Testing Forecast Rationality for Measures of Central Tendency

Timo Dimitriadis[1,2]

[1]Heidelberg University, Germany

[2]Heidelberg Institute for Theoretical Studies, Germany

Revisiting and Improving Prediction Tools for Central Banks

February 25, 2022

based on the working paper:

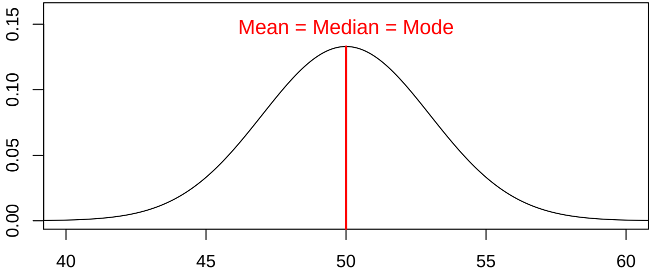
Dimitriadis, T., Patton A.J., and Schmidt, P. (2019). Testing Forecast Rationality for Measures of Central Tendency. arXiv: 1910.12545 [econ.EM].

Private Income Forecasts

- SCE survey question:

  "What do you **believe** your annual earnings will be in 4 months?"

- Assume your **beliefs** are summarized by this distribution



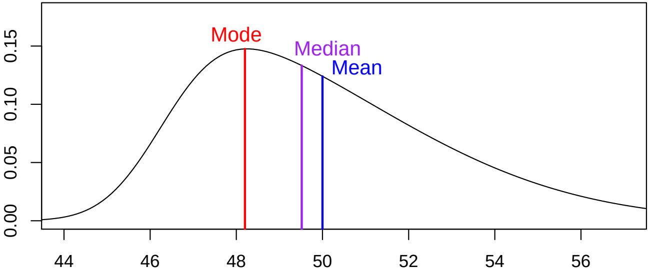
- **Research Question: Which *measure* do rational respondents report?**

Private Income Forecasts

- SCE survey question:

  "What do you **believe** your annual earnings will be in 4 months?"

- Assume your **beliefs** are summarized by this distribution



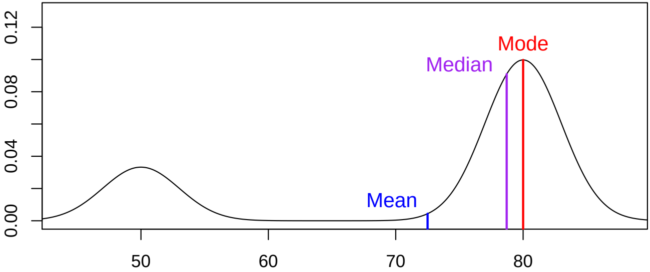
- **Research Question: Which *measure* do rational respondents report?**

Private Income Forecasts

- SCE survey question:

  "What do you **believe** your annual earnings will be in 4 months?"

- Assume your **beliefs** are summarized by this distribution



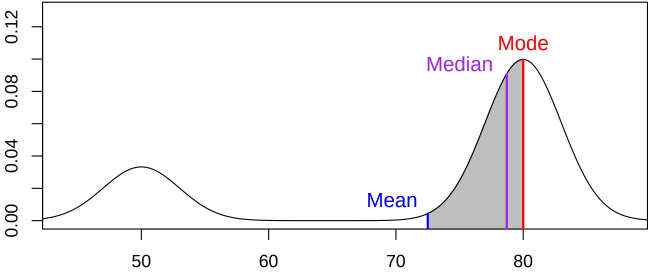
- **Research Question: Which *measure* do rational respondents report?**

Private Income Forecasts

- SCE survey question:

  ”What do you **believe** your annual earnings will be in 4 months?”

- Assume your **beliefs** are summarized by this distribution



- **Research Question: Which *measure* do rational respondents report?**
- Reasonable answers: the **mean**, **median**, **mode** or **anything in between**.

Economic Examples

(1) **SCE Labor Market Survey** of the New York Fed survey question:

"What do you **believe** your annual earnings will be in 4 months?"[1]

(2) **Greenbook GDP Forecasts:**

"The staff [...] prepares **projections** about how the economy will fare [...]."[2]

(3) **Survey of Professional Forecasters (SPF):**

"The forecasts [...] are the forecasters' **projections** [...]"[3]

1 Source: Questionnaire for SCE Labor Market Survey. https://www.newyorkfed.org/medialibrary/media/research/microeconomics/interactive/downloads/sce-labor-questionnaire.pdf

2 Source: https://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data

3 Source: Croushore, D. and Stark, T. (2019) Fifty Years of the Survey of Professional Forecasters. Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/-/media/frbp/assets/economy/articles/economic-insights/2019/q4/eiq419.pdf

A General Class of Central Tendency Forecasts

- General class of **central tendency measures**: any convex combination of the mean, median, and mode:

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- These are the famous "**three Ms**" of statistics.
- The first two Ms are widely studied, but the third:
  - Dalenius (1965, *JRSSA*): "The Mode – A Neglected Statistical Parameter"
  - New results on loss and identification functions for the mode
  - The mode is the most intuitive but also the most complicated measure.

An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

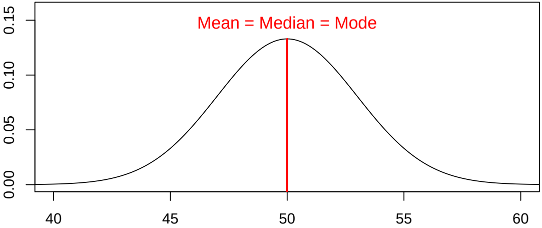
$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- But, the weights can be
  - **un-identified**



An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- But, the weights can be
  - **un-identified**
  - **weakly identified**

An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- But, the weights can be
  - **un-identified**
  - **weakly identified**
  - **strongly identified**

An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- But, the weights can be
  - **un-identified**
  - **weakly identified**
  - **strongly identified**
  - **partially identified**

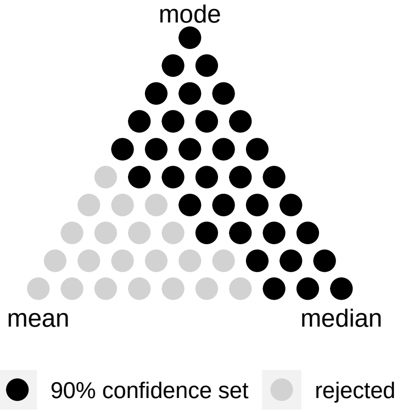
An Identification Problem

- We would like to estimate the weights $\boldsymbol{w}$ in

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}}$$

- But, the weights can be
  - **un-identified**
  - **weakly identified**
  - **strongly identified**
  - **partially identified**
- A valid testing approach must accommodate all these possibilities
- We construct **confidence sets** through **inverted test statistics** from Stock and Wright (2000, *ECMA*)

Illustration of the Confidence Set



A Mode Forecast Rationality Test

A **Mean** Forecast Rationality Tests

- Y_{t+1}: variable of interest
- X_t: 1-step ahead forecast for Y_{t+1}
- $\mathcal{F}_t$: information available to the forecaster at time t
- $\boldsymbol{h}_t \in \mathbb{R}^k$: $\mathcal{F}_t$-measurable vector of instruments

