Federal Reserve income survey forecasts, for low- and high-income respondents who are below or above the age of 40 with clustered covariance estimator.

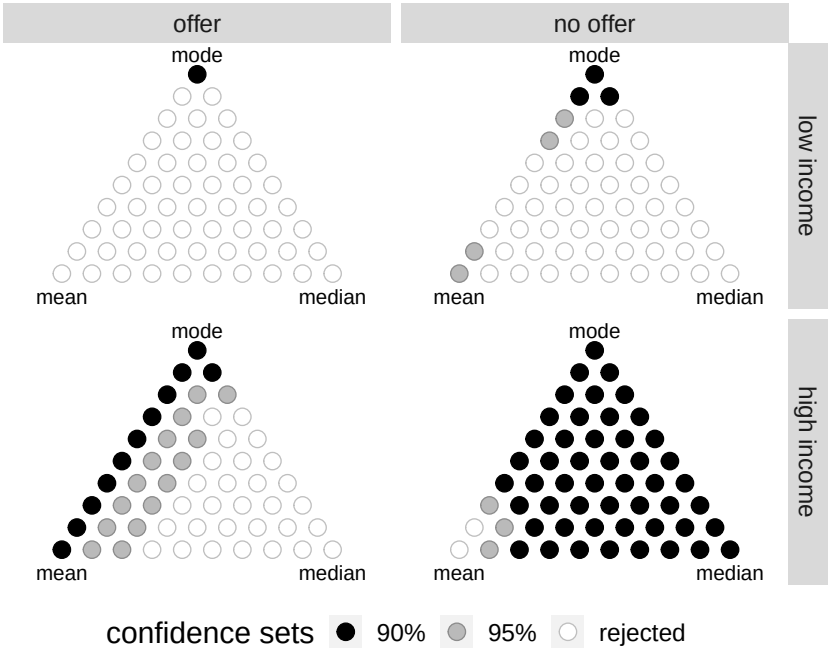


Figure S.6: Confidence sets for income survey forecasts, stratified by income and job offer. This figure shows the measures of centrality that rationalize the New York Federal Reserve income survey forecasts, for low- and high-income respondents in the private sector or not with clustered covariance estimator.

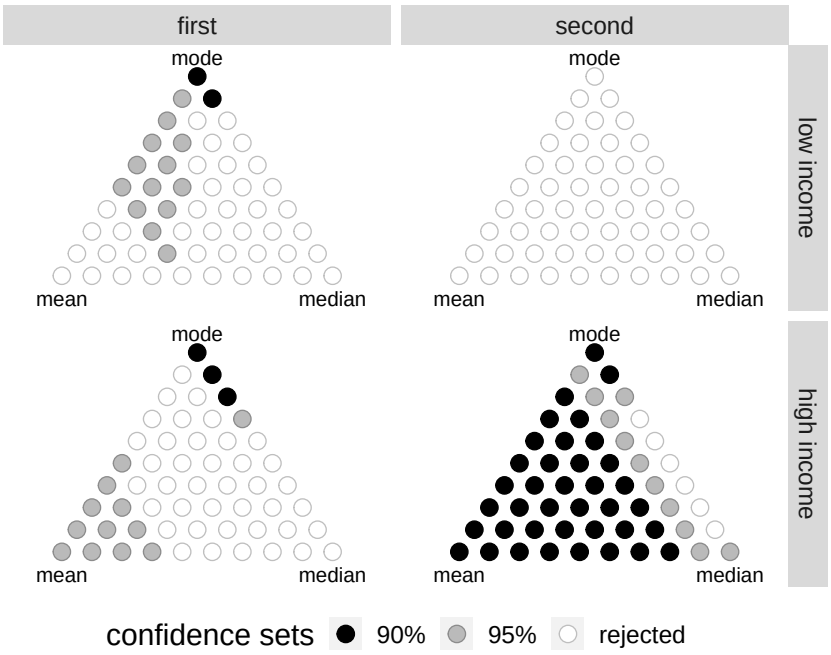


Figure S.7: Confidence sets for income survey forecasts, stratified by income and elicitation round. This figure shows the measures of centrality that rationalize the New York Federal Reserve income survey forecasts, for low- and high-income respondents in their first and second panel round with clustered covariance estimator.

