

Testing Forecast Rationality for Measures of Central Tendency*

Timo Dimitriadis[†] Andrew J. Patton[‡] Patrick W. Schmidt[§]

This Version: May 15, 2023
First Version: February 15, 2019

Abstract

Rational respondents to economic surveys may report as a point forecast any measure of the central tendency of their (possibly latent) predictive distribution, for example the mean, median, mode, or any convex combination thereof. We propose tests of forecast rationality when the measure of central tendency used by the respondent is unknown. We overcome an identification problem that arises when the measures of central tendency are equal or in a local neighborhood of each other, as is the case for (exactly or nearly) symmetric distributions. As a building block, we also present novel tests for the rationality of mode forecasts. We apply our tests to income forecasts from the Federal Reserve Bank of New York’s Survey of Consumer Expectations. We find these forecasts are rationalizable as mode forecasts, but not as mean or median forecasts. We also find heterogeneity in the measure of centrality used by respondents when stratifying the sample by past income, age, job stability, and past forecast accuracy.

Keywords: forecast evaluation, partial identification, survey forecasts, mode forecasts

J.E.L. Codes: C53, D84, E27

*We thank Jonas Brehmer, Tobias Fissler, Rafael Frongillo, Tilmann Gneiting, Arnaud Maurel, Adam Rosen, Melanie Schienle, Jörg Stoye, as well as seminar participants at Boston University, Duke University, KIT Karlsruhe, HITS Heidelberg, Monash University, UC Riverside, Universität Hohenheim, Universität Heidelberg, University of Sydney, CSS Workshop in Zürich, HKMetrics Workshop in Mannheim, ISI World Statistics Congress in Kuala Lumpur, the 2019 Statistische Woche in Trier, the 2019 EC² Conference in Oxford, and the 2020 Econometric Society World Congress. Code to replicate the empirical work in this paper in the form of the R package `fcrat` is available at <https://github.com/Schmidtpk/fcrat>.

[†]Alfred Weber Institute of Economics, Heidelberg University and Heidelberg Institute for Theoretical Studies, Heidelberg, Germany, e-mail: timo.dimitriadis@awi.uni-heidelberg.de

[‡]Department of Economics, Duke University, Durham, USA, e-mail: andrew.patton@duke.edu

[§]University of Zurich, Zurich, Switzerland, e-mail: patrickwolfgang.schmidt@uzh.ch

1 Introduction

Economic surveys are a rich source of information about future economic conditions, yet most economic surveys are vague about the specific statistical quantity the respondent should report. For example, the New York Federal Reserve’s labor market survey asks respondents “What do you believe your annual earnings will be in four months?” A reasonable response to this question is the respondent reporting her mathematical expectation of future earnings, or her median, or her mode; all common measures of the central tendency of a distribution. When these measures coincide, as they do for symmetric unimodal distributions, this ambiguity does not affect the information content of the forecast. When these measures differ, the specific measure adopted by the respondent can influence its use in other applications, and testing rationality of forecasts becomes difficult. Subjective forecast distributions have been found to be asymmetric for many important economic variables as GDP growth ([Adrian et al., 2019](#); [Bekaert and Popov, 2019](#)), inflation rates ([Garcia and Manzanares, 2007](#)), firm earnings ([Foster, 1986](#); [Givoly and Hayn, 2000](#); [Gu and Wu, 2003](#)) and consumer expenses ([Howard et al., 2022](#)), motivating the need for reliable forecast evaluation methods for the different measures of central tendency.

While the assumption of a mean forecast¹ is common in the economic and statistical literature (see, e.g., [Coibion and Gorodnichenko, 2015](#); [Bordalo et al., 2020](#)), it may be mistaken in some applications. For example, [Knüppel and Schultefrankfeld \(2012\)](#) conclude that point forecasts of inflation published by central banks almost always correspond to the modes of the forecast densities, and [Reifschneider and Tulip \(2019\)](#) suggest that forecasts from the U.S. Board of Governors are best interpreted as mode forecasts. [Howard et al. \(2022\)](#) and [Zhao \(2022\)](#) find that some survey forecasts are more consistent with the respondents’ modes than means or medians. In an early experimental study, [Peterson and Miller \(1964\)](#) found that respondents could accurately predict the mode and median if incentivized correctly, but had difficulty reporting accurate estimates of the mean, and a more recent study by [Kröger and Pierrot \(2019b\)](#) reports that participants asked to summarize their predictive distribution responded most frequently with their mode. We propose

¹We use the phrase “mean forecasts,” or similar, as shorthand for the forecaster reporting the mean of her predictive distribution as her point forecast.

that other measures of central tendency deserve further consideration, especially given the growing evidence that point forecasts may reflect the mode.

Given the ambiguity around which specific measure of central tendency is used by survey respondents, we consider a *class* of such measures defined by the set of convex combinations of the mean, median, and mode.² Similar to [Elliott et al. \(2005\)](#), we propose a testing framework that nests the mean as a special case, but unlike that paper’s approach we allow for alternative forecasts *within* a class of measures of central tendency, rather than measures that represent other aspects of the predictive distribution (such as non-central quantiles or expectiles). Our approach faces an identification problem: for symmetric distributions, the combination weight vector is unidentified, for “mildly” asymmetric distributions, the weight vector is only weakly identified, and even for strongly asymmetric distributions the weight vector may only be partially identified. Economic variables may fall into any of these cases, and a valid testing approach must accommodate these measures of central tendency being equal, unequal, or in a local neighborhood of each other. We use the work of [Stock and Wright \(2000\)](#) to obtain asymptotically valid confidence sets for the combination weights and to test forecast rationality.

Depending on the skewness of the respondent’s predictive distribution, our approach may allow the researcher to distinguish mean, median, and mode forecasts based on point forecasts and realizations alone. An alternative approach was proposed in [Engelberg et al. \(2009\)](#), who combine point and density forecasts to determine whether the point forecasts are the mean, median, or mode of the predictive density. In their study of respondents to the Federal Reserve Bank of Philadelphia’s Survey of Professional Forecasters, [Engelberg et al. \(2009\)](#) find that most point forecasts are consistent with the bounds derived for all three functionals. [Zhao \(2022\)](#) applies the same methodology to inflation forecasts from the Federal Reserve Bank of New York’s Survey of Consumer Expectations and finds the mode of the density forecast to be most consistent with the point forecasts, but the mean and median are also valid for a majority of forecasts. Our testing approach accommodates the fact that these functionals may be hard, or impossible, to separately identify in data.

²These are the three measures described in introductory statistics textbooks ([McClave et al., 2017](#)), in previous studies, ([Engelberg et al., 2009](#)), in central bank publications (the [Bank of England’s](#) quarterly inflation reports), and in psychological work, (p. 86, [Kahneman et al., 1982](#)). Nevertheless, our approach can easily be extended to consider a broader set of measures of centrality.

Before implementing the above test for rationality for a general forecast of central tendency, we must first overcome a lack of rationality tests for mode forecasts. Rationality tests for mean forecasts go back to at least [Mincer and Zarnowitz \(1969\)](#), see [Elliott and Timmermann \(2016\)](#) for a recent survey, while rationality tests for quantile forecasts (nesting median forecasts as a special case) are considered in [Christoffersen \(1998\)](#) and [Gaglianone et al. \(2011\)](#). A critical impediment to similar tests for mode forecasts is that the mode is not an “elicitable functional” ([Heinrich, 2014](#)), meaning that it cannot be obtained as the solution to an expected loss minimization problem.^{3,4} We obtain a test for mode forecast rationality by first proposing novel results on the *asymptotic elicibility* of the mode. We define a functional to be asymptotically elicitable if there exists a sequence of elicitable functionals that converges to the target functional. We consider the (elicitable) “generalized modal interval,” defined in detail in Section 2.2, and show that it converges to the mode for a general class of probability distributions. We combine these results with recent work on mode regression ([Kemp and Silva, 2012](#); [Kemp et al., 2020](#)) and nonparametric kernel methods to obtain mode forecast rationality tests analogous to well-known tests for mean and median forecasts. In addition to size control, we show that the proposed test has non-trivial asymptotic power against both fixed and local alternative hypotheses.

We evaluate the finite sample performance of the new mode rationality test and of the proposed method for obtaining confidence sets for measures of centrality through an extensive simulation study. We use cross-sectional and time-series data generating processes with a range of asymmetry levels. We find that our proposed mode forecast rationality test has satisfactory size properties, even in small samples, and exhibits strong power across different misspecification designs. Our simulation design allows us to consider the four identification cases that can arise in practice: strongly identified (skewed data), where the mean, median and mode differ; unidentified (symmetric unimodal data), where all centrality measures coincide; weakly identified (mildly skewed data), where the centrality measures differ but are close to each other; and partially identified (skewed

³[Gneiting \(2011\)](#) provides an overview of elicibility and identifiability of statistical functionals and shows that several important functionals such as variance, Expected Shortfall, and mode are not elicitable. [Fissler and Ziegel \(2016\)](#) introduce the concept of higher-order elicibility, which facilitates the elicitation of vector-valued (stacked) functionals such as the variance and Expected Shortfall, though not the mode ([Dearborn and Frongillo, 2020](#)).

⁴For categorical data, rationality tests for mode forecasts were established in [Das et al. \(1999\)](#) and extended in [Madeira \(2018\)](#). Our focus is on continuously distributed target variables, and so we cannot use their results.

location-scale data), where one centrality measure is a convex combination of the other two. We find that in the symmetric case, the resulting confidence sets contain, correctly, the entire set of convex combinations of mean, median and mode. In the asymmetric cases, our rationality test is able to identify the combination weights corresponding to the issued centrality forecast.

We apply the new tests to the income survey responses from the Survey of Consumer Expectations conducted by the Federal Reserve Bank of New York. In the full sample of respondents, we find the intriguing result that we can reject rationality with respect to the mean or median, however we cannot reject rationality when interpreting these as mode forecasts, suggesting that survey participants report the anticipated *most likely* outcome rather than the average or median. When allowing for cross-respondent heterogeneity, we find that forecasts from low-income younger survey respondents cannot be rationalized using any measure of central tendency, while forecasts from high-income respondents, regardless of their age, are rationalizable for many different measures of centrality. We also find evidence of learning between survey rounds (Kim and Binder, 2023) for high-income respondents, but much less so for low-income respondents. We compare our rationality test results with those obtained using the approach of Elliott et al. (2005) (EKT), which allows for rational optimism or pessimism. We find no cases where allowing for optimism or pessimism “overturns” a rejection of rationality based on a measure of centrality. As a stark example, we find that forecasts from younger low-income respondents cannot be rationalized as any centrality measure, nor as a measure in the EKT framework.

Our paper is related to the large literature on forecasting under asymmetric loss, see Granger (1969), Christoffersen and Diebold (1997), Elliott et al. (2005), Patton and Timmermann (2007) and Elliott et al. (2008) amongst others. The work in these papers is motivated by the fact that forecasters may wish to use a loss function other than the omnipresent squared-error loss function. The use of asymmetric loss functions generally leads to point forecasts that differ from the mean (though this is not always true, see Gneiting, 2011 and Patton, 2020), and generally these point forecasts are not interpretable as measures of central tendency. For example, Christoffersen and Diebold (1997) show that the linex loss function implies an optimal point forecast that is a weighted sum of the mean and variance, while Elliott et al. (2008) find that their sample of macroeconomic

forecasters report an expectile with asymmetry parameter around 0.4. Instead of moving from the mean to a point forecast that is not a measure of location, our novel approach considers moving only *within* a general class of central tendency measures.

The remainder of the paper is structured as follows. In Section 2 we propose new forecast rationality tests for the mode based on the concepts of asymptotic elicibility and identifiability. Section 3 presents forecast rationality tests for general measures of central tendency, allowing for weak and partial identification. Section 4 presents simulation results on the finite-sample properties of the proposed tests, and Section 5 presents the empirical application using income survey forecasts. Proofs are presented in Appendix A. A supplemental appendix contains additional technical details, simulation results, and empirical analyses of the Federal Reserve Board’s “Greenbook” forecasts of US GDP growth and random walk forecasts of exchange rates.