A **Mean** Forecast Rationality Tests

- Y_{t+1}: variable of interest
- X_t: 1-step ahead forecast for Y_{t+1}
- $\mathcal{F}_t$: information available to the forecaster at time t
- $\boldsymbol{h}_t \in \mathbb{R}^k$: $\mathcal{F}_t$-measurable vector of instruments
- Mean forecast rationality:

$$\mathbb{H}_0^* : X_t = \mathbb{E}[Y_{t+1} | \mathcal{F}_t] \quad \text{a.s.} \tag{1}$$

- This is often tested by using instruments $\boldsymbol{h}_t = (1, X_t)$:

$$\mathbb{H}_0 : \mathbb{E}\left[(X_t - Y_{t+1})\, \boldsymbol{h}_t^\top\right] = 0 \tag{2}$$

- Intuition: the forecast error $\varepsilon_t = X_t - Y_{t+1}$ is zero on average

A **Mean** Forecast Rationality Tests

- Y_{t+1}: variable of interest
- X_t: 1-step ahead forecast for Y_{t+1}
- $\mathcal{F}_t$: information available to the forecaster at time t
- $\boldsymbol{h}_t \in \mathbb{R}^k$: $\mathcal{F}_t$-measurable vector of instruments
- Mean forecast rationality:

$$\mathbb{H}_0^* : X_t = \mathbb{E}[Y_{t+1} | \mathcal{F}_t] \quad \text{a.s.} \tag{1}$$

- This is often tested by using instruments $\boldsymbol{h}_t = (1, X_t)$:

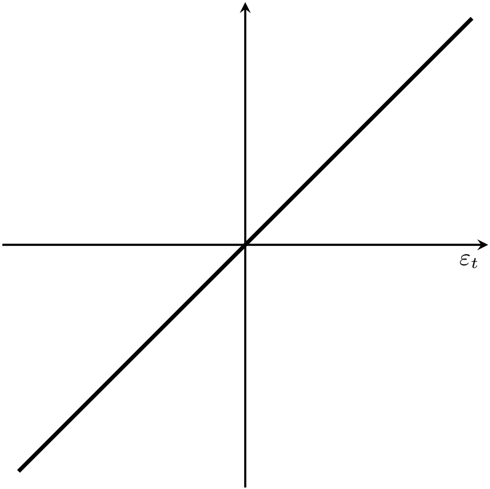
$$\mathbb{H}_0 : \mathbb{E}\left[(X_t - Y_{t+1})\, \boldsymbol{h}_t^\top\right] = 0 \tag{2}$$

- Intuition: the forecast error $\varepsilon_t = X_t - Y_{t+1}$ is zero on average
- Using the **strict identification function** $V_t^{\text{Mean}}(\varepsilon_t) = \varepsilon_t$:

$$\mathbb{H}_0 : \mathbb{E}\left[V_t^{\text{Mean}}(\varepsilon_t)\, \boldsymbol{h}_t^\top\right] = 0 \tag{3}$$

Mean Identification Function

$$V_t^{\text{Mean}}(\varepsilon_t) = \varepsilon_t$$



A **Median** Forecast Rationality Test

- We want to test:

$$\mathbb{H}_0^* : X_t = \text{Median}[Y_{t+1}|\mathcal{F}_t] \quad \text{a.s.} \tag{4}$$

- Instead, we test the necessary hypothesis:

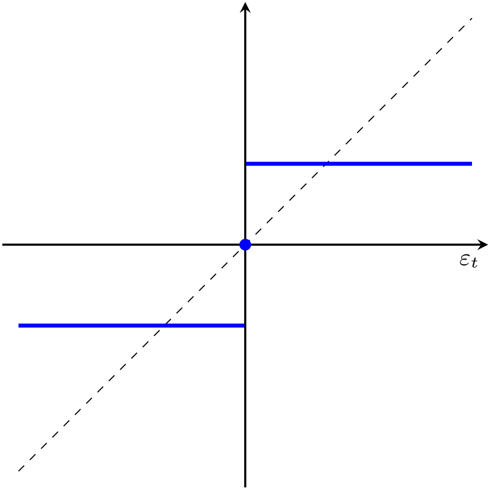
$$\mathbb{H}_0 : \mathbb{E}\left[V_t^{\text{Med}}(\varepsilon_t)\ \boldsymbol{h}_t^\top\right] = 0 \tag{5}$$

- Main difference: **median** identification function:

$$V_t^{\text{Med}}(\varepsilon_t) = \mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}} \tag{6}$$

Median Identification Function

$$V_t^{\text{Med}}(\varepsilon_t) = \mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}}$$



What about the **Mode**?

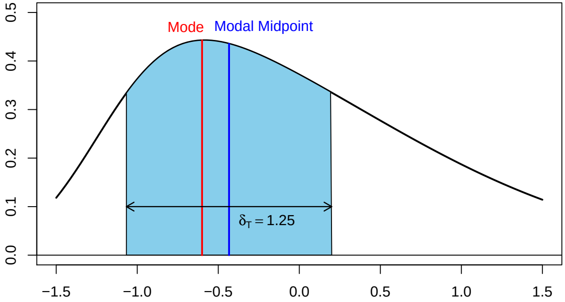
- No identification function exists for the mode (Heinrich, 2014, *Biometrika*)

What about the **Mode**?

- No identification function exists for the mode (Heinrich, 2014, *Biometrika*)
- We consider **asymptotic identifiability** of the mode through the (smoothed) **modal midpoint**

$$\mathrm{MMP}(\delta) = \arg\max_x \mathbb{P}\Big(Y \in [x - \delta/2, x + \delta/2]\Big) \quad (7)$$

- The interval of fixed length δ which contains the highest probability.

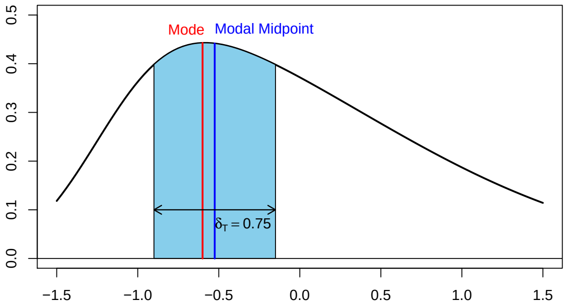


What about the **Mode**?

- No identification function exists for the mode (Heinrich, 2014, *Biometrika*)
- We consider **asymptotic identifiability** of the mode through the (smoothed) **modal midpoint**

$$\mathrm{MMP}(\delta) = \arg\max_x \mathbb{P}\Big(Y \in [x - \delta/2, x + \delta/2]\Big) \tag{7}$$

- The interval of fixed length δ which contains the highest probability.

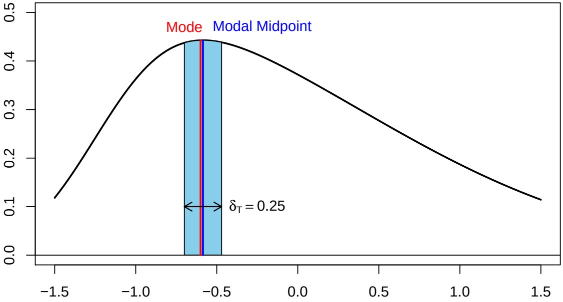


What about the **Mode**?