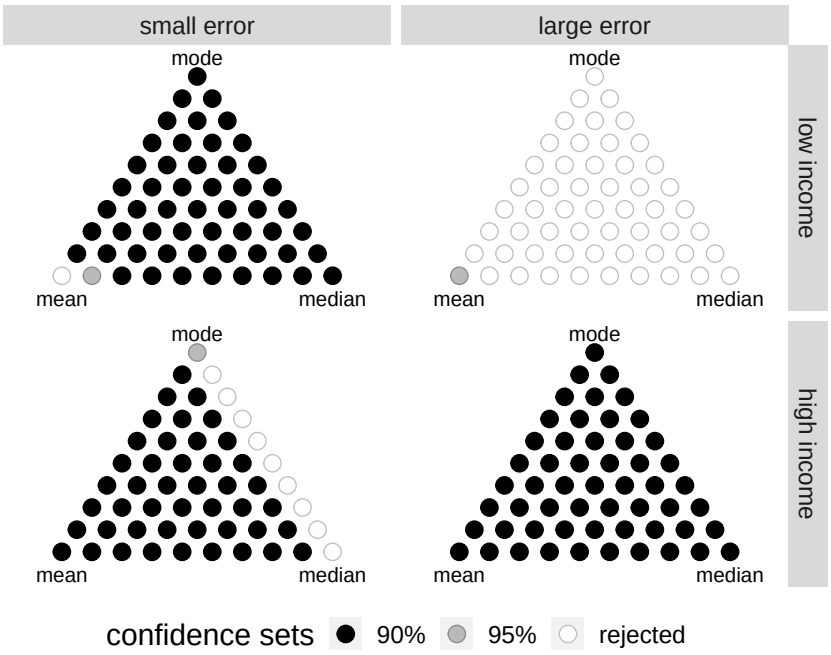


Figure S.8: Confidence sets for income survey forecasts, stratified by income and shock size. This figure shows the measures of centrality that rationalize the New York Federal Reserve income survey forecasts, for low- and high-income respondents with small and large income shocks with clustered covariance estimator.

S.7 Additional results for the Survey of Consumer Expectations

Table S.5 shows the test results for the subsamples considered in Section 5 that were omitted in Table 4.

Table S.5: Summary of p-values for rationality tests in different samples

	n	Mean	Median	Mode	Quantiles	Expectiles
Offer, low income	429	0.01	0.00	0.64	0.00 [0.63, 0.71]	0.01 [0.50, 0.70]
No offer, low income	1471	0.11	0.00	0.95	0.00 [0.57, 0.61]	0.05 [0.42, 0.53]
Offer, high income	328	0.20	0.00	0.70	0.03 [0.63, 0.71]	0.08 [0.40, 0.63]
No offer, high income	1560	0.11	0.10	0.96	0.11 [0.62, 0.66]	0.04 [0.43, 0.54]
First round, low income	1263	0.03	0.00	0.84	0.00 [0.58, 0.62]	0.03 [0.50, 0.60]
Second round, low income	638	0.05	0.00	0.06	0.00 [0.59, 0.65]	0.03 [0.32, 0.51]
First round, high income	1238	0.09	0.00	0.42	0.03 [0.63, 0.68]	0.03 [0.45, 0.57]
Second round, high income	650	0.52	0.20	0.36	0.29 [0.60, 0.66]	0.28 [0.40, 0.56]
Large error, low income	376	0.08	0.00	0.00	0.00 [0.53, 0.61]	0.03 [0.34, 0.56]
Small error, low income	262	0.04	0.30	0.76	0.50 [0.62, 0.72]	0.46 [0.17, 0.43]
Large error, high income	267	0.87	0.72	0.49	0.95 [0.53, 0.63]	0.73 [0.36, 0.58]
Small error, high income	383	0.56	0.08	0.10	0.03 [0.62, 0.70]	0.29 [0.41, 0.60]

Notes: The first four columns of this table present the sample size and p-values from tests of rationality when interpreting the point forecasts as forecasts of the mean, median, or mode. The last two columns present p-values from tests of rationality when interpreting the point forecasts as quantiles or expectiles following Elliott et al. (2005), and 90% confidence intervals for the asymmetry parameter, given in square brackets.

S.8 Additional Empirical Applications

In this section, we present two additional economic applications, to "Greenbook" forecasts of US GDP growth, and to random walk forecasts of exchange rates. In some of this applications we find evidence against mode rationality but not against mean rationality. This confirms that our proposed mode forecast rationality test has non-trivial power in relevant applications.