2 Eliciting and Evaluating Mode Forecasts

2.1 General Forecast Rationality Tests

Let $Z_t = (Y_t, X_t, \tilde{\mathbf{h}}_t)$ be a stochastic process defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Y_{t+1} denotes the (scalar) variable of interest, $\tilde{\mathbf{h}}_t$ denotes a vector of variables known to the forecaster at the time she issues her point forecast for Y_{t+1} , which is denoted X_t .⁵ We define the information set $\mathcal{F}_t = \sigma\{Y_s, X_s, \tilde{\mathbf{h}}_s; s \leq t\}$ as the σ -field containing all information known to the forecaster at time t . We denote the distribution of Y_{t+1} given \mathcal{F}_t by F_t , and with corresponding density f_t . Neither the forecaster’s information set, \mathcal{F}_t nor her predictive distribution, F_t , are assumed known to the econometrician. In conducting the forecast rationality test, we assume that the econometrician uses an \mathcal{F}_t -measurable $(k \times 1)$ vector \mathbf{h}_t , which can be thought of as a subset of $\tilde{\mathbf{h}}_t$.⁶ Note that since F_t is not observed, we do not know whether this distribution is strongly skewed, mildly skewed, or symmetric, and thus our inference method must be valid for all of these possibilities. Conditional

⁵Given our focus on survey forecasts, where the underlying model used by the respondents, if any, is unknown, we take the forecasts as given, putting this paper in the general framework of [Giacomini and White \(2006\)](#), as opposed to that of [West \(1996\)](#).

⁶The interpretability of the outcome of a forecast rationality test, including ours, critically depends on the “test vector” or “instrument vector,” \mathbf{h}_t , being observable to the forecaster. In our empirical analysis we only use vectors that are guaranteed to satisfy this requirement.

expectations are denoted $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$. We use \mathcal{P} to denote a class of distributions.

We start by considering rationality tests (also known as calibration tests; [Nolde and Ziegel, 2017](#)) for the mean, i.e. we assume that the forecasts X_t are one-step ahead *mean* forecasts for Y_{t+1} . We are interested in testing if these forecasts are rational, which would imply the null hypothesis:

$$\mathbb{H}_0 : X_t = \mathbb{E}[Y_{t+1}|\mathcal{F}_t] \forall t \text{ a.s.} \quad (2.1)$$

We test this hypothesis using the “identification function” for the mean, which is simply the difference between the forecast and the realized value, i.e., the forecast error:⁷

$$V_{\text{Mean}}(X_t, Y_{t+1}) = X_t - Y_{t+1} =: \varepsilon_t. \quad (2.2)$$

Specifically, the null hypothesis in (2.1) implies that the identification function $V_{\text{Mean}}(X_t, Y_{t+1})$ is uncorrelated with any instrument vector $\mathbf{h}_t \in \mathcal{F}_t$, which provides a testable moment condition. Under the above null hypothesis and subject to standard regularity conditions, it is straight forward to show that $T^{-1/2} \sum_{t=1}^T V_{\text{Mean}}(X_t, Y_{t+1}) \mathbf{h}_t \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mean}})$ as $T \rightarrow \infty$, and that

$$J_T = \frac{1}{T} \left(\sum_{t=1}^T V_{\text{Mean}}(X_t, Y_{t+1}) \mathbf{h}_t^\top \right) \widehat{\Omega}_{T, \text{Mean}}^{-1} \left(\sum_{t=1}^T V_{\text{Mean}}(X_t, Y_{t+1}) \mathbf{h}_t \right) \xrightarrow{d} \chi_k^2 \quad (2.3)$$

as $T \rightarrow \infty$, where $\widehat{\Omega}_{T, \text{Mean}}$ is a consistent estimator of Ω_{Mean} . This result facilitates testing whether given forecasts X_t are rational mean forecasts for the realizations Y_{t+1} by using the test statistic J_T in equation (2.3) to test for uncorrelatedness of the identification function $V_{\text{Mean}}(X_t, Y_{t+1})$ and the instrument vector \mathbf{h}_t . As in most other tests in the literature, this is of course only a test of a necessary condition for forecast rationality, and the conclusion may be sensitive to the choice of instruments, \mathbf{h}_t .⁸

⁷The identification function for a point forecast can be obtained as the first derivative of any loss function that elicits that forecast. The quadratic loss function elicits the mean, and so its identification function is simply the forecast error, up to scale and sign. In econometrics the forecast error is usually defined as $Y_{t+1} - X_t$, that is, as the negative of the identification function in equation (2.2). Given the important role that forecast identification functions play in this paper, we adopt the definition for ε_t given in equation (2.2), and we refer to ε_t as the forecast error.

⁸Like most of the forecast evaluation literature, we assume that the vector of instruments is of fixed and finite length. A [Bierens \(1982\)](#)-type test, where the length of the vector diverges with the sample size, is considered for forecast evaluation in, for example, [Corradi and Swanson \(2002\)](#).

The test statistic in equation (2.3) is only informative about rationality if the forecasts are interpreted as being for the *mean* of Y_{t+1} . The decision-theoretic framework of identification functions and consistent loss functions is fundamental for generalizations to other measures of central tendency, such as the median and the mode. For a general real-valued functional $\Gamma : \mathcal{P} \rightarrow \mathbb{R}$, a strict identification function $V_\Gamma(Y, x)$ is defined by being zero in expectation if and only if x equals the functional $\Gamma(F)$. Strict identification functions are generally obtained as the derivatives of strictly consistent loss functions, which are defined as having the functional $\Gamma(F)$ as their unique minimizer (in expectation). A functional is called *identifiable* if a strict identification function exists, and is called *elicitable* if a strictly consistent loss function exists. See [Gneiting \(2011\)](#) for a general introduction to elicibility and identifiability.

The forecast error $X_t - Y_{t+1}$ is a strict identification function for the mean, and a strict identification function for the median is given by the step function:

$$V_{\text{Med}}(X_t, Y_{t+1}) = \mathbb{1}_{\{Y_{t+1} < X_t\}} - \mathbb{1}_{\{Y_{t+1} > X_t\}}. \quad (2.4)$$

We obtain a test of median forecast rationality by replacing V_{Mean} and $\widehat{\Omega}_{T, \text{Mean}}$ by V_{Med} and $\widehat{\Omega}_{T, \text{Med}}$ in equation (2.3).

2.2 The Mode Functional

In contrast to the mean and the median, rationality tests for mode forecasts are more challenging to consider. The underlying reason is that there do not exist identification functions for the mode for random variables with continuous Lebesgue densities ([Heinrich, 2014](#); [Dearborn and Frongillo, 2020](#)). In this section we simplify notation and refer to the target variable and forecast as Y and x . We define the mode for random variables with continuous Lebesgue densities as the global maxima of the density function.⁹ We make the following distinction in the notion of *unimodality*.

⁹More generally, the mode is often defined as the limit, as $\delta \rightarrow 0$, of the modal midpoint functional MMP_δ , given in equation (2.5) below ([Gneiting, 2011](#); [Dearborn and Frongillo, 2020](#)). This definition coincides with the global maxima of the density function for distributions with continuous Lebesgue density; and it coincides with the points of maximal probability for discrete distributions ([Heinrich, 2014](#)). Note that our definition of the mode as the global maxima of a density function is only valid for distributions with *continuous* Lebesgue densities as otherwise the density function can be modified on singletons (null sets in the distribution) without altering the underlying probability measure.

Definition 2.1. *An absolutely continuous distribution is defined as (a) weakly unimodal if it has a unique and well-defined mode, and (b) strongly unimodal if it is weakly unimodal and does not have further local modes.*

Heinrich-Mertsching and Fissler (2022) show that even for the class of strongly unimodal distributions, neither strictly consistent loss functions nor strict identification functions exist for the mode. Gneiting (2011) notes that it is sometimes stated informally that the mode is an optimal point forecast under the loss function $L_\delta(x, Y) = \mathbf{1}_{\{|x-Y| \leq \delta\}}$ for some fixed $\delta > 0$. In fact, this loss function elicits the midpoint of the modal interval (also known as the modal midpoint, or MMP) of length 2δ . The MMP of $Y \sim P$ is defined as the midpoint of the interval of length 2δ that contains the highest probability:

$$\text{MMP}_\delta(P) = \sup_{x \in \mathbb{R}} \mathbb{P}(Y \in [x - \delta, x + \delta]). \quad (2.5)$$

More formally, it holds that for any $\delta > 0$ small enough, the modal midpoint is well-defined for all distributions with unique and well-defined mode, and it holds that $\lim_{\delta \downarrow 0} \text{MMP}_\delta(P) = \text{Mode}(P)$ (Gneiting, 2011). In a similar manner, Eddy (1980), Kemp and Silva (2012) and Kemp et al. (2020) propose estimation of the mode by estimating the modal interval with an asymptotically shrinking length. We formalize these ideas in the decision-theoretical framework in the following definition.

Definition 2.2. *The functional $\Gamma : \mathcal{P} \rightarrow \mathbb{R}$ is asymptotically elicitable (identifiable) relative to \mathcal{P} if there exists a sequence of elicitable (identifiable) functionals $\Gamma_\delta : \mathcal{P} \rightarrow \mathbb{R}$, such that $\Gamma_\delta(P) \rightarrow \Gamma(P)$ as $\delta \rightarrow 0$ for all $P \in \mathcal{P}$.*

As the modal midpoint converges to the mode, this establishes *asymptotic elicibility* for the mode functional for the class of weakly unimodal probability distributions with continuous Lebesgue densities. Unfortunately, this does not directly allow for asymptotic *identifiability* of the mode as any pseudo-derivative of the loss function L_δ equals zero. We establish asymptotic identifiability of the mode through the *generalized modal midpoint*.

Definition 2.3. *Given a kernel function $K(\cdot)$ and bandwidth parameter δ , the generalized modal*

midpoint, $\Gamma_\delta^K(P)$, of $Y \sim P$ is defined as

$$\Gamma_\delta^K(P) = \arg \min_{x \in \mathbb{R}} \mathbb{E} [L_\delta^K(x, Y)], \quad \text{where} \quad L_\delta^K(x, Y) = -\frac{1}{\delta} K\left(\frac{x - Y}{\delta}\right). \quad (2.6)$$

The familiar modal midpoint is nested in this definition by using a rectangular kernel for $K(u) = \mathbb{1}_{\{|u| \leq 1\}}$, and this definition allows for smooth generalizations. As this definition involves the argmin of a function, we first establish that this is well-defined and that it converges to the mode functional. The following theorem also considers identifiability of the generalized modal midpoint.

Theorem 2.4. *Let K be a strictly positive kernel function on the real line that is log-concave, i.e. $\log(K(u))$ is a concave function, and additionally let $\int K(u)du = 1$ and $\int |u|K(u)du < \infty$. Let \mathcal{P} be the class of absolutely continuous and weakly unimodal distributions with bounded and Lipschitz-continuous density and let $\tilde{\mathcal{P}} \subset \mathcal{P}$ be the subclass of strongly unimodal distributions.*

- (a) *The functional Γ_δ^K induced by the loss function (2.6) is well-defined for all $\delta > 0$ and $P \in \mathcal{P}$.*
- (b) *It holds that $\Gamma_\delta^K(P) \rightarrow \text{Mode}(P)$ as $\delta \rightarrow 0$ for all $P \in \mathcal{P}$.*
- (c) *If K is differentiable, for all (fixed) $\delta > 0$ and $P \in \tilde{\mathcal{P}}$, it holds that the function*

$$V_\delta^K(x, Y) = \frac{\partial}{\partial x} L_\delta^K(x, Y) = -\frac{1}{\delta^2} K'\left(\frac{x - Y}{\delta}\right) \quad (2.7)$$

is a strict identification function for Γ_δ^K . In particular, the generalized modal midpoint is identifiable and the mode is asymptotically identifiable with respect to $\tilde{\mathcal{P}}$.

This theorem shows that for the classes of weakly (strongly) unimodal distributions, the generalized modal midpoint is elicitable (and identifiable), and consequently, the mode is asymptotically elicitable (and identifiable). While $V_\delta^K(x, Y)$ being an identification function for the generalized modal midpoint is obvious from Definition 2.3, Theorem 2.4(c) establishes its strictness.