- No identification function exists for the mode (Heinrich, 2014, *Biometrika*)
- We consider **asymptotic identifiability** of the mode through the (smoothed) **modal midpoint**

$$\mathrm{MMP}(\delta) = \arg\max_x \mathbb{P}\Big(Y \in [x - \delta/2, x + \delta/2]\Big) \tag{7}$$

- The interval of fixed length δ which contains the highest probability.



What about the **Mode**?

- We consider **asymptotic identifiability** of the mode through the (smoothed) **modal midpoint**

$$\text{SMMP}(K, \delta) = \arg\max_x \mathbb{E}\left[\frac{1}{\delta} K\left(\frac{Y - x}{\delta}\right)\right] \tag{8}$$

  - e.g. the Gaussian density function $K(\cdot) = \phi(\cdot)$
  - a rectangular kernel recovers the MMP from the previous slides

What about the **Mode**?

- We consider **asymptotic identifiability** of the mode through the (smoothed) **modal midpoint**

$$\mathrm{SMMP}(K,\delta) = \arg\max_x \mathbb{E}\left[\frac{1}{\delta} K\left(\frac{Y-x}{\delta}\right)\right] \tag{8}$$

  - e.g. the Gaussian density function $K(\cdot) = \phi(\cdot)$
  - a rectangular kernel recovers the MMP from the previous slides

- Asymptotic Identification Function for the smoothed modal midpoint:

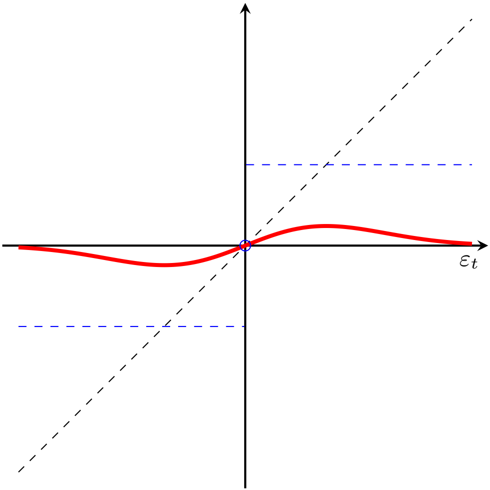
$$V^{\text{Mode}}_{t,\delta_T}(\varepsilon_t) = \frac{1}{\delta_T^2} K'\left(\frac{-\varepsilon_t}{\delta_T}\right) \tag{9}$$

  - Let $\delta_T \to 0$ converge "slowly" with the sample size T

Details on (Asymptotic) Elicitability

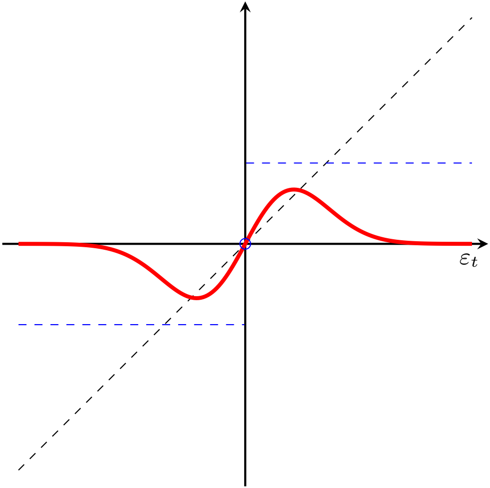
(Asymptotic) Mode Identification Function

$$V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t) = \frac{1}{\delta_T^2} K'\left(\frac{-\varepsilon_t}{\delta_T}\right)$$



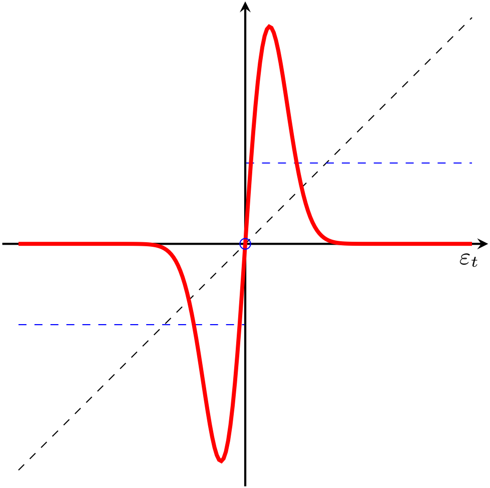
(Asymptotic) Mode Identification Function

$$V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t) = \frac{1}{\delta_T^2} K'\left(\frac{-\varepsilon_t}{\delta_T}\right)$$



(Asymptotic) Mode Identification Function

$$V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t) = \frac{1}{\delta_T^2} K'\left(\frac{-\varepsilon_t}{\delta_T}\right)$$



A (Nonparametric) Mode Forecast Rationality Test

- Recall: We want to test

$$\mathbb{H}_0^* : X_t = \text{Mode}[Y_{t+1}|\mathcal{F}_t] \quad \text{a.s.} \tag{10}$$

- We test instead:

$$\mathbb{H}_0 : \mathbb{E}\left[V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t)\, \boldsymbol{h}_t^\top\right] = 0 \tag{11}$$

A (Nonparametric) Mode Forecast Rationality Test

- Recall: We want to test

$$\mathbb{H}_0^* : X_t = \text{Mode}[Y_{t+1}|\mathcal{F}_t] \quad \text{a.s.} \tag{10}$$

- We test instead:

$$\mathbb{H}_0 : \mathbb{E}\left[V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t)\, \boldsymbol{h}_t^\top\right] = 0 \tag{11}$$

Theorem 1

Under $\mathbb{H}_0^*$ and given Assumption 1, it holds that

$$\delta_T^{3/2} T^{-1/2} \sum_{t=1}^{T-1} V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t)\, \boldsymbol{h}_t^\top \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mode}}), \tag{12}$$

and

$$J_T = \frac{1}{T\delta_T^3} \left(\sum_t V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t)\, \boldsymbol{h}_t^\top\right) \widehat{\Omega}_{T,\text{Mode}}^{-1} \left(\sum_t V_{t,\delta_T}^{\text{Mode}}(\varepsilon_t)\, \boldsymbol{h}_t\right) \xrightarrow{d} \chi_k^2. \tag{13}$$

- Use the test statistic J_T and compare it to χ_k^2 critical values.

Assumption 1 | Bandwidth Choice | Covariance Estimation | Test Power

Forecasts of Central Tendency

Identifying the Measure of Central Tendency

- Test whether forecasts are rational for **some measure of central tendency**.
- Central tendency: convex combination of the **functionals**:

$$X_t^{\text{central}}(\boldsymbol{w}) = w_1 X_t^{\text{mean}} + w_2 X_t^{\text{median}} + w_3 X_t^{\text{mode}} \tag{14}$$

- Convex combinations not elicitable/identifiable. What now?