S.8.1 Greenbook forecasts of U.S. GDP growth

First, we consider one-quarter-ahead forecasts of U.S. GDP growth produced by the staff of the Board of Governors of the Federal Reserve (the so-called "Greenbook" forecasts), from 1967Q2 until 2015Q2, a total of 192 observations.[1] These forecasts are prepared in preparation for each meeting of the Federal Open Market Committee, and substantial resources are devoted to them, see e.g. Romer and Romer (2000). Greenbook

[1] Greenbook forecasts are only available to the public after a five-year lag.

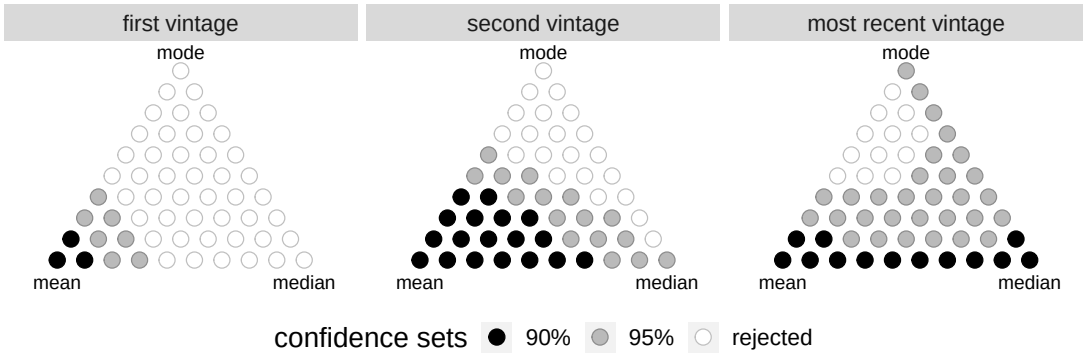


Figure S.9: Confidence sets for Greenbook GDP forecasts. This figure shows the measures of centrality that "rationalize" the Federal Reserve Board's "Greenbook" forecasts of U.S. GDP growth. The three panels use three different measures of GDP growth in a given quarter. Black dots indicate that the measure is inside the Stock-Wright 90% confidence set, grey dots indicate that the measure is inside the 95% confidence set, and white dots indicate that rationality for that measure of centrality can be rejected at the 5% level. All panels use a constant and the forecast as test instruments.

forecasts are available several times each quarter; for this analysis we take the single forecast closest to the middle date in each quarter. Broadly similar results are found when using the first, or last, forecast within each quarter.

Figure S.9 presents the confidence set for the measures of centrality that can be rationalized for these forecasts. As GDP growth is measured with error and official values are often revised, we present results for three different "vintages" of the realized value: the first, second and most recent release. For the first and second vintages, we see that only measures of centrality "close to" the mean can be rationalized as optimal, while the mode, median and similar measures can all be rejected. This is particularly noteworthy given the known lower power at the mode vertex. Using the most recent vintage for GDP growth, both the mean and median, and centrality measures between and near those, are included in the confidence set. That the Greenbook GDP forecasts are rational when interpreted as mean forecasts, but not when taken as mode or median forecasts, is consistent with the Fed staff using econometric models for these forecasts, as such models almost invariably focus on the mean.[2]

S.8.2 Random walk forecasts of exchange rates

For our final empirical application we revisit the famous result of Meese and Rogoff (1983), that exchange rate movements are approximately unpredictable when evaluated by the squared-error loss function, implying

[2] Reifschneider and Tulip (2019) discuss the ambiguity in the specific centrality measure reported in the Greenbook forecasts, but write that they are "typically viewed as modal forecasts" by the Federal Reserve staff. Our results suggest that they are better interpreted as mean forecasts.