For a fixed $\delta > 0$, strict identifiability is doomed to fail when both the underlying distribution and the kernel function have bounded support as the expected identification function equals zero for values far outside both supports. Theorem 2.4(c) shows that employing log-concave kernels with

infinite support circumvents this problem. While kernels with bounded support generally exhibit a superior performance in nonparametric statistics, this proposition motivates the use of kernels with infinite support like the Gaussian density function. Furthermore, the Gaussian density function, among many others, satisfies the required log-concavity of the kernel function.¹⁰

2.3 Forecast Rationality Tests for the Mode

Having established the asymptotic identifiability of the mode in the previous section, we now consider rationality testing of mode forecasts, i.e. testing the following null hypothesis,

$$\mathbb{H}_0 : X_t = \text{Mode}(Y_{t+1}|\mathcal{F}_t) \quad \forall t \text{ a.s.} \quad (2.8)$$

While classical, \sqrt{T} -consistent rationality tests based on strict identification functions are unavailable for the mode due to its non-identifiability, we next propose a rationality test for mode forecasts based on an asymptotically shrinking bandwidth parameter δ_T . Consider the (asymptotically valid) identification function $V_{\delta_T}^K$ with bandwidth δ_T , and multiplied by the instruments \mathbf{h}_t ,

$$\psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) := V_{\delta_T}^K(X_t, Y_{t+1})\mathbf{h}_t = -\frac{1}{\delta_T^2}K'\left(\frac{X_t - Y_{t+1}}{\delta_T}\right)\mathbf{h}_t. \quad (2.9)$$

We make the following assumptions. For the remainder of the paper all limits are taken as $T \rightarrow \infty$, unless stated otherwise.

Assumption 2.5.

(A1) The sequence $(\varepsilon_t, \mathbf{h}_t)$ for $t \in \mathbb{N}$ is α -mixing of size $-r/(r-1)$ for some $r > 1$.

(A2) It holds that $\mathbb{E}[|\mathbf{h}_t|^{2r+\delta}] < \infty$ for all $t \in \mathbb{N}$ for some $\delta > 0$.

(A3) The matrix $\mathbb{E}[\mathbf{h}_t\mathbf{h}_t^\top]$ has full rank for all $t \in \mathbb{N}$.

(A4) The limit, Ω_{Mode} , of $\Omega_{T,\text{Mode}} := \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{h}_t\mathbf{h}_t^\top f_t(0)] \int K'(u)^2 du$ is positive definite.

¹⁰Theorem 2.4(c) also holds if log-concavity of the underlying density, instead of the kernel function, holds. This illustrates that kernels with bounded support can be employed at the cost of restricting the class of distributions.

(A5) For all $t \in \mathbb{N}$, the conditional distribution of $\varepsilon_t = X_t - Y_{t+1}$ given \mathcal{F}_t is absolutely continuous with density $f_t(\cdot)$ which is three times continuously differentiable with bounded derivatives.

(A6) $K : \mathbb{R} \rightarrow \mathbb{R}$, $u \mapsto K(u)$ is a non-negative and continuously differentiable kernel function such that: (i) $\int K(u)du = 1$, (ii) $\int uK(u)du = 0$, (iii) $\sup K(u) \leq c < \infty$, (iv) $\sup K'(u) \leq c < \infty$, (v) $\int u^2K(u)du < \infty$, (vi) $\int |K'(u)|du < \infty$, (vii) $\int uK'(u)du < \infty$.

(A7) δ_T is a strictly positive and deterministic sequence such that (i) $T\delta_T \rightarrow \infty$, and (ii) $T\delta_T^7 \rightarrow 0$.

The above assumptions are a combination of standard assumptions from rationality testing and nonparametric statistics. Conditions (A1) and (A2) facilitate the use of a law of large numbers and of a central limit theorem for martingale difference arrays (MDA) (that generalize MD sequences to triangular arrays; see Davidson (1994)) and allows for both possibly non-stationary time-series and cross-sectional applications. Notice that the rate in the mixing condition (A1) is relatively weak as we only need it for a law of large numbers while we apply a central limit theorem for MDA in the proof of Theorem 2.6 below. In cross-sectional applications with independent observations, this assumption can be replaced (and weakened) by the classical Lindeberg condition (see e.g. White (2001), Section 5.2). Notice that as the kernel function K' is bounded, we do not require existence of any moments of Y_t or X_t , which makes this more flexible than rationality testing for mean forecasts. The full rank condition (A3) prevents the instruments from being perfectly colinear which in turn prevents the asymptotic covariance matrix from being singular. Condition (A4) guarantees that the asymptotic covariance matrix is well behaved for non-stationary data. Assumption (A5) assumes a relatively smooth behavior of the conditional density function which is required to apply a Taylor expansion common to the nonparametric literature. Conditions (A6) and (A7) are standard kernel and bandwidth conditions from the nonparametric literature. We discuss specific kernel and bandwidth choices in Supplemental Appendices S.2 and S.3 respectively.

Theorem 2.6. Under Assumption 2.5 and $\mathbb{H}_0 : X_t = \text{Mode}(Y_{t+1}|\mathcal{F}_t) \forall t$ a.s., it holds that

$$\delta_T^{3/2} T^{-1/2} \sum_{t=1}^T \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mode}}), \quad (2.10)$$

where Ω_{Mode} is the limit of $\Omega_{T,\text{Mode}} = \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\mathbf{h}_t \mathbf{h}_t^\top f_t(0)] \int K'(u)^2 du$ as $T \rightarrow \infty$.

We obtain a test for the rationality of mode forecasts by drawing on the literature on nonparametric estimation. Unsurprisingly, therefore, the rate of convergence is slower than \sqrt{T} ; Assumption (A7) requires $\delta_T \propto T^{-\kappa}$ with $\kappa \in (1/7, 1)$, which implies that the fastest convergence rate approaches $T^{2/7}$ and is obtained by setting $\delta_T \approx T^{-1/7}$.¹¹

Theorem 2.4 guarantees the *strictness* of the identification function $V_\delta(x, Y)$ only if K has infinite support and is strictly increasing (decreasing) left (right) of its mode. These conditions are satisfied by the familiar Gaussian kernel, which we adopt for our analysis.¹²

Following Kemp and Silva (2012) and Kemp et al. (2020), we estimate the covariance matrix by its sample counterpart,

$$\widehat{\Omega}_{T,\text{Mode}} = \frac{1}{T} \sum_{t=1}^T \delta_T^{-1} K' \left(\frac{X_t - Y_{t+1}}{\delta_T} \right)^2 \mathbf{h}_t \mathbf{h}_t^\top. \quad (2.11)$$

The following theorem shows consistency of the asymptotic covariance estimator, its proof is presented in Supplemental Appendix S.1.

Theorem 2.7. *Given Assumption 2.5, it holds that $\widehat{\Omega}_{T,\text{Mode}} - \Omega_{T,\text{Mode}} \xrightarrow{P} 0$.*

We can now define the Wald test statistic:

$$J_T = \left(\delta_T^{3/2} T^{-1/2} \sum_{t=1}^T \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) \right)^\top \widehat{\Omega}_{T,\text{Mode}}^{-1} \left(\delta_T^{3/2} T^{-1/2} \sum_{t=1}^T \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) \right). \quad (2.12)$$

The following statement follows directly from Theorem 2.6 and Theorem 2.7.

Corollary 2.8. *Under Assumption 2.5 and the null hypothesis $\mathbb{H}_0 : X_t = \text{Mode}(Y_{t+1} | \mathcal{F}_t) \forall t$ a.s., it holds that $J_T \xrightarrow{d} \chi_k^2$.*

¹¹Under additional assumptions, the speed of convergence of a nonparametric estimator may be increased via the use of higher-order kernel functions, see e.g. Li and Racine (2006). However, as the generalized modal midpoint introduced in Definition 2.3 requires a log-concave kernel to be well-defined and unique (see Theorem 2.4), and as this assumption is automatically violated for higher-order kernels, we do not consider them here.

¹²We also considered the relatively efficient quartic (or biweight) kernel but did not observe a change in power relative to the Gaussian kernel. See Supplemental Appendix S.2 for further discussion of the kernel choice and simulation results.

This corollary justifies an asymptotic test at level $\alpha \in (0, 1)$ which rejects \mathbb{H}_0 when $J_T > Q_k(1 - \alpha)$, where $Q_k(1 - \alpha)$ denotes the $(1 - \alpha)$ quantile of the χ_k^2 distribution.¹³ Note that the bandwidth parameter, δ_T , is introduced only to conduct the test of mode forecast rationality; the forecast itself, X_t , is, under the null, the true conditional mode of the target variable, not a (smoothed) modal midpoint.

We now turn to the behavior of our test statistic J_T under a local alternative hypothesis,

$$\mathbb{H}_{A,loc} : \frac{1}{T} \sum_{t=1}^T f'_t(0) \mathbf{h}_t = c \cdot a_T + o_P(1), \quad (2.13)$$

for some constant $c \in \mathbb{R}^k$ and some possibly stochastic sequence a_T to be specified in the following Theorem 2.9, which characterizes the behavior of our mode rationality test under the local alternative.

Theorem 2.9. *If $T\delta_T^3 \rightarrow \infty$, then under Assumption 2.5 and the alternative $\mathbb{H}_{A,loc}$ in (2.13):*

- (a) *If $a_T T^{1/2} \delta_T^{3/2} \xrightarrow{P} 1$, then $J_T \xrightarrow{d} \chi_k^2(c^\top \Omega_{Mode}^{-1} c)$, where $\chi_k^2(\tilde{c})$ denotes a χ_k^2 distribution with non-centrality parameter $\tilde{c} \in \mathbb{R}$.*
- (b) *If $a_T T^{1/2} \delta_T^{3/2} \xrightarrow{P} 0$, then $J_T \xrightarrow{d} \chi_k^2$.*
- (c) *If $(a_T T^{1/2} \delta_T^{3/2})^{-1} \xrightarrow{P} 0$, then $\mathbb{P}(J_T \geq \bar{c}) \rightarrow 1$ for any $\bar{c} > 0$, i.e., we have uniform power.*

This theorem shows that the power of our mode rationality test is essentially driven by the condition that $\frac{1}{T} \sum_{t=1}^T f'_t(0) \mathbf{h}_t$ is non-zero, which implies that X_t cannot be the mode of F_t . It essentially means that the instruments \mathbf{h}_t are correlated with the conditional density slope at zero, $f'_t(0)$. This is analogous to the case for standard mean and median rationality tests which have power against the alternative that $\mathbb{E}[V(X_t, Y_{t+1}) \mathbf{h}_t] \neq 0$, where $V(X_t, Y_{t+1})$ is the identification function for the mean or median.¹⁴ When the instruments include a constant, a sufficient condition

¹³Our focus on one-step ahead forecasts allows for the application of a central limit theorem for MDAs, which only requires the existence of second moments. Multi-step ahead forecasts, on the other hand, are usually handled via CLTs for processes with more memory (e.g., mixing processes) at a cost of imposing stronger moment conditions. We leave this extension for future research.

¹⁴See e.g. Theorem 2 of [Giacomini and White \(2006\)](#), where their Comment 6 and [Nolde and Ziegel \(2017\)](#) point out that the theory can be directly adapted to rationality testing.

for a global alternative hypothesis, where a_T is constant, is $f'_t(0) \geq c > 0$ ($f'_t(0) \leq -c < 0$) for all $t \in \mathbb{N}$, i.e. when all forecasts are issued to the left (right) of the mode. This can be interpreted as uniform power against the class of strongly unimodal distributions. Our test will have low power to reject forecasts that are far in the tail, where the density is (almost) flat, however our test is intended for forecasts of a measure of central tendency, and such forecasts will lie broadly in the central region of the distribution where this concern does not arise.

Part (a) of Theorem 2.9 shows that if $\frac{1}{T} \sum_{t=1}^T f'_t(0) \mathbf{h}_t$ converges at rate $T^{-1/2} \delta_T^{-3/2}$, then the asymptotic distribution of J_T stabilizes as a non-central χ^2 -distribution, implying that for a fixed alternative, our test statistic diverges at rate $T^{1/2} \delta_T^{3/2}$, which is approximately $T^{2/7}$ in practice. In contrast, it is easy to show that rationality tests for identifiable functionals as the mean and median have local power against alternatives that converge with rate $T^{-1/2}$, which is of course faster than the rate $T^{-1/2} \delta_T^{-3/2}$ of the mode test. We demonstrate in the simulations and the applications that our mode test nevertheless has satisfactory power in practice in typically encountered sample sizes.