Identifying the Measure of Central Tendency

- **Solution:** Convex combination of **identification functions**:

$$\phi_{t,T}(\theta) = \boldsymbol{h}_t \ \theta^\top \begin{pmatrix} \widehat{\mathbf{w}}_{T,\mathrm{Mean}} & \varepsilon_t \\ \widehat{\mathbf{w}}_{T,\mathrm{Med}} & \left(\mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}}\right) \\ \widehat{\mathbf{w}}_{T,\mathrm{Mode}} & \delta_T^{-1/2} K' \left(\frac{-\varepsilon_t}{\delta_T}\right) \end{pmatrix}, \qquad (15)$$

  where $\widehat{\mathbf{w}}_{T,\bullet}$ are possibly estimated weights.

- Motivated by a forecaster minimizing a convex combination of loss functions.

- Formally, we test the null hypothesis:

$$\mathbb{H}_0 : \quad \exists\, \theta_0 \in \Theta \quad \text{s.t.} \quad \lim_{T \to \infty} \mathbb{E}\left[\phi_{t,T}(\theta_0)\right] = 0. \qquad (16)$$

Infeasible Estimation of the Combination Weights

- The GMM objective function:

$$S_T(\theta) = \left[T^{-1/2} \sum_{t=1}^{T-1} \phi_{t,T}(\theta)\right]^\top \widehat{\Sigma}_T^{-1}(\theta) \left[T^{-1/2} \sum_{t=1}^{T-1} \phi_{t,T}(\theta)\right]$$

- Think if $\phi_{t,T}(\theta)$ as a moment condition in the GMM context
- Intuition: try to find a value of θ such that $T^{-1/2} \sum_t \phi_{t,T}(\theta)$ is close to zero
- The standard approach: estimate θ by minimization of $S_T(\theta)$

Infeasible Estimation of the Combination Weights

- The GMM objective function:

$$S_T(\theta) = \left[T^{-1/2} \sum_{t=1}^{T-1} \phi_{t,T}(\theta)\right]^\top \widehat{\Sigma}_T^{-1}(\theta) \left[T^{-1/2} \sum_{t=1}^{T-1} \phi_{t,T}(\theta)\right]$$

- Think if $\phi_{t,T}(\theta)$ as a moment condition in the GMM context
- Intuition: try to find a value of θ such that $T^{-1/2} \sum_t \phi_{t,T}(\theta)$ is close to zero
- The standard approach: estimate θ by minimization of $S_T(\theta)$
- θ might be strongly, weakly, partially or un-identified
- But: We cannot consistently estimate θ

Infeasible Estimation of the Combination Weights

- The GMM objective function:

$$S_T(\theta) = \left[T^{-1/2} \sum\nolimits_{t=1}^{T-1} \phi_{t,T}(\theta)\right]^\top \widehat{\Sigma}_T^{-1}(\theta) \left[T^{-1/2} \sum\nolimits_{t=1}^{T-1} \phi_{t,T}(\theta)\right]$$

- Think if $\phi_{t,T}(\theta)$ as a moment condition in the GMM context
- Intuition: try to find a value of θ such that $T^{-1/2} \sum_t \phi_{t,T}(\theta)$ is close to zero
- The standard approach: estimate θ by minimization of $S_T(\theta)$
- θ might be strongly, weakly, partially or un-identified
- But: We cannot consistently estimate θ
- We get asymptotically valid confidence sets through **inverted test statistics** (Stock and Wright, 2000, *ECMA*)

Confidence Sets for the Combination Weights

Theorem 2

Given the null hypothesis,

$$\mathbb{H}_0 : \quad \exists\, \theta_0 \in \Theta \text{ s.t. } \lim_{T \to \infty} \mathbb{E}\left[\phi_{t,T}(\theta_0)\right] = 0, \tag{17}$$

and given Assumption 2, for the **true, but unknown** parameter(s) θ_0,

$$S_T(\theta_0) \xrightarrow{d} \chi^2_k. \tag{18}$$

Confidence Sets for the Combination Weights

Theorem 2

Given the null hypothesis,

$$\mathbb{H}_0 : \quad \exists\, \theta_0 \in \Theta \text{ s.t. } \lim_{T \to \infty} \mathbb{E}\left[\phi_{t,T}(\theta_0)\right] = 0, \tag{17}$$

and given Assumption 2, for the **true, but unknown** parameter(s) θ_0,

$$S_T(\theta_0) \xrightarrow{d} \chi^2_k. \tag{18}$$

Corollary

Given Assumption 2, the set

$$\widehat{\Theta}^*_T = \left\{\theta \in \Theta : \quad S_T(\theta) \leq \chi_k^{(-1)}(1-\alpha)\right\} \tag{19}$$

is an asymptotically valid $100(1-\alpha)\%$ **confidence set**.

Assumption 2 Confidence Set Power

Illustration of the Results

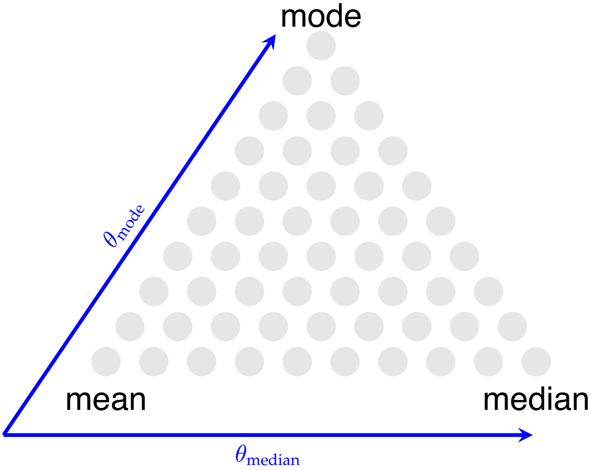


Illustration of the Results

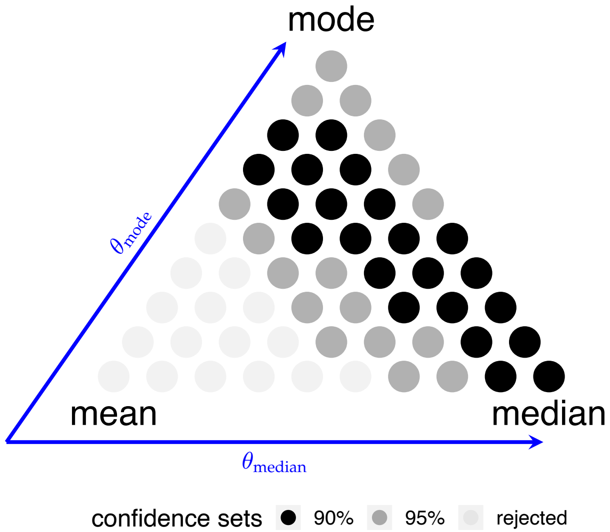
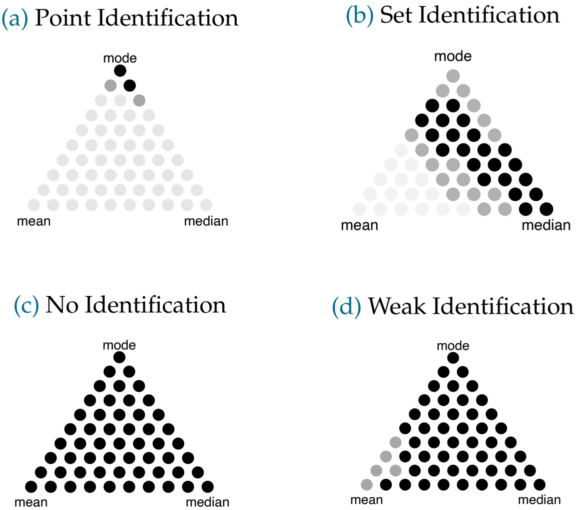


Illustration of the Identification Problem



Simulation Study

- Both our methods work well in finite samples.