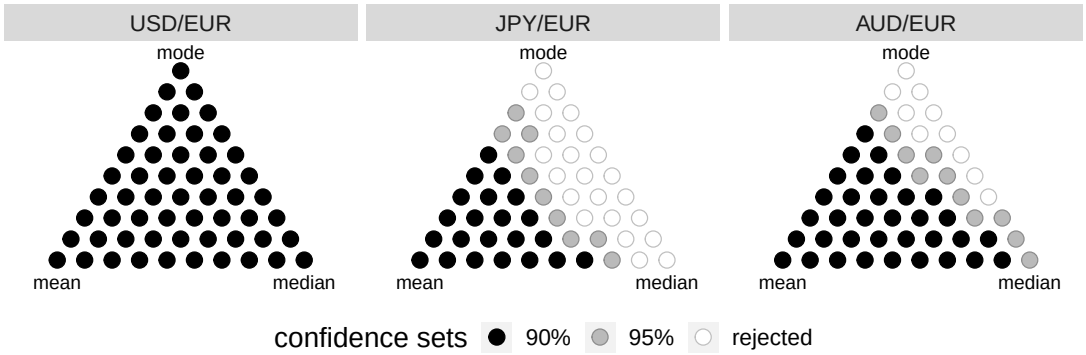


Figure S.10: Confidence sets for random walk forecasts of exchange rates. This figure shows the measures of centrality that "rationalize" the random walk forecast of daily exchange rates movements. Black dots indicate that the measure is inside the Stock-Wright 90% confidence set, grey dots indicate that the measure is inside the 95% confidence set, and white dots indicate that rationality for that measure of centrality can be rejected at the 5% level. All panels use a constant and the forecast as test instruments.

that the lagged exchange rate is an optimal mean forecast. See Rossi (2013) for a more recent survey of the literature on forecasting exchange rates. We use daily data from the European Central Bank's "Statistical Data Warehouse" on the USD/EUR, JPY/EUR and AUD/EUR exchange rates, over the period January 2000 to July 2020, a total of $5,265$ trading days. Note that our sample period has no overlap with that of Meese and Rogoff (1983), and so their conclusions about the mean-optimality of the random walk forecast need not hold in our data.

Figure S.10 presents the results of our tests for rationality, all of which use a constant and the forecast as the instrument set. The middle and right panels reveal that for the JPY/EUR and AUD/EUR exchange rates the lagged exchange rate is not rejected as a mean forecast, while it is rejected when taken as a mode or median forecast. Thus the rationality of the random walk forecast critically depends, for these exchange rates, on whether it is interpreted as a mean, median or mode forecast. For the USD/EUR exchange rate we cannot reject rationality with respect to *any* of convex combination of these measures of central tendency, implying that the random walk forecast is consistent with rationality under any of these measures.[3] The mean vertex being included in the confidence set for all three exchange rates, indicating no evidence against rationality of the random walk model when interpreted as a mean forecast, is consistent with the conclusion of Meese and Rogoff (1983).

[3]Results for the GBP/EUR and CAD/EUR exchange rates are identical to those for the USD/EUR.

Table S.6: Summary of p-values for rationality tests in different samples

Data	n	Mean	Median	Mode	Quantiles	Expectiles
FRBNY income forecasts	3916	0.01	0.00	0.77	0.00 [0.61, 0.64]	0.01 [0.49, 0.56]
Greenbook GDP forecasts	192	0.35	0.06	0.03	0.02 [0.46, 0.57]	0.20 [0.45, 0.61]
Random walk: USD/EUR	5264	0.94	0.32	0.17	0.35 [0.48, 0.51]	0.86 [0.48, 0.51]
JPY/EUR	5264	0.91	0.02	0.04	0.43 [0.47, 0.50]	0.74 [0.48, 0.51]
AUD/EUR	5264	0.88	0.09	0.03	0.96 [0.51, 0.53]	0.62 [0.48, 0.52]

Notes: The first four columns of this table present the sample size and p-values from tests of rationality when interpreting the point forecasts as forecasts of the mean, median, or mode. The last two columns present p-values from tests of rationality when interpreting the point forecasts as quantiles or expectiles following Elliott et al. (2005), and 90% confidence intervals for the asymmetry parameter are given in square brackets. The first row uses the FRBNY consumer survey data introduced in Section 5 of the main paper. The second row uses Federal Reserve "Greenbook" forecasts of US GDP growth discussed in Section S.8.1. The last three rows uses daily data on exchange rates, discussed in Section S.8.2.

S.8.3 Summary of results across empirical applications

Table S.6 summarizes the rationality tests for the mean, median, and mode functionals, as well as the Elliott et al. (2005) rationality test that allows for optimism and pessimism. For the Federal Reserve Bank of New York (FRBNY) survey data, mode rationality is not rejected, but all other functionals are rejected. For the Greenbook GDP forecasts, the mode is rejected a the 5% level, but the mean and expectiles close to the mean are consistent with the data. For the exchange rates random walk forecasts, the mean and central quantiles and expectiles are consistent with the data, while mode rationality is rejected for two of the three exchange rates at the 5% level.