Part (b) of Theorem 2.9 further shows that if $\frac{1}{T} \sum_{t=1}^T f'_t(0) \mathbf{h}_t$ converges to zero *faster* than $T^{-1/2} \delta_T^{-3/2}$, then our test behaves as under the null and has no power. Finally, part (c) implies that our test has unit power asymptotically when $\frac{1}{T} \sum_{t=1}^T f'_t(0) \mathbf{h}_t$ converges to zero *slower* than $T^{-1/2} \delta_T^{-3/2}$, which nests the classical case of a fixed, global alternative.

Beyond the rationality tests proposed here, the concept of asymptotic elicibility is of interest in its own right. Asymptotic elicibility may facilitate forecast comparison and elicitation of novel measures of uncertainty. See for example [Eyting and Schmidt \(2021\)](#) for an elicitation procedure for the maximum, a functional that generally is not elicitable ([Bellini and Bignozzi, 2015](#)).

3 Evaluating Forecasts of Measures of Central Tendency

We define a class of measures of central tendency nesting the mean, median, and mode, and we propose tests of whether a forecast is rational with respect to *any* element of the class. Formally, we consider convex combinations of loss functions pertaining to the mean, median, and mode: $L_{\text{Mean}}(x, y) = (x - y)^2$, $L_{\text{Med}}(x, y) = |x - y|$, $L_{\text{Mode}, \delta}(x, y) = -\delta^{1/2} K((x - y)/\delta)$, for some kernel

K .¹⁵ Each vector in the unit simplex, $\Theta := \{\theta \in \mathbb{R}^3 : \|\theta\|_1 = 1, \theta \geq 0\}$, generates an optimal forecast

$$X_t^*(\theta) = \arg \min_{X \in \mathcal{F}_t} \mathbb{E}_t \left[\theta^\top [L_{\text{Mean}}(X, Y_{t+1}), L_{\text{Med}}(X, Y_{t+1}), L_{\text{Mode}, \delta}(X, Y_{t+1})]^\top \right], \quad (3.1)$$

where we minimize over all \mathcal{F}_t -measurable X . At the vertices of Θ , this nests the mean, median, and generalized modal midpoint, the latter tending to the mode as $\delta \rightarrow 0$; see Theorem 2.4.

Remark 3.1. *Intuitively one might be inclined to consider a convex combination of the functional values rather than of the associated loss functions, however such functionals are generally neither elicitable nor identifiable (see Proposition S.4.1 in the Supplemental Appendix), rendering infeasible an extension of the rationality tests introduced in Section 2.1 to this case. However it can be shown (Proposition S.4.2) that any forecast following equation (3.1) lies between the mean, median and generalized modal midpoint, though with combination weights for loss functions that generally differ from the combination weights for the functional values.*

Assuming that a forecaster generates her forecasts $X_t = X_t^*(\theta_0)$ according to (3.1), we aim to determine the set of values for θ_0 for which the given forecasts are optimal. Tests of forecast optimality are commonly conducted via an unconditional moment condition obtained by interacting an \mathcal{F}_t -measurable ($k \times 1$) vector of instruments \mathbf{h}_t with the first order condition of the optimal forecast (Elliott et al., 2005; Nolde and Ziegel, 2017). The latter also arise in the rationality tests introduced in (2.3) and (2.9). We continue in this tradition and test rationality across our general class of central tendency measures by testing the following hypothesis:

$$\mathbb{H}_0 : \quad \exists \theta_0 \in \Theta \text{ s.t. } \lim_{T \rightarrow \infty} T^{-1/2} \mathbb{E} [\phi_{t,T}(\theta_0)] = 0 \quad \forall t \leq T, \quad (3.2)$$

$$\text{where } \phi_{t,T}(\theta) = \mathbf{h}_t \theta^\top \begin{pmatrix} w_{\text{Mean}} V_{\text{Mean}}(X_t, Y_{t+1}) \\ w_{\text{Med}} V_{\text{Med}}(X_t, Y_{t+1}) \\ w_{\text{Mode}} V_{\text{Mode}, \delta_T}(X_t, Y_{t+1}) \end{pmatrix}, \quad \text{for } \theta \in \Theta. \quad (3.3)$$

¹⁵Note that $L_{\text{Mode}, \delta}$ differs from L_δ^K in equation (2.6) by a scaling factor of $\delta^{3/2}$. This arises from the convergence rate presented in Theorem 2.6.

The identification functions V_{Mean} and V_{Med} are given in (2.2) and (2.4), and $V_{\text{Mode},\delta}(x, y) = -\delta^{-1/2}K'((x - y)/\delta)$ is an appropriately scaled version of (2.7) that guarantees asymptotic normality in Theorem 2.6 and Theorem 3.3. The limit as $T \rightarrow \infty$ is required in (3.2) to capture the mode through its asymptotic elicibility. Notice that $\phi_{t,T}$ is a triangular array, depending on t and T , through its dependence on the bandwidth δ_T .

In equation (3.3), we allow for normalizations of each identification function using the scalar weights w_{Mean} , w_{Med} and w_{Mode} . This allows one to adjust the importance of each loss function in order to construct tests that are robust to linear data transformations. To do so, we use the inverse of the standard deviations of the respective identification functions in our empirical analysis, though one could instead use equal weights, or some other choice. We define $\hat{\phi}_{t,T}(\theta)$ as in equation (3.3), but using sample-dependent weights $\hat{w}_{T,\text{Mean}}$, $\hat{w}_{T,\text{Med}}$ and $\hat{w}_{T,\text{Mode}}$.

Consider the GMM objective function based on $\hat{\phi}_{t,T}(\theta)$:

$$S_T(\theta) = \left[T^{-1/2} \sum_{t=1}^T \hat{\phi}_{t,T}(\theta) \right]^\top \hat{\Sigma}_T^{-1}(\theta) \left[T^{-1/2} \sum_{t=1}^T \hat{\phi}_{t,T}(\theta) \right], \quad (3.4)$$

where $\hat{\Sigma}_T^{-1}(\theta)$ denotes an $O_P(1)$ positive definite weighting matrix, which may depend on the parameter θ . Unlike the problem in Elliott et al. (2005), the unknown parameter in our framework (the weight vector θ) cannot be assumed to be well identified. For example, for symmetric distributions, the combination weights are completely unidentified. For distributions that exhibit only mild asymmetry a weak identification problem arises. For asymmetric distributions where one measure is a convex combination of the other two (a situation that arises naturally in location-scale processes) we have partial identification of the weight vector. The distribution of economic variables may or may not exhibit asymmetry, and so addressing this identification problem is a first-order concern.

The possibility that the true parameter θ_0 is unidentified, partially identified, or weakly identified implies that the objective function $S_T(\theta)$ may be flat or almost flat in a neighborhood of θ_0 , ruling out consistent estimation of θ_0 . Stock and Wright (2000) show that, under regularity conditions, we can nevertheless construct asymptotically valid confidence bounds for θ_0 , by showing that the objective function S_T evaluated at θ_0 continues to exhibit an asymptotic χ^2 distribution. This

facilitates the construction of asymptotically valid confidence bounds even in a setting where the parameter vector may be strongly identified, weakly identified, or unidentified.¹⁶

We further impose the following regularity conditions on our process.

Assumption 3.2. (A) $\mathbb{E} [|\varepsilon_t|^{2+\delta}] < \infty$ and $\mathbb{E} [|\mathbf{h}_t|^{2+\delta} |\varepsilon_t|^{2+\delta}] < \infty$, (B) $\widehat{w}_{T,\text{Mean}} \xrightarrow{P} w_{\text{Mean}}$, $\widehat{w}_{T,\text{Med}} \xrightarrow{P} w_{\text{Med}}$, and $\widehat{w}_{T,\text{Mode}} \xrightarrow{P} w_{\text{Mode}}$ for some positive weights w_{Mean} , w_{Med} and w_{Mode} ; (C) The limit of $\Sigma_T(\theta_0)$ defined in (3.8) is positive definite. (D) There exists a $\theta_0 \in \Theta$ and triangular arrays $\phi_{t,T}^*(\theta_0)$ and $u_{t,T}(\theta_0)$, such that

$$\phi_{t,T}(\theta_0) = \phi_{t,T}^*(\theta_0) + u_{t,T}(\theta_0), \quad \text{where} \quad (3.5)$$

- (a) $\{T^{-1/2}\phi_{t,T}^*(\theta_0), \mathcal{F}_{t+1}\}$ is a martingale difference array,
- (b) $T^{-1} \sum_{t=1}^T \|u_{t,T}(\theta_0)\|^2 \xrightarrow{P} 0$, and $\sum_{t=1}^T \mathbb{E} [\|T^{-1/2}u_{t,T}(\theta_0)\|^{2+\delta}] \rightarrow 0$,
- (c) $T^{-1} \sum_{t=1}^T u_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top \xrightarrow{P} 0$ and $T^{-1} \sum_{t=1}^T \mathbb{E} [u_{t,T}(\theta_0)\phi_{t,T}(\theta_0)^\top] \rightarrow 0$.

While conditions (A), (B), and (C) are standard, a discussion of (D) is in order. Assumption (D) strengthens the null hypothesis in (3.2) by restricting the dependence structure of the array $\phi_{t,T}(\theta_0)$ such that a CLT can be applied. The decomposition in equation (3.5) implies that the array $T^{-1/2}\phi_{t,T}(\theta_0)$ is an *approximate* MDA in the sense that $T^{-1/2}\phi_{t,T}(\theta_0)$ can be decomposed into a MDA $T^{-1/2}\phi_{t,T}^*(\theta_0)$ and some asymptotically vanishing array $T^{-1/2}u_{t,T}(\theta_0)$. This decomposition is required as the two standard assumptions—mixing and *exact* MDA conditions—are too restrictive for our application. First, the assumption that $\{T^{-1/2}\phi_t(\theta_0), \mathcal{F}_{t+1}\}$ is a MDA does not hold for the baseline case that X_t is an optimal mode forecast (see the proof of Theorem 2.6 for details). Second, imposing mixing conditions is too weak for our case as CLTs for mixing processes generally require finite moments of some order $r > 2$, which is not fulfilled for the mode case as these moments diverge through the bandwidth parameter δ_T .

The intermediate case of Assumption 3.2 (D) allows to apply a CLT based on finite second moments only, and can easily be shown to hold at the three vertices, where the forecast is the

¹⁶Alternative approaches to estimate the confidence sets under partial identification include Kleibergen (2005), Chernozhukov et al. (2007), Beresteanu and Molinari (2008) and Chen et al. (2018) among others.

mean, median or mode. Specifically, 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 $\{T^{-1/2}\phi_{t,T}(\theta_0), \mathcal{F}_{t+1}\}$ is obviously a MDA. When X_t is the true conditional mode of Y_{t+1} , (i.e. $\theta_0 = (0, 0, 1)$), set

$$u_{t,T}(\theta_0) = \mathbb{E}_t[\phi_{t,T}(\theta_0)] = -\omega_{\text{Mode}}\delta_T^{-1/2}\mathbb{E}_t\left[K'\left(\frac{\varepsilon_t}{\delta_T}\right)\right]\mathbf{h}_t. \quad (3.6)$$

Then, the conditions on $u_{t,T}(\theta_0)$ in Assumption 3.2 (D) are fulfilled as shown in Lemma S.1.8 in the Supplemental Appendix.

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 $\{T^{-1/2}\phi_{t,T}(\theta_0), \mathcal{F}_{t+1}\}$ is again a MDA. When X_t is a convex combination with non-zero weight on the mode, Assumption 3.2 (D) is difficult to verify.

Theorem 3.3 below presents the asymptotic distribution of the process $T^{-1/2}\sum_{t=1}^T\widehat{\phi}_{t,T}(\theta_0)$ at the true parameter θ_0 .