Simulation Study Mode Rationality Test
Simulation Study Measure of Centrality

Applications

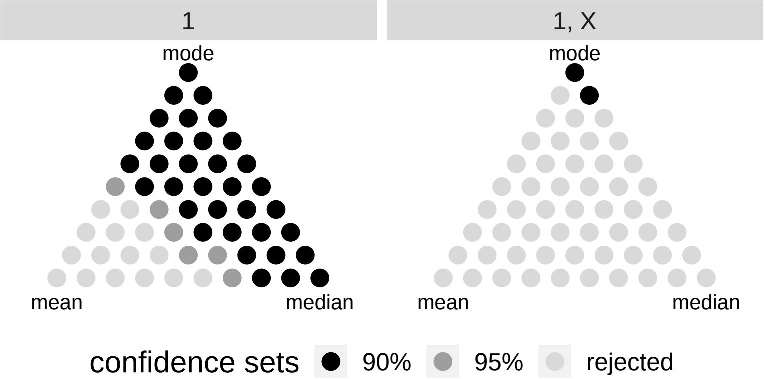
Application I: SCE Income Forecasts

- Data from the *Survey of Consumer Expectations* of the New York Fed.
- $T = 3,916$ responses from more than $2,000$ individuals; interviewed between March 2015 and March 2018.
- Survey question:

  "What do you **believe** your annual earnings will be in 4 months?"

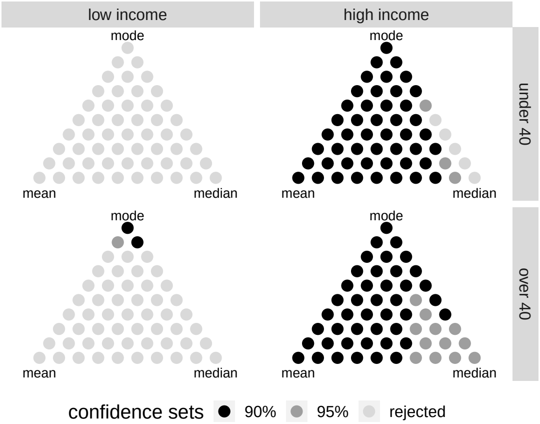
- They also report their realized income 4 months later.

SCE Income Forecasts



- best rationalized as **mode** forecasts, i.e. as the **most likely outcome**.

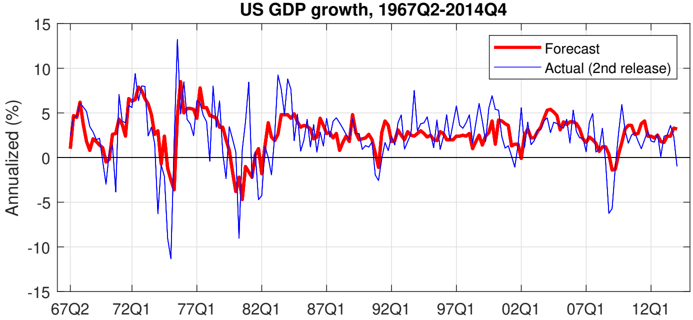
SCE Income: Stratified by Past Income and Age



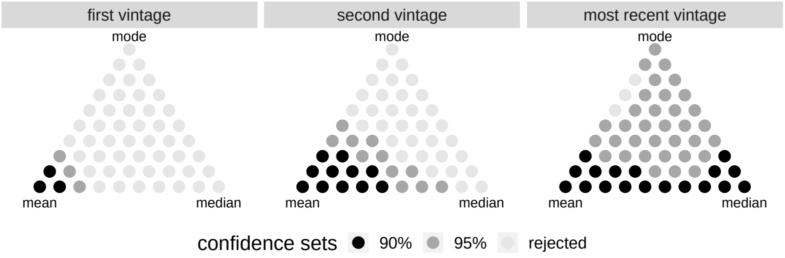
- Low Income & under 40 respondents **cannot rationally forecast their income**.
- More stratifications in the paper.

Application II: GDP Greenbook Forecasts

- We use one-quarter-ahead *Greenbook* forecasts of US GDP growth produced by the Fed Board
- The sample is from 1967Q2 to 2014Q1; $T = 187$ observations



GDP Greenbook Forecasts

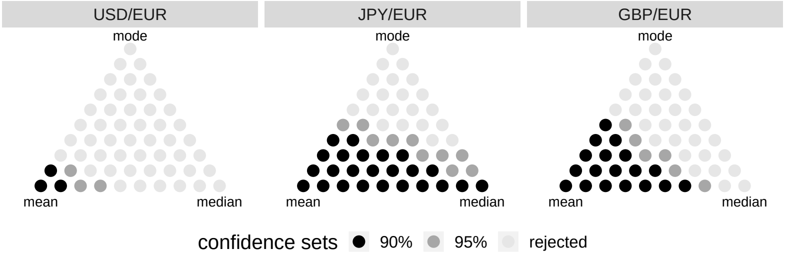


- always rationalizable as **mean forecasts**
- in the last vintage also as median forecasts

Application III: No-Change Exchange Rate Forecasts

- Meese and Rogoff (1983, *JIE*): exchange rate changes are unpredictable when using the **squared-error loss** function.
- This implies that the lagged exchange rate $X_t = Y_t$ is a rational (optimal) **mean** forecast for Y_{t+1}.
- Are the lagged exchange rates also rational median and mode forecasts?
- We use daily data on USD/EUR, JPY/EUR and GBP/EUR exchange rates from May 2000 to June 2019; $T = 4,978$ observations.
- We use $\boldsymbol{h}_t = (1, Y_t - Y_{t-1})$ as instruments for stationarity reasons.

Random Walk Forecasts of Exchange Rates



- no-change exchange rate forecasts are rationalizable as **mean forecasts**.

Conclusion

- Economic surveys and forecasting tasks generally request a **point forecast**:
  - they are vague about the specific quantity to be reported
  - reasonable respondents can interpret this request in many ways

Conclusion

- Economic surveys and forecasting tasks generally request a **point forecast**:
  - they are vague about the specific quantity to be reported
  - reasonable respondents can interpret this request in many ways
- We propose new methods to test forecast rationality of some unknown measure of central tendency:
  - we use a convex combination nesting the mean, median, and mode
  - we establish asymptotic elicitability/identifiability of the mode
  - we overcome an inherent identification problem
  - our method provides confidence sets of rational centrality measures

Conclusion

- Economic surveys and forecasting tasks generally request a **point forecast**:
  - they are vague about the specific quantity to be reported
  - reasonable respondents can interpret this request in many ways
- We propose new methods to test forecast rationality of some unknown measure of central tendency:
  - we use a convex combination nesting the mean, median, and mode
  - we establish asymptotic elicitability/identifiability of the mode
  - we overcome an inherent identification problem
  - our method provides confidence sets of rational centrality measures
- Three applications: micro, macro, finance

Thank you for your attention!