Theorem 3.3. *Under Assumptions 2.5, 3.2 and the null hypothesis, $\mathbb{H}_0 : \exists \theta_0 \in \Theta$ such that $\lim_{T \rightarrow \infty} T^{-1/2}\mathbb{E}[\phi_{t,T}(\theta_0)] = 0$ for all $t \leq T$, it holds that*

$$T^{-1/2}\sum_{t=1}^T\widehat{\phi}_{t,T}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta_0)), \quad (3.7)$$

where $\Sigma(\theta_0)$ is the limit as $T \rightarrow \infty$ of

$$\begin{aligned} \Sigma_T(\theta_0) := & \frac{1}{T}\sum_{t=1}^T\mathbb{E}\left[\theta_{10}^2w_{\text{Mean}}\mathbf{h}_t\mathbf{h}_t^\top w_{\text{Mean}}\varepsilon_t^2 + \theta_{20}^2w_{\text{Med}}\mathbf{h}_t\mathbf{h}_t^\top w_{\text{Med}}(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}})^2\right. \\ & + \theta_{30}^2w_{\text{Mode}}\mathbf{h}_t\mathbf{h}_t^\top w_{\text{Mode}}f_t(0)\int K'(u)^2du \\ & \left. + 2\theta_{10}\theta_{20}w_{\text{Mean}}\mathbf{h}_t\mathbf{h}_t^\top w_{\text{Med}}\varepsilon_t(\mathbb{1}_{\{\varepsilon_t>0\}} - \mathbb{1}_{\{\varepsilon_t<0\}})\right]. \end{aligned} \quad (3.8)$$

Under the null hypothesis, Assumption 3.2 (D) implies that $T^{-1/2}\phi_{t,T}(\theta_0)$ (and hence also $T^{-1/2}\widehat{\phi}_{t,T}(\theta_0)$) is an approximate MDA, i.e. this array is approximately (as $T \rightarrow \infty$) uncorrelated. Consequently, we do not need to rely on HAC covariance estimation, and can instead estimate the asymptotic covariance matrix using the simple sample covariance matrix:

$$\widehat{\Sigma}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \widehat{\phi}_{t,T}(\theta) \widehat{\phi}_{t,T}(\theta)^\top. \quad (3.9)$$

The next theorem shows consistency of the outer product covariance estimator, its proof is presented in Supplemental Appendix S.1.

Theorem 3.4. *Given Assumptions 2.5 and 3.2, it holds that $\widehat{\Sigma}_T(\theta_0) - \Sigma_T(\theta_0) \xrightarrow{P} 0$.*

Corollary 3.5. *Given Assumptions 2.5 and 3.2, it holds that $S_T(\theta_0) \xrightarrow{d} \chi_k^2$.*

Following Stock and Wright (2000), this corollary allows one to construct asymptotically valid confidence regions for θ_0 with coverage probability $(1 - \alpha)\%$ by considering the set

$$\{\theta \in \Theta : S_T(\theta) \leq Q_k(1 - \alpha)\}, \quad (3.10)$$

where $Q_k(1 - \alpha)$ denotes the $(1 - \alpha)$ quantile of the χ_k^2 distribution.

Given the above results, we obtain a test for forecast rationality for a general measure of central tendency by evaluating the GMM objective function using a dense grid of convex combination parameters $\theta_j \in \Theta$ for $j = 1, \dots, J$. An asymptotically valid confidence set is given by the values of θ_j for which $S_T(\theta_j) \leq Q_k(1 - \alpha)$. These values represent the centrality measures that “rationalize” the observed sequence of forecasts and realizations, in that rationality cannot be rejected for these measures of centrality. It is possible that the confidence set is empty, in which case we reject rationality at the α significance level for the *entire class* of general centrality measures.

The power of the rationality test depends on the instrument choice. This property is shared with many other tests in the literature (e.g., Elliott et al., 2005; Patton and Timmermann, 2007; Schmidt et al., 2021). Good instruments span the information set of the forecaster and are not strongly correlated with each other. Additional instruments generally improve power asymptotically (or shrink the identified set, in the partial identification case), but can deteriorate power in finite samples. An informative instrument, and one used as far back as Mincer and Zarnowitz (1969), is the forecast X_t itself. In our simulations and applications we found the forecast to be a powerful instrument. Additionally, it is guaranteed to be in the information set of the forecaster, and so is a valid instrument.

The forecast evaluation problem and approach considered here is related to, but distinct from, [Elliott et al. \(2005\)](#). These authors consider the case that a respondent’s point forecast corresponds to some quantile (or expectile) of her predictive distribution. They employ a parametric loss function (“lin-lin” for quantiles, “quad-quad” for expectiles), $L(Y, X; \tau)$ with a scalar unknown parameter (τ) characterizing the asymmetry of the loss. [Elliott et al. \(2005\)](#) use GMM to estimate the τ that best describes the sequence of forecasts and realizations, and test whether forecast rationality holds at the estimated value for τ . Economically, our approach differs from [Elliott et al. \(2005\)](#) in that we consider forecasts only as measures of centrality, allowing for a wide range of such measures, while that paper considers only a single centrality measure nested within a wide range of asymmetric forecasts. Statistically, our approach differs as we are forced to address the feature that our parameter may be partially, weakly, or un-identified, which precludes point estimation.

We use convex combinations of identification functions to parametrically nest the mean, the median, and (asymptotically) the mode. Alternative approaches are possible, e.g., via the L^p norm. Our convex combination approach has the advantage of separating the parameter of interest θ from the bandwidth parameter δ_T , and further does not suffer from technical difficulties such as bandwidth parameters in the exponent, or identification functions with singularities at the mode.

4 Simulation Study

This section studies the finite-sample performance of the methods proposed above. [Section 4.1](#) presents simulations for the mode rationality test and [Section 4.2](#) analyzes the test for rationality for general measures of central tendency.

4.1 Rationality tests for mode forecasts

To evaluate the finite-sample properties of our mode rationality test, we simulate data from the following data generating processes (DGPs). We consider two cross-sectional DGPs, one homoskedastic and the other heteroskedastic, and two time-series DGPs, a simple AR(1) and an AR(1)-GARCH(1,1). We simulate data using the following unified framework:

$$Y_{t+1} = Z_t^\top \zeta + \sigma_{t+1} \xi_{t+1}, \quad \text{where} \quad \xi_{t+1} \stackrel{iid}{\sim} \mathcal{SN}(0, 1, \gamma), \quad (4.1)$$

where $\mathcal{SN}(0, 1, \gamma)$ is a skewed standard Normal distribution, and Z_t denotes a vector of covariates (possibly including lagged values of Y_{t+1}), ζ denotes a parameter vector and σ_{t+1} denotes a conditional variance process. Using this general formulation, the four cases we consider are:

- (1) Homoskedastic cross-sectional data: $Z_t \stackrel{iid}{\sim} \mathcal{N}((1, 1, -1, 2), \text{diag}(0, 1, 1, 0.1))$, where $\zeta = (1, 1, 1, 1)$ and $\sigma_{t+1} = 1$.
- (2) Heteroskedastic cross-sectional data: As in case (1), but $\sigma_{t+1} = 0.5 + 1.5(t + 1)/T$.
- (3) AR(1) data: $Z_t = Y_t$ with $\zeta = 0.5$ and $\sigma_{t+1} = 1$.
- (4) AR(1)-GARCH(1,1) data: As in case (3), but $\sigma_{t+1}^2 = 0.1 + 0.8\sigma_t^2 + 0.1\sigma_t^2\xi_t^2$.

All the above DGPs are based on a skewed Gaussian residual distribution with skewness parameter γ . This choice nests the case of a standard Gaussian distribution at $\gamma = 0$, and in this case all measures of centrality coincide. As the skewness parameter increases in magnitude, the measures of centrality increasingly diverge.

For the general DGP in (4.1), optimal one-step ahead mode forecasts are given by

$$X_t = \text{Mode}(Y_{t+1} | \mathcal{F}_t) = Z_t^\top \zeta + \sigma_{t+1} \text{Mode}(\xi_t), \quad (4.2)$$

where $\text{Mode}(\xi_t)$ depends on the skewness parameter γ .

We consider a range of skewness parameters, $\gamma \in \{0, 0.1, 0.25, 0.5\}$, and sample sizes $T \in \{100, 500, 2000, 5000\}$. In all cases we use 10,000 replications. To evaluate the size of our test in finite samples, we generate optimal mode forecasts through equation (4.2) and apply the mode forecast rationality test based on three choices of instruments: we use the instrument choices $\mathbf{h}_{t,1} = 1$ and $\mathbf{h}_{t,2} = (1, X_t)$ for all DGPs; for the two cross-sectional DGPs, our third choice of instruments is $\mathbf{h}_{t,3} = (1, X_t, Z_{t,1})$, while for the two time-series DGPs we use $\mathbf{h}_{t,3} = (1, X_t, Y_{t-1})$.¹⁷

¹⁷Notice that the choice of instruments $(1, X_t, Y_t)$ is invalid for the AR-GARCH DGP for $\gamma = 0$ as in that case X_t and Y_t are perfectly colinear. Consequently, we use the lag Y_{t-1} as instrument here.

Table 1: Size of the mode rationality test in finite samples

	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	4.7	4.7	5.9	9.1	5.1	5.1	6.1	8.2	5.5	5.3	6.1	7.7
500	5.2	5.9	8.0	8.9	5.3	5.7	7.2	7.8	5.2	5.7	7.3	7.2
2000	5.3	6.9	8.7	7.2	5.3	6.2	7.4	6.8	5.4	6.1	6.8	6.3
5000	5.3	6.1	7.2	6.4	5.1	5.8	6.4	5.7	5.1	5.5	6.4	5.6
	Panel B: Heteroskedastic data											
100	5.1	5.1	6.5	9.3	5.3	5.2	6.2	8.4	5.6	5.7	6.1	7.6
500	6.0	5.9	8.4	8.2	5.7	5.5	7.6	7.4	5.6	5.9	7.2	6.8
2000	5.1	6.8	8.4	6.9	5.2	6.2	7.4	6.3	5.3	6.0	7.1	6.2
5000	5.3	7.2	9.2	6.5	5.1	6.1	7.7	6.0	4.8	5.9	7.1	5.9
	Panel C: Autoregressive data											
100	4.8	4.8	6.0	8.3	5.5	5.1	5.9	7.5	5.8	5.2	5.8	7.1
500	5.6	5.9	7.8	9.3	5.4	5.7	7.1	8.0	5.3	5.2	6.4	7.6
2000	5.5	6.3	8.0	6.8	5.3	5.8	7.1	6.2	5.2	5.7	6.5	5.7
5000	5.1	6.4	7.2	6.5	5.2	6.5	6.5	5.9	5.2	5.9	6.3	5.9
	Panel D: AR-GARCH data											
100	4.7	4.8	5.7	9.7	5.0	4.9	5.8	8.6	5.4	5.4	5.8	8.2
500	5.5	6.4	8.6	9.3	5.7	6.1	7.2	8.4	5.6	5.4	6.8	7.8
2000	5.3	6.5	8.3	6.7	4.8	5.5	7.2	6.2	5.0	5.6	6.6	6.2
5000	5.3	6.6	7.7	6.3	5.2	5.9	6.7	6.1	5.3	6.0	6.6	5.6

Notes: This table presents the empirical rejection rates (in percent) of the mode rationality test using various sample sizes, various levels of skewness in the residual distribution, different choices of instruments, and the four DGPs described in equation (4.1). The nominal significance level is 5%.

Table 1 presents the finite-sample sizes of the test under the different DGPs, sample sizes, instrument choices, and skewness parameters. In all cases we use a Gaussian kernel and set the nominal size to 5%. Results for nominal test sizes of 1% and 10% are given in Table S.1 and Table S.2 in the Supplemental Appendix.