References I

Chen, X., Christensen, T. M., and Tamer, E. (2018). Monte carlo confidence sets for identified sets. *Econometrica*, 86(6):1965–2018.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models1. *Econometrica*, 75(5):1243–1284.

Dalenius, T. (1965). The mode–a neglected statistical parameter. *Journal of the Royal Statistical Society. Series A (General)*, 128(1):110–117.

Eddy, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.*, 8(4):870–882.

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72(4):1107–1125.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4):1680–1707.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, 101(1):245–251.

References II

Kemp, G. C., Parente, P. M., and Silva, J. S. (2019). Dynamic vector mode regression. *Journal of Business & Economic Statistics*. forthcoming.

Kemp, G. C. and Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, 170(1):92 – 101.

Kröger, S. and Pierrot, T. (2019). What point of a distribution summarises point predictions? WZB Discussion Paper, No. SP II 2019-212, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin, http://hdl.handle.net/10419/206533.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14:3–24.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076.

Patton, A. J. and Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480):1172–1184.

Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.*, 16(2):629–647.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5):1055–1096.

Some Related Work

- **Mode estimation:** Parzen (1962, *AMS*), Eddy (1980, *AoS*), Romano (1988, *AoS*), Kemp and Santos Silva (2012, *JoE*), Kemp et al. (2019, *JBES*)
  - we draw on this for implementing mode forecast rationality test
- **Mode forecasting and elicitability**: Gneiting (2011, *JASA*), Heinrich (2014, *Biometrika*), Fissler and Ziegel (2016, *AoS*)
  - we draw on these in defining asymptotic elicitibility of the mode
- **Survey forecasts and responses:** Manski (2004, *ECMA*), Engleberg, Manski and Williams (2009, *JBES*), Kröger and Peirrot (2019, *WP*)
  - motivates our consideration of multiple measures of central tendency
- **Rationality under unknown loss:** Elliott, Komunjer and Timmermann (2005, *REStud*) (EKT), Patton and Timmermann (2007, *JASA*)
  - like EKT we nest the mean & median as special cases
  - unlike EKT we consider central tendency measures

Elicitability and Identifiability

- A statistical functional Γ is said to be **elicitable** (Gneiting, 2011, *JASA*) if, for $Y \sim F$, there exists a **loss function** L where

$$\Gamma(F) = \arg\min_x \mathbb{E}\left[L(x, Y)\right]$$

- Alternatively, it is **identifiable** if there exists an **identification function** V such that

$$0 = E\left[V(x, Y)\right] \quad \Leftrightarrow \quad x = \Gamma(F)$$

  - one can think of $V(x, Y)$ as $\partial L(x, Y) / \partial x$
  - for the mean, $V(x, Y) = (x - Y) =$ the forecast error
  - for the median, $V(x, Y) = \mathbf{1}\{Y > x\} - \mathbf{1}\{Y < x\}$

- If the forecast is elicitable, then a natural way to test rationality of forecast x is via the first-order condition:

$$0 = E\left[V(x, Y) \cdot \mathbf{h}\right] \text{ for } \mathbf{h} \in \mathcal{F}$$

Asymptotic Elicitability of the Mode

The mode is asymptotically elicitable

Definition: A functional Γ is *asymptotically elicitable* relative to the class of distributions $\mathcal{P}$ if there exists a sequence of elicitable functionals Γ_k such that $\Gamma_k(P) \to \Gamma(P)$ for all $P \in \mathcal{P}$.

Proposition: Let $\mathcal{P}$ denote the class of distributions consisting of absolutely continuous unimodal distributions with bounded density and assume K is is positive, smooth, log-concave, and $\int K(u)\,du = 1$. Then

$$\Gamma_\delta = \arg\min_x \; E\left[\frac{1}{\delta} K\left(\frac{Y-x}{\delta}\right)\right]$$

is well defined for all $\delta > 0$, and

$$\Gamma_\delta \to \operatorname{Mode}(P) \;\text{ as } \delta \to 0 \text{ for all } P \in \mathcal{P}.$$

Asymptotic Elicitability of the Mode

The Relationship to Mincer-Zarnowitz Tests

- Mincer and Zarnowitz (1969) propose to fit the (mean) regression

$$Y_{t+1} = \alpha + \beta X_t + \varepsilon_t, \quad \text{where} \quad \mathbb{E}[\varepsilon_t | \mathcal{F}_t] = 0, \tag{20}$$

  and to test $\mathbb{H}_0 : (\alpha, \beta) = (0, 1)$.

- Corresponding first-order moment conditions

$$\mathbb{E}\left[(X_t - Y_{t+1}) \cdot (1, X_t)^\top \right] = 0. \tag{21}$$

- This is a special case of rationality tests for $\boldsymbol{h}_t = (1, X_t)$.
- A similar MZ-test for the mode could be used by using a *mode regression* of Kemp and Santos Silva (2012, *JoE*) and Kemp et al. (2019, *JBES*)

Assumption 1

(A1) $\left(\varepsilon_t, \boldsymbol{h}_t\right)$ *is a stationary and ergodic sequence*

(A2) $\mathbb{E}\left[||\boldsymbol{h}_t||^{2+\delta}\right] < \infty$

(A3) *the matrix* $\mathbb{E}\left[\boldsymbol{h}_t \boldsymbol{h}_t^\top\right]$ *has full rank*

(A4) *the conditional distribution of* $\varepsilon_t = X_t - Y_{t+1}$ *given* $\mathcal{F}_t$ *is absolutely continuous with three times continuously differentiable density* $f_t(\cdot)$ *and bounded derivatives.*

(A5) $K : \mathbb{R} \to \mathbb{R}$, $u \mapsto K(u)$ *is a continuously differentiable kernel function such that: (i)* $\int K(u)\mathrm{d}u = 1$, *(ii)* $\sup K(u) \leq c < \infty$, *(iii)* $\sup K'(u) \leq c < \infty$, *(iv)* $\int uK(u)\mathrm{d}u = 0$, *(v)* $\int u^2 K(u)\mathrm{d}u = c < \infty$, *(vi)* $\int \left(K'(u)\right)^2 \mathrm{d}u = M < \infty$.

(A6) δ_T *is a strictly positive and non-stochastic bandwidth such that for* $T \to \infty$, *(i)* $T\delta_T \to \infty$, *(ii)* $T\delta_T^7 \to 0$

Mode Rationality Test

Bandwidth Selection

- Theory requires $\delta_T \propto T^{-a}$ with $a \in (1/7, 1)$. Optimal rate if $a = 1/7 + \epsilon$.
- Kemp et al. (2019, *JBES*) used the rule:

$$\delta_T = 3.2 \times \widehat{MAD}\left[|X_t - Y_{t+1}|\right] \times T^{-0.143}$$

- We found we needed to generalize this to allow the bandwidth to vary with the degree of **skewness**: bandwidth needs to shrink for skewed data
- We use

$$\delta_T = 3.2 \times \exp\left\{-3.2\,|\hat{\gamma}|\right\} \times \widehat{MAD}\left[|X_t - Y_{t+1}|\right] \times T^{-0.143}$$

where

$$\hat{\gamma} = 3\frac{\hat{E}\left[X_t - Y_{t+1}\right] - \widehat{Median}\left[X_t - Y_{t+1}\right]}{\widehat{Var}\left[X_t - Y_{t+1}\right]^{1/2}}$$

is Pearson's *second coefficient of skewness*.