We find that our mode rationality test leads to finite-sample rejection rates that are generally close to the nominal test size, across all of the different choices of DGPs, instruments, and skewness parameters. (In the Supplemental Appendix we present similar results for different significance levels and kernel choices.) Table 1 reveals that an increasing degree of skewness in the underlying

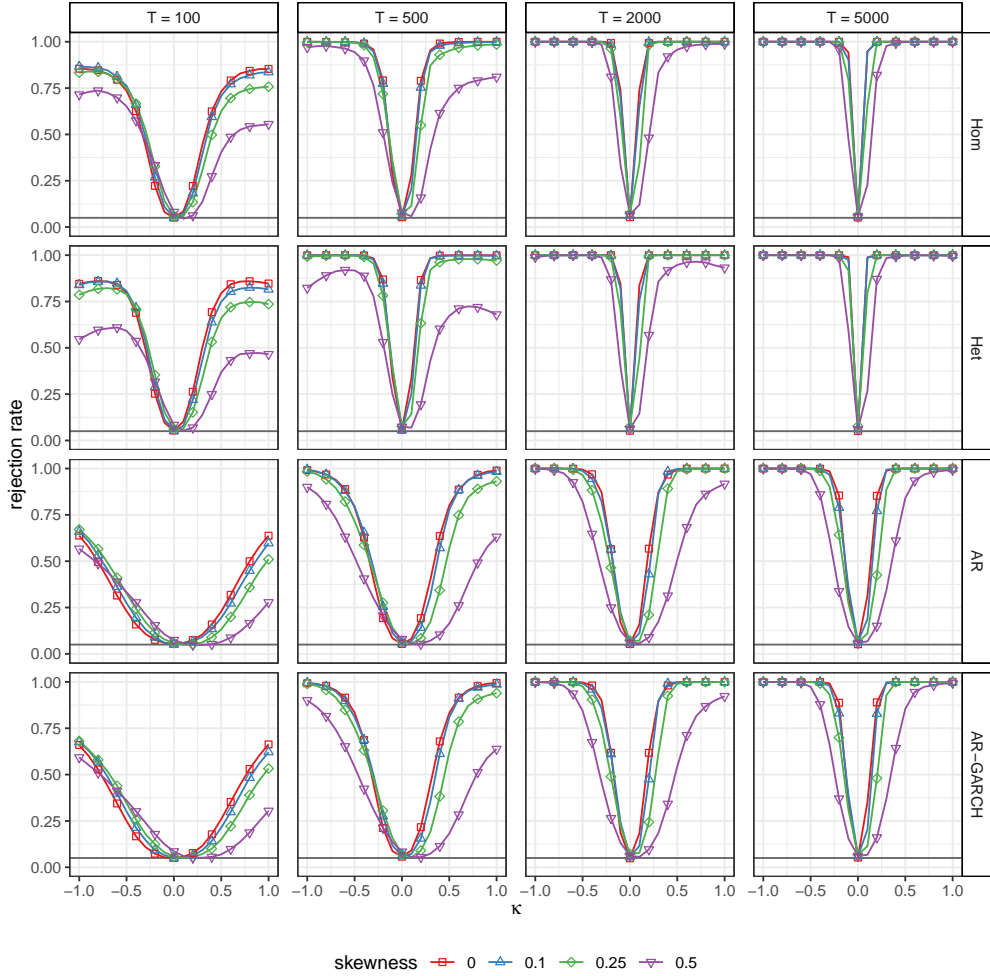


Figure 1: Power study for the “bias” simulation. This figure plots the empirical rejection frequencies against the degrees of misspecification κ for different sample sizes in the vertical panels and for the four DGPs in the horizontal panels. The misspecification follows the design in equation (4.3) and we utilize the instrument vector $(1, X_t)$ and a nominal significance level of 5%.

ing conditional distribution negatively influences the tests’ performance. This is explained by the fact that for more skewed data we choose a smaller bandwidth parameter (following the rule of thumb described in Supplemental Appendix S.3), resulting in less efficient estimates. Consequently, for highly skewed distributions, the mode rationality test requires larger sample sizes in order to converge to the nominal test size.

To analyze the power of the mode forecast rationality test we use the DGPs from (4.1) and consider two forms of sub-optimal forecasts:

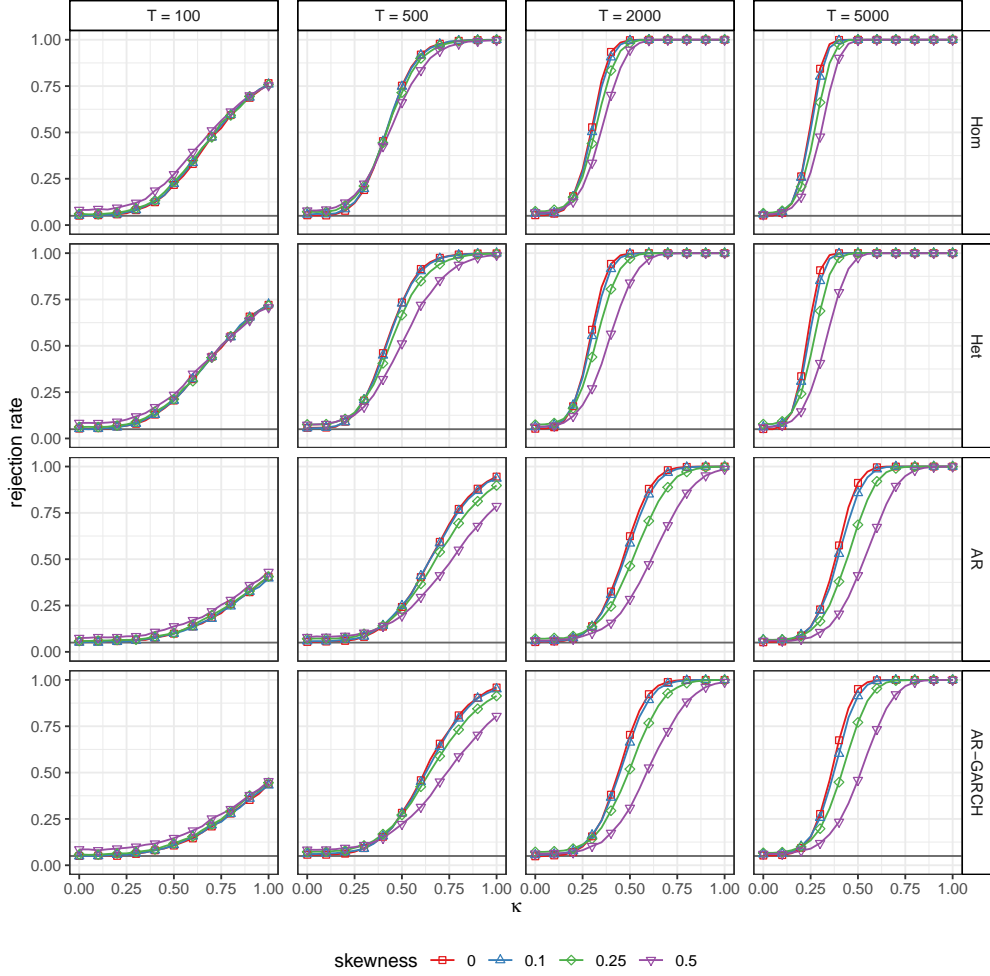


Figure 2: Power for the “noise” simulation. This figure plots the empirical rejection frequencies against the degrees of misspecification κ for different sample sizes in the vertical panels and for the four DGPs in the horizontal panels. The misspecification follows the design in equation (4.4) and we utilize the instrument vector $(1, X_t)$ and a nominal significance level of 5%.

$$\text{Bias: } \tilde{X}_t = X_t + \kappa\sigma_X, \quad \text{where } \sigma_X = \sqrt{\text{Var}(X_t)} \text{ and } \kappa \in (-1, 1) \quad (4.3)$$

$$\text{Noise: } \tilde{X}_t = X_t + \mathcal{N}(0, \kappa\sigma_X^2), \quad \text{where } \sigma_X = \sqrt{\text{Var}(X_t)} \text{ and } \kappa \in (0, 1) \quad (4.4)$$

The first type of misspecification introduces a deterministic bias, where the degree of misspecification depends on the misspecification parameter κ . We standardize the bias using the unconditional standard deviation of the optimal forecasts, $\sqrt{\text{Var}(X_t)}$. The second type of misspecification introduces independent noise, and the magnitude of the noise is regulated through the parameter κ : for $\kappa = 1$, the signal-to-noise ratio is one, and as κ shrinks to zero the noise vanishes.

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

Figures 1 and 2 reveal that the proposed mode rationality test exhibits, as expected, increasing power for an increasing degree of misspecification. Also as expected, larger sample sizes lead to tests with greater power, although even the two smaller sample sizes exhibit reasonable power, particularly in the case of biased forecasts. The figures also reveal that increasing degree of skewness yields to a slight loss of power (given a fixed degree of misspecification). This is driven through the bandwidth choice, where larger values of the (empirical) skewness result in a smaller bandwidth, and consequently a lower test power (analogous to the bias-variance trade-off in the nonparametric estimation literature). Overall, these results show that even though our mode rationality test converges at a slower-than-parametric rate (approximately $T^{2/7}$), it nevertheless has considerable power for the empirically relevant sample sizes of $T \geq 500$; as e.g., in our application in Section 5.

4.2 Rationality tests for an unknown measure of central tendency

In this section we examine the small sample behavior of the asymptotic confidence sets for the measures of central tendency, described in Section 3. As in the previous section, we consider the four DGPs described in and after equation (4.1) and the same varying sample sizes T , skewness parameters γ , and instruments \mathbf{h}_t . We generate optimal one-step ahead forecasts for the mean, median and mode as the true conditional mean, median and mode of Y_{t+1} given \mathcal{F}_t :

$$X_t^{\text{Mean}} = \zeta^\top Z_t + \sigma_{t+1} \text{Mean}(\xi_{t+1}), \quad (4.5)$$

$$X_t^{\text{Med}} = \zeta^\top Z_t + \sigma_{t+1} \text{Median}(\xi_{t+1}), \quad (4.6)$$

$$X_t^{\text{Mode}} = \zeta^\top Z_t + \sigma_{t+1} \text{Mode}(\xi_{t+1}). \quad (4.7)$$

We use the notation $\mathbf{X}_t = (X_t^{\text{Mean}}, X_t^{\text{Med}}, X_t^{\text{Mode}})^\top$ and consider convex combinations of these functionals $X_t = \mathbf{X}_t^\top \beta$, using the following specifications: (a) Mean: $\beta = (1, 0, 0)^\top$, (b) Mode: $\beta =$

$(0, 0, 1)^\top$, (c) Median: $\beta = (0, 1, 0)^\top$, (d) Mean-Mode: $\beta = (1/2, 0, 1/2)^\top$, (e) Mean-Median: $\beta = (1/2, 1/2, 0)^\top$, (f) Median-Mode: $\beta = (0, 1/2, 1/2)^\top$, (g) Mean-Median-Mode: $\beta = (1/3, 1/3, 1/3)^\top$.

For the interpretation of simulation results below, recall that the functional identified by a convex combination of loss functions for the mean, median and the mode (with weights θ) is some convex combination of the mean, median and mode forecasts, but with possibly different combination weights, β . For example, an equal-weighted combination of the mean and the mode forecasts is not necessarily identified by an equal-weighted combination of the mean and mode loss functions, rather it will generally be some other weighted combination of these functions. This makes directly interpreting values of θ difficult, except, of course, when all but one of its elements are equal to zero.

Note also that in some applications, for any given forecast combination weight vector β_0 , there may be infinitely many other forecast combination weights $\tilde{\beta}$, each with a corresponding loss function combination weight $\tilde{\theta}$, that lead to the same forecast. For example, in the right-skewed DGP used here, the three centrality measures are ordered Mode < Median < Mean, and given the functional forms of the optimal forecasts for this DGP presented in equations (4.5)-(4.7), this implies that when the forecast is the median, and so $\beta_0 = (0, 1, 0)^\top$, any weight vector $\tilde{\beta} = (\tilde{\beta}_{\text{Mean}}, \tilde{\beta}_{\text{Med}}, \tilde{\beta}_{\text{Mode}})^\top \in \Theta$ that satisfies

$$\tilde{\beta}_{\text{Mean}} = \frac{\text{Median}(\xi_{t+1}) - \text{Mode}(\xi_{t+1})}{\text{Mean}(\xi_{t+1}) - \text{Median}(\xi_{t+1})} \cdot \tilde{\beta}_{\text{Mode}} \quad (4.8)$$

will lead to a combination forecast that coincides with the median forecast. Equation (4.8) defines a line through the unit simplex, starting at the median vertex and ending somewhere on the edge connecting the mean and mode vertices.¹⁸ This again highlights the identification problems that can arise in our analysis of forecasts of measures of central tendency. In contrast, the mean and mode centrality measures in this DGP each have a unique identification function combination weight vector, equal to the associated forecast combination weight vector.

¹⁸When the DGP is conditionally location-scale, one of the three centrality measures can always be expressed as a (constant) convex combination of the other two. In this DGP, this is the median, but in other applications it may be any of the three measures. When the conditional distribution exhibits variation in higher-order moments or other “shape” parameters, this restriction will generally not hold, and the variation may or may not be sufficient to separately identify the three centrality measures.

Analyzing the coverage properties of this method requires knowledge of the set of identification function weights θ corresponding to the forecast weights β used to construct the forecast. In general, θ is not known in closed-form; we use one thousand draws from each DGP and skewness level to numerically obtain the identification function weights corresponding to each set of forecast combination weights.

Table 2 shows the empirical coverage rates for the confidence sets of centrality measures for the seven simulated convex combinations of functionals, different sample sizes, skewness parameters, and DGPs. When the set of identification function weights corresponding to a particular forecast combination weight vector is not a singleton, we choose the mid-point of the line that defines this set.¹⁹ In the left panel of Table 2 the data is unimodal and symmetric, and the measures of central tendency coincide, making all seven forecasts identical. The test outcomes, however, can differ as each row uses a different set of moment conditions to evaluate forecast rationality. We see that in all cases the coverage rates are very close to the nominal 90% level. In the right panel of Table 2 the data is asymmetric and the measures of central tendency differ. The coverage rates remain close to the nominal 90% level, especially for larger sample sizes.

Figure 3 illustrates the average rejection (coverage) rates based on 90% confidence sets for the central tendency measures across a richer set of combination weights. (We omit the mean-median combination forecast from this figure in the interest of space.) This figure uses the AR-GARCH DGP; equivalent results for the homoskedastic cross-sectional DGP are shown in Figure S.2 in the Supplemental Appendix. Each point in the triangles corresponds to a tested centrality measure, i.e. to one value of θ , and for each point we compute how often it is contained in the 90% confidence set. We depict a coverage rate between 85% and 100% by a black point, a coverage rate between 50% and 85% by a grey point and anything below 50% by a white point. We use a cut-off of 85% to include points with coverage rates “close” to the nominal rate of 90%. We use a sample size $T = 2000$, instruments $\mathbf{h}_t = (1, X_t)$ and a Gaussian kernel. The upper panel presents results for the DGPs with zero skewness and the lower panel considers skewness of $\gamma = 0.5$.

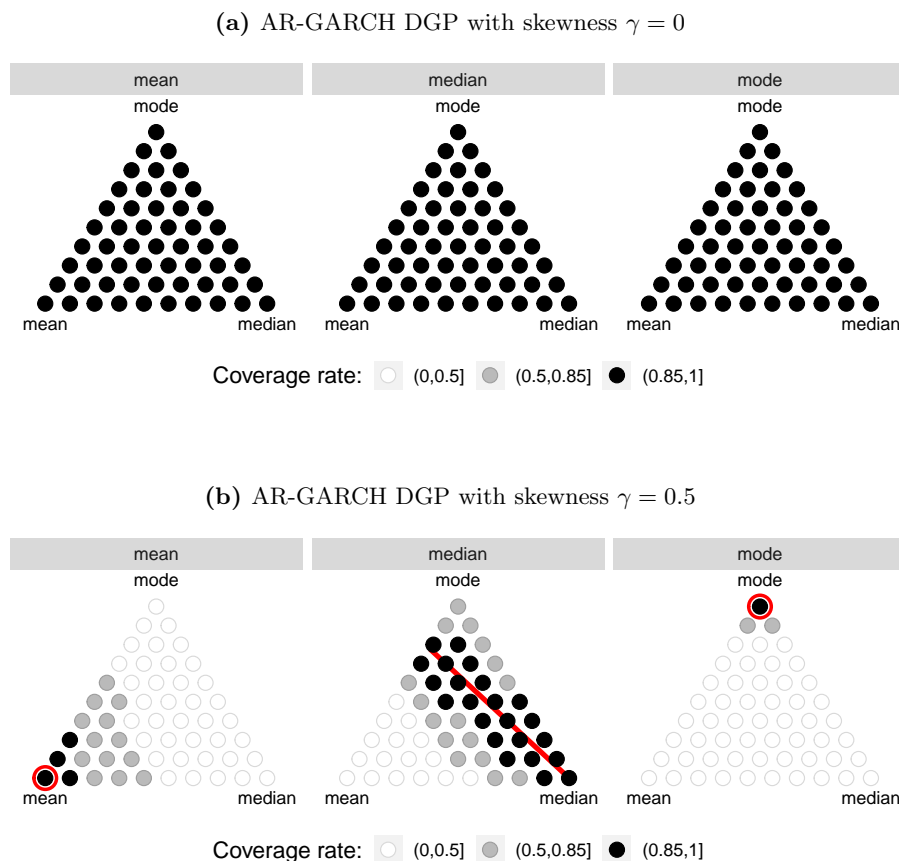
¹⁹For this DGP, θ is a singleton only when the forecast is the mean or the mode. Table S.3 and Table S.4 in the Supplemental Appendix show finite-sample rejection rates for three values of θ : the end-points of the line defining the set, and the mid-point reported in Table 2. In all cases we find very similar results for all three points.

Table 2: Coverage of the central tendency confidence sets in finite samples

Sample size	Symmetric data				Skewed data			
	100	500	2000	5000	100	500	2000	5000
Centrality measure	Panel A: Homoskedastic iid data							
Mean	89.3	90.0	89.8	89.2	89.4	90.2	89.6	90.2
Mode	90.1	89.1	89.8	90.1	85.7	86.0	87.6	88.8
Median	89.5	89.0	89.6	89.6	90.3	91.9	91.4	91.2
Mean-Mode	90.1	89.0	89.4	90.0	90.7	92.2	92.0	91.9
Mean-Median	89.8	89.6	89.6	89.5	90.4	91.1	90.3	90.4
Median-Mode	89.9	89.0	89.6	89.8	90.1	90.9	90.7	91.4
Mean-Median-Mode	90.1	88.9	89.5	89.8	90.9	92.3	92.3	91.8
	Panel B: Heteroskedastic data							
Mean	89.2	89.9	89.6	90.3	88.7	89.6	89.4	89.5
Mode	89.5	89.5	89.9	89.8	85.8	86.4	87.9	88.7
Median	89.0	89.8	90.1	89.9	90.0	90.9	90.8	90.7
Mean-Mode	89.2	89.5	89.8	90.0	90.0	91.1	91.0	90.9
Mean-Median	89.2	89.6	90.0	90.5	89.3	90.2	89.7	90.0
Median-Mode	89.3	89.4	89.8	90.0	89.0	90.0	90.3	90.8
Mean-Median-Mode	89.2	89.7	89.9	90.1	90.3	91.2	91.1	91.3
	Panel C: Autoregressive data							
Mean	89.2	90.1	89.9	90.3	89.4	89.4	89.8	90.3
Mode	89.2	89.8	89.3	89.9	85.4	86.2	88.0	88.7
Median	89.0	89.8	89.1	89.7	90.4	90.7	92.3	90.7
Mean-Mode	88.9	89.9	89.2	89.7	90.5	91.6	92.5	91.3
Mean-Median	88.8	90.2	89.3	89.9	90.0	89.8	91.2	89.7
Median-Mode	89.1	89.9	89.1	89.9	89.4	90.5	91.5	91.1
Mean-Median-Mode	89.0	89.7	89.2	89.6	90.7	91.7	92.7	91.7
	Panel D: AR-GARCH data							
Mean	89.2	90.4	89.9	90.0	89.5	89.9	89.9	89.8
Mode	90.0	89.0	89.6	89.9	85.9	86.1	88.4	88.9
Median	88.8	89.4	89.3	89.5	90.1	91.2	91.3	90.8
Mean-Mode	89.6	89.1	89.6	89.6	90.3	91.3	92.3	91.3
Mean-Median	88.7	89.7	89.3	89.4	90.1	90.7	90.8	89.9
Median-Mode	90.1	89.0	89.5	89.7	89.4	90.6	92.0	91.1
Mean-Median-Mode	89.6	89.2	89.3	89.7	90.4	91.7	92.3	91.2

Notes: This table presents empirical coverage rates (in percent) of nominal 90% confidence sets for the forecasts of central tendency. We report the results for symmetric and skewed data ($\gamma = 0$ or 0.5), for various sample sizes and the four DGPs described in equation (4.1). As instruments we use $\mathbf{h}_t = (1, X_t)$.

Figure 3: Coverage rates of the confidence regions for central tendency measures



This figure shows coverage rates of 90% confidence regions for the measures of central tendency for the AR-GARCH DGP. The true forecasted functional is given in the text above each triangle. The circles 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 $\mathbf{h}_t = (1, X_t)$ and use a Gaussian kernel.

Panel (a) of Figure 3 uses a unimodal DGP with zero skewness, and so all measures of central tendency and all convex combinations thereof coincide. This implies that all three of these triangles are identical; we include them here for ease of comparison with the lower panel, where the optimal forecasts differ. (The points inside each triangle need not be identical, as they represent tests for optimality of different centrality measures.) Under symmetry and unimodality, every point in the

triangle should be contained in the confidence set with probability 90%, and the figure is consistent with this, thus confirming the procedure’s coverage level in this simulation design.

Panel (b) of Figure 3 considers nonzero skewness, and the measures of central tendency differ. In the two point-identified cases (mean and mode), we use a red circle to highlight the single centrality measure that corresponds to the forecast. The case for the median is only partially identified and we use a red line to indicate the set of centrality measures that correspond to the forecast.

The left plot of Panel (b) considers optimal mean forecasts and exhibits the expected behavior: the mean, and convex combinations close to the mean, are usually contained in the confidence set, whereas points far away are usually excluded. Note that the mode rationality test has sufficient power, in this setting, against a true mean forecast to consistently reject it. A similar, but more pronounced, picture can be observed for optimal mode forecasts in the right plot.

The middle plot in Panel (b) of Figure 3 reveals, as expected, that the optimal median forecast is generally not rejected when testing using the identification function for the median (revealed by the dot at the median vertex being black). This plot further shows that the convex combinations of mean, median and mode that coincide with the median (see equation (4.8), represented in Figure 3 by a red line) are also included in the confidence set, as are convex combinations that are close to these, leading to a region of black or grey dots around the red line.

5 Evaluating Survey Forecasts of Individual Income

We apply our proposed tests to survey responses to the Federal Reserve Bank of New York’s Survey of Consumer Expectations.^{20,21} We focus on the Labor Market Survey component, which is conducted each March, July, and November, and which asks participants a variety of questions,

²⁰Source: Survey of Consumer Expectations, © 2013-2019 Federal Reserve Bank of New York (FRBNY). The SCE data are available without charge at <http://www.newyorkfed.org/microeconomics/SCE> and may be used subject to license terms posted there. FRBNY disclaims any responsibility or legal liability for this analysis and interpretation of Survey of Consumer Expectations data.

²¹In Supplemental Appendix S.8 we consider two other empirical applications. The first is to the “Greenbook” forecasts of US GDP growth produced by economists at the Board of Governors of the Federal Reserve. In that application we find that the Greenbook forecasts are rational mean forecasts, but not mode or median forecasts. Our second application is to random walk forecasts of exchange rates, in the spirit of Meese and Rogoff (1983). We find random walk forecasts to be rational mean forecasts for three different exchange rates, while the mode and median are rejected for two of the three exchange rates.

Table 3: Evaluating income survey forecasts

vector \mathbf{h}_t	Centrality measure		
	Mean	Median	Mode
1	0.008	0.401	0.737
1, X	0.014	0.000	0.768
1, X , lagged income	0.031	0.000	0.970
1, X , government sector	0.064	0.000	0.949
1, X , private sector	0.047	0.000	0.898
1, X , job offers	0.006	0.000	0.913

Notes: This table presents p -values from tests of rationality of individual income forecasts from the New York Federal Reserve’s Survey of Consumer Expectations. The columns present test results when interpreting the point forecasts as forecasts of the mean, median or mode. The rows present results for different choices of instrument vectors \mathbf{h}_t used in the test: 1 is the constant, X is the forecast, “lagged income” is the respondent’s income at the time of the forecast, “government sector” and “private sector” are indicators for the self-reported industry in which the respondent works, “job offers” is an indicator for whether the respondent received any job offers in the previous four months.

including about their current earnings and their beliefs about their earnings in four months (i.e., the date of the next survey). Using adjacent surveys over the period March 2015 to March 2018, we obtain a sample of 3,916 pairs of forecasts (X_t) and realizations (Y_{t+1}).^{22,23} In testing the rationality of these forecasts we initially assume that all participants report the same, unknown, measure of centrality as their forecast. In the next subsection, we explore potential heterogeneity in the measure of centrality used by different respondents.