Mode Rationality Test

Covariance Estimation

- Asymptotic Covariance

$$\Omega_{Mode} = \mathbb{E}\left[\boldsymbol{h}_t \boldsymbol{h}_t^\top f_t(0)\right] \int K'(u)^2 \mathrm{d}u \tag{22}$$

- Covariance Estimator (following Kemp and Santos Silva (2012, *JoE*) and Kemp et al. (2019, *JBES*))

$$\widehat{\Omega}_{T,Mode} = \frac{1}{T} \sum_{t=1}^{T} \delta_T^{-1} K' \left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 \boldsymbol{h}_t \boldsymbol{h}_t^\top \tag{23}$$

Theorem 3

Under Assumption 1, it holds that

$$\widehat{\Omega}_{T,Mode} \xrightarrow{P} \Omega_{Mode}. \tag{24}$$

- The matrix (22) is known from the asymptotic theory of quantile regression.
- Engle and Manganelli (2004, *JBES*) use an indicator function instead of $K(\cdot)$.

Mode Rationality Test

Test Power

- Our mode rationality test has power against the general alternative

$$\mathbb{H}_A : \mathbb{E}\left[f_t'(0)\boldsymbol{h}_t\right] \neq 0 \quad \text{for all } t = 1, \dots, T. \tag{25}$$

Theorem 4

Assume that Assumption 1 holds and that $T\delta_T^3 \to \infty$. Then, under the alternative hypothesis $\mathbb{H}_A$, it holds that

$$\mathbb{P}\left(J_T \geq c\right) \to 1 \tag{26}$$

for any constant $c \in \mathbb{R}$.

Mode Rationality Test

Assumption 2

Besides Assumption 1, it holds that

(B1) There exist $\theta_0 \in \Theta$ and sequences $\phi^*_{t,T}(\theta_0)$ and $u_{t,T}(\theta_0)$, such that

$$\tilde{\phi}_{t,T}(\theta_0) := \phi^*_{t,T}(\theta_0) + u_{t,T}(\theta_0), \tag{27}$$

and

(a) $\left\{T^{-1/2}\phi^*_{t,T}(\theta_0), \mathcal{F}_{t+1}\right\}$ is a martingale difference sequence.

(b) $T^{-1}\sum_{t=1}^{T}\left(u_{t,T}(\theta_0)\lambda\right)^2 \xrightarrow{P} 0$ and $\sum_{t=1}^{T}\mathbb{E}\left[||T^{-1/2}u_{t,T}(\theta_0)||^{2+\delta}\right] \to 0$

(c) $T^{-1}\sum_{t=1}^{T} u_{t,T}(\theta_0)\tilde{\phi}_{t,T}(\theta_0) \xrightarrow{P} 0$ and $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[u_{t,T}(\theta_0)\tilde{\phi}_{t,T}(\theta_0)\right] \to 0$

(B2) $\mathbb{E}\left[\varepsilon_t^2\right] < \infty$

(B3) $\widehat{\boldsymbol{W}}_{T,\text{Mean}} \xrightarrow{P} \boldsymbol{W}_{\text{Mean}}$, $\widehat{\boldsymbol{W}}_{T,\text{Med}} \xrightarrow{P} \boldsymbol{W}_{\text{Med}}$, and $\widehat{\boldsymbol{W}}_{T,\text{Mode}} \xrightarrow{P} \boldsymbol{W}_{\text{Mode}}$ for some positive definite matrices $\boldsymbol{W}_{\text{Mean}}$, $\boldsymbol{W}_{\text{Med}}$ and $\boldsymbol{W}_{\text{Mode}}$.

Identifying the Measure of Centrality

Discussion of Assumption (B1)

- Assumption (B1) of an approximate MDS is weaker than the standard MDS assumption,

$$\exists\, \theta_0 \in \Theta \text{ s.t. } \left\{T^{-1/2}\tilde{\phi}_t(\theta_0), \mathcal{F}_{t+1}\right\} \text{ is a MDS.} \tag{28}$$

- The classically imposed (weaker) MDS assumption

$$\exists\, \theta_0 \in \Theta \text{ s.t. } \mathbb{E}\left[\tilde{\phi}_t(\theta_0)\right] = 0, \tag{29}$$

  is too weak for our case.

  - Given (29), in order to apply a CLT for stationary and ergodic (or strong mixing) assumptions without the MDS assumption, we need that moments of order $2+\delta$ (or $r>2$) are finite, which is not fulfilled for the mode case as these moments diverge arbitrarily slowly through the bandwidth parameter δ_T.

Identifying the Measure of Centrality

Discussion of Assumption (B1) II

Assumption (B1) can easily be shown to hold for the three vertices, where the forecast is the mean, median or mode:

- When X_t is a mean or median forecast (i.e. $\theta_0 = (1,0,0)$ or $\theta_0 = (0,1,0)$), set $u_{t,T}(\theta_0) = 0$ and $\left\{T^{-1/2}\tilde{\phi}_{t,T}(\theta_0), \mathcal{F}_{t+1}\right\}$ is obviously a MDS.
- When X_t is the true conditional mode of Y_{t+1}, (i.e. $\theta_0 = (0,0,1)$), set

$$u_{t,T}(\theta_0) = \mathbb{E}_t\left[\tilde{\phi}_{t,T}(\theta_0)\right] = (T\delta_T)^{-1/2}\mathbb{E}_t\left[K'\left(\frac{-\varepsilon_t}{\delta_T}\right)\right]\boldsymbol{h}_t^\top. \tag{30}$$

- When X_t is a convex combination of a mean and median forecast, i.e. $\theta_0 = (\xi, 1-\xi, 0)$ for some $\xi \in [0,1]$, we set $u_{t,T}(\theta_0) = 0$ and $\left\{T^{-1/2}\tilde{\phi}_{t,T}(\theta_0), \mathcal{F}_{t+1}\right\}$ is again a MDS.
- When X_t is a convex combination with non-zero weight on the mode Assumption (B1) is difficult to verify.

Identifying the Measure of Centrality

Practical Choice of the Instruments

- The instruments
  - must be $\mathcal{F}_t$-measurable (known to the forecaster at time t)
  - should span as much information as possible in $\mathcal{F}_t$
  - should be uncorrelated
- $\boldsymbol{h}_t = (1, X_t)$ is an obvious starting point for all applications
- Examples of further variables that might be interesting:
  - Income Application: covariates of the forecaster, past reported income
  - GDP Application: forecast revisions, macro-economic variables
  - FOREX Application: macro-economic variables

Simulation Setup: Bias

- Generate data from a **skewed normal** AR(1)-GARCH(1,1) model:

$$Y_{t+1} = 0.5Y_t + \sigma_{t+1}\,\varepsilon_{t+1}, \tag{31}$$

$$\sigma^2_{t+1} = 0.1 + 0.1Y_t^2 + 0.8\sigma_t^2 \tag{32}$$

$$\varepsilon_t \sim \mathcal{SN}(0,1,\eta) \tag{33}$$

- Optimal mode forecast:

$$\tilde{X}_t = 0.5Y_t + \sigma_{t+1}\,\mathrm{Mode}(\varepsilon_{t+1}) \tag{34}$$

- Misspecified forecasts:

$$\textbf{Bias:} \quad X_t = \tilde{X}_t + \kappa_1\sigma_X, \quad \text{where} \quad \sigma_X = \sqrt{\mathrm{Var}(\tilde{X}_t)} \tag{35}$$

$$\textbf{Noise:} \quad X_t = \tilde{X}_t + \mathcal{N}(0, \kappa_2\sigma_X^2). \tag{36}$$

- We vary the following:
  - misspecification: $\kappa_1 \in [-1,1]$ and $\kappa_2 \in [0,1]$
  - skewness: $\eta \in \{0, 0.1, 0.25, 0.5\}$
  - sample size: $T \in \{100, 500, 2000\}$
  - instruments: $\boldsymbol{h}_t \subseteq \{1, X_t, Y_{t-1}, X_{t-1} - Y_{t-1}\}$

Size

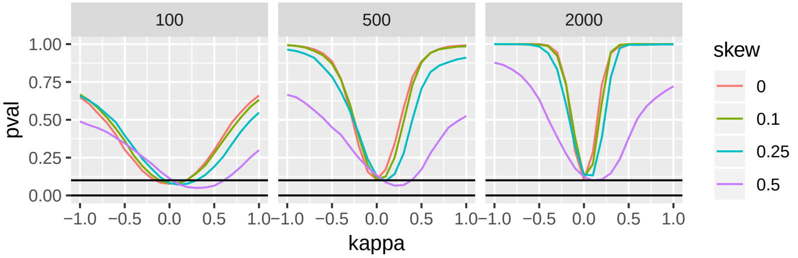
- nominal size = 0.1

	$\boldsymbol{h}_t = 1$				$\boldsymbol{h}_t = (1, X_t)$			
skewness	0	0.1	0.25	0.5	0	0.1	0.25	0.5
$T = 100$	0.085	0.088	0.094	0.126	0.094	0.086	0.096	0.116
$T = 500$	0.112	0.113	0.124	0.135	0.111	0.107	0.116	0.125
$T = 2000$	0.107	0.111	0.123	0.125	0.104	0.180	0.114	0.111

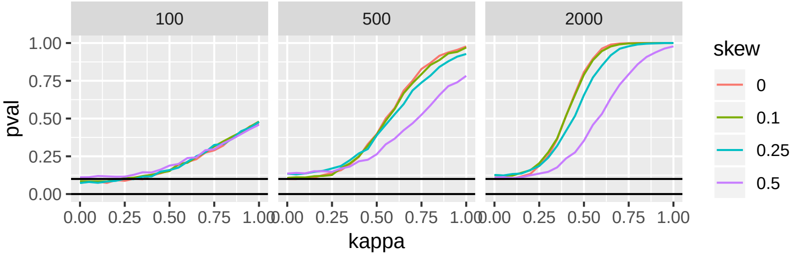
	$\boldsymbol{h}_t = (1, X_t, Y_{t-1})$				$\boldsymbol{h}_t = (1, X_t, Y_{t-2}, e_{t-1})$			
skewness	0	0.1	0.25	0.5	0	0.1	0.25	0.5
$T = 100$	0.082	0.082	0.084	0.102	0.074	0.068	0.074	0.088
$T = 500$	0.108	0.107	0.104	0.118	0.103	0.103	0.102	0.111
$T = 2000$	0.102	0.103	0.114	0.109	0.101	0.101	0.114	0.109

Power

$\boldsymbol{h}_t = (1, X_t)$



$\boldsymbol{h}_t = (1, X_t)$



Measure of Centrality: Simulation Setup

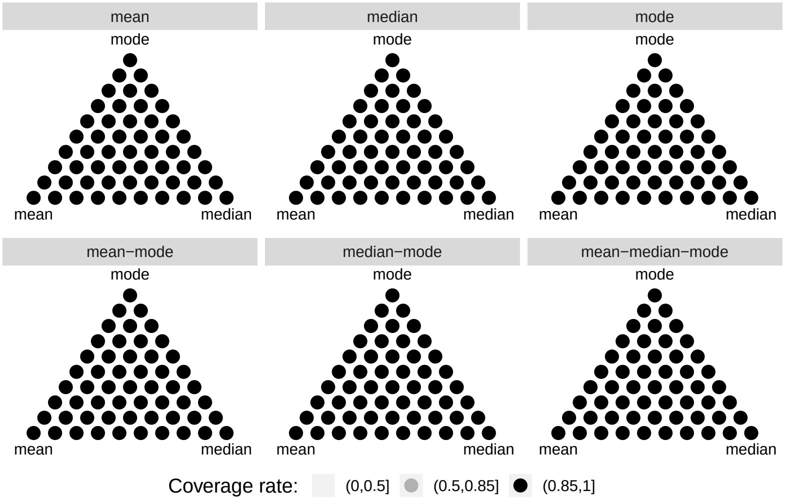
- Generate data from a **skewed normal** AR(1)-GARCH(1,1) model:

$$Y_{t+1} = 0.5Y_t + \sigma_{t+1}\,\varepsilon_{t+1}, \tag{37}$$
$$\sigma^2_{t+1} = 0.1 + 0.1Y_t^2 + 0.8\sigma_t^2 \tag{38}$$
$$\varepsilon_t \sim \mathcal{SN}(0, 1, \eta) \tag{39}$$

- We choose:
  - sample size: $T = 2000$
  - instruments: $\boldsymbol{h}_t = (1, X_t)$
  - skewness: $\eta \in \{0, 0.5\}$
- Optimal Forecasts:

$$X_t^{Mode} = 0.5Y_t + \sigma_{t+1}\,\mathrm{Mode}(\varepsilon_{t+1}) \tag{40}$$
$$X_t^{Mean} = 0.5Y_t + \sigma_{t+1}\,\mathrm{Mean}(\varepsilon_{t+1}) \tag{41}$$
$$X_t^{Median} = 0.5Y_t + \sigma_{t+1}\,\mathrm{Median}(\varepsilon_{t+1}) \tag{42}$$
$$X_t^{Mean-Med-Mode} = 1/3X_t^{Mean} + 1/3X_t^{Median} + 1/3X_t^{Mode} \tag{43}$$
$$X_t^{Mean-Mode} = 1/2X_t^{Mean} + 1/2X_t^{Mode} \tag{44}$$
$$X_t^{Median-Mode} = 1/2X_t^{Median} + 1/2X_t^{Mode} \tag{45}$$

Measure of Centrality Coverage Rates, skew $= 0$



Measure of Centrality Coverage Rates, skew $= 0.5$