Table 3 presents the results of rationality tests for three measures of central tendency, and for a variety of instrument sets. The first instrument set includes just a constant, and the rationality test simply tests whether the forecast errors have unconditional mean, median or mode, respectively, of zero. The other instrument sets additionally include the forecast (X_t) itself, and other information about the respondent collected in the survey. We consider the respondent’s income at the time of making the forecast, indicators for the respondent’s type of employer,²⁴ and whether the respondent

²²We drop observations that include forecasts or realizations of annualized income below \$1,000 or above \$1 million, which represent less than 1% of the initial sample. We also drop observations where the ratio of the realization to the forecast, or its inverse, is between 9 and 13, to avoid our results being affected by misplaced decimal points or by the failure to report annualized income (leading to proportional errors of around 10 to 12 respectively).

²³Most respondents in our sample appear just once: the 3,916 forecast-realization pairs come from 2,628 unique individuals. Our econometric approach does not require repeated samples and so this causes no difficulty, and our results are qualitatively unchanged if we use only the first forecast-realization pair from each respondent.

²⁴The survey includes the categories government, private (for-profit), non-profit, family business, and “other.” The first two categories dominate the responses and so we only consider indicators for those.

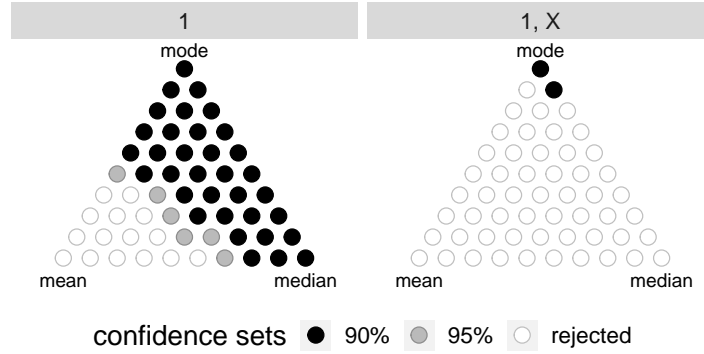


Figure 4: Confidence sets for income survey forecasts. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts. The circles that comprise each the triangle correspond to specific convex combinations of the vertices, which are the mean, median and mode functionals. 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. The left panel uses only a constant as the instrument; the right panel uses a constant and the forecast.

received any job offers in the past four months.

The first row of Table 3 shows that when only a constant is used, rationality of the survey forecasts can be rejected for the mean, but cannot be rejected for the other two measures of central tendency. When we additionally include the forecast as an instrument, we can reject rationality as mean or median forecasts, but we cannot reject rational mode forecasts. We are similarly able to reject rationality as mean and median forecasts when we include additional covariates, but find no evidence against rationality when these forecasts are interpreted as mode forecasts.

Figure 4 shows the convex combinations of mean, median and mode forecasts that lie in the confidence set constructed using the methods for weakly-identified GMM estimation in [Stock and Wright \(2000\)](#).²⁵ In the left panel we see that when using only a constant as the instrument, we are able to reject rationality for the mean, and for measures of centrality “close” to the mean, but we are unable to reject rationality for the median or mode and measures in a neighborhood of these. This is consistent with the entries in the first row of Table 3, which correspond directly to the three vertices in Figure 4. In the right panel of Figure 4, when the instrument set includes a constant and

²⁵The interpretation of this figure is slightly different to that of Figure 3: in that figure the shade of each dot indicated the proportion of times, across simulations, that point was included in the confidence set, allowing us to study the finite-sample coverage rates of our procedure. In Figure 4 the shade of each dot indicates whether, for this sample, that point is included in the 90% confidence set, the 95% confidence set, or is outside the 95% confidence set, the latter indicating a rejection of rationality at the 5% significance level.

the forecast, we see that only the mode and centrality measures very close to the mode are included in the confidence set; all other forecasts can be rejected at the 5% level. Overall, we conclude that the responses to the New York Fed’s income survey, taken on aggregate, are consistent with rationality when interpreted as mode forecasts, but not when interpreted as forecasts of the mean, median, or convex combinations of these measures.

When interpreting the results in Figure 4, and similar figures below, it is worth keeping in mind that the power to detect sub-optimal forecasts is not uniform across values of θ : sampling variation in the mean and median vanishes at rate $T^{-1/2}$, while for the mode it vanishes only at rate approximately $T^{-2/7}$ (see Theorem 2.9). This implies that for comparably sub-optimal forecasts, power will be lower at the mode vertex than at the mean or median vertices. This unavoidable variation in power means that the information conveyed by inclusion in the confidence set differs across values of θ .²⁶

The analysis of individual income survey forecasts above used 3,916 pairs of forecasts and realizations from a total of 2,628 unique survey respondents. This naturally raises the question of whether there is heterogeneity in the measure of centrality used by different respondents.²⁷ Given that our survey respondents generally only appear once or twice in our sample, allowing for arbitrary heterogeneity is not empirically feasible. In our analysis below we instead stratify survey respondents by observable characteristics and test forecast rationality separately for each subsample. Stratifying the sample may reduce power, due to the smaller sample sizes available, or it may reveal irrationality that is undetected in a joint analysis, for example if one subsample deviates strongly from rationality while the remaining subsamples are rational.

5.1 Forecast rationality and income level

Firstly, we consider stratifying our sample by income. This is motivated by the possibility that, in addition to a different *level* of future income, low-income respondents face a different *shape*

²⁶Stock and Wright (2000) suggest caution when interpreting a small but nonempty confidence set, as such an outcome is consistent with both a correctly-specified model (a rational forecast, in our case) estimated precisely and also with a misspecified model (irrational forecast) facing low power. These two interpretations clearly have very different economic implications, but cannot be disentangled empirically. Given the lower power at the mode vertex, the latter explanation may be relevant here.

²⁷Heterogeneity in the underlying predictive distributions, F_t , is accommodated by our approach.

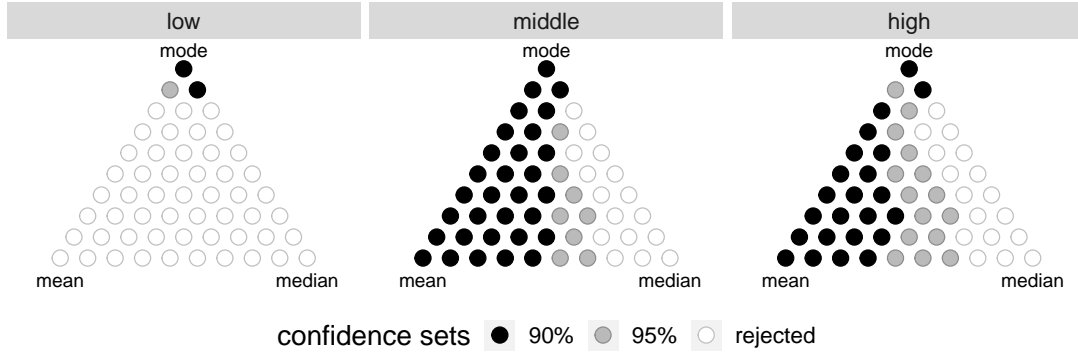


Figure 5: Confidence sets for income survey forecasts, stratified by income. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts, for low-, middle- and high-income respondents. Groups are formed using terciles of lagged reported income. 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.

of future income, compared with high-income respondents. This analysis may also reveal that respondents at different income levels use different centrality measures to summarize their predictive distribution. Figure 5 presents confidence sets for forecast rationality of measures of centrality for low-, middle-, and high-income respondents based on terciles of the distribution of reported income.²⁸ We see that for low-income respondents, only the mode and measures very close to the mode are contained in the confidence set. For middle- and high-income respondents, the mode, mean, and centrality measures “close” to the mean and mode are included in the confidence set. This finding is consistent with all respondents using the mode, and only the distribution for low-income respondents allowing for separate identification of the mode and the mean. Specifically, if the conditional distribution of future earnings is more skewed for low-income respondents than for high-income respondents, then the measures of centrality are better separated for the former, which allows for better identification of the functional used by survey respondents. As income is commonly correlated with numeracy and education levels, rejecting mean- and median-rationality for lower-income respondents is consistent with work in the experimental literature: Kröger and Pierrot (2019a) observe high-numeracy individuals to be better at mean and median forecasting.²⁹

²⁸Qualitatively similar results are found if we stratify the sample into just two groups based on median income.

²⁹Given the panel structure of our data, the target variable may be subject to common unpredictable shocks, leading to forecast errors correlated across individuals within the same wave. In Supplemental Appendix Section S.6, we repeat all of our analyses using a covariance estimator clustered at the time level and find very similar results.

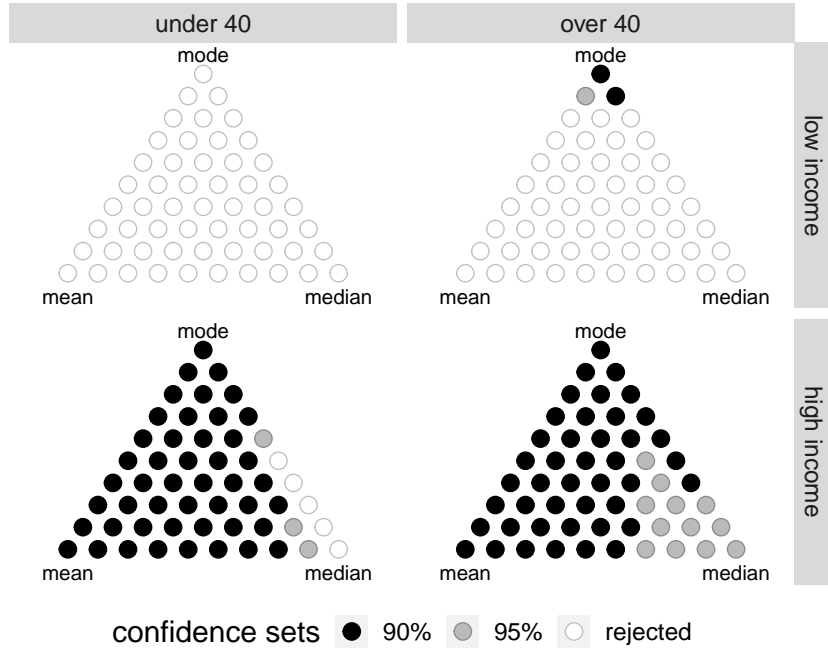


Figure 6: Confidence sets for income survey forecasts, stratified by age and income. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts, for low- and high-income respondents who are below or above the age of 40. Income groups are formed using the median lagged reported income. 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.

5.2 Forecast rationality and age

We next stratify the sample both by income and by age, motivated by the possibility that younger respondents have less experience in the workforce and may be less able to predict their future earnings. The latest versions of the FRBNY survey contain a field for whether the respondent is above or below 40 years of age, while earlier versions asked for the respondent’s specific age. To maximize coverage, we adopt the age 40 split contained in the later versions of the survey. We have 2,457 (1,332) respondents over (under) that age. We use the median income within each age category to define “low” and “high” income groups.

Figure 6 presents the striking result that forecasts from younger low-income workers cannot be rationalized using *any* measure of centrality; all points lie outside the 95% confidence set. Forecasts from older low-income respondents can be rationalized as only mode forecasts. In contrast, forecasts

from both young and old high-income workers can be rationalized for a variety of measures of centrality.³⁰ This figure suggests that younger low-income workers make systematic errors when attempting to predict their income in the coming four months. Asymmetric loss functions, such as those used in [Elliott et al. \(2005\)](#) (EKT), have the potential to explain these forecasts, and we consider this in Section 5.6 below. Foreshadowing those results, we also reject rationality of forecasts from younger low-income respondents in the EKT framework.

5.3 Forecast rationality and job stability

An important component of income uncertainty is job search and new job offers, see [Mueller and Spinnewijn \(2022\)](#) for a recent review of survey expectations in the job search literature. Respondents naturally have different prospects of changing jobs, and these changes impact their predictive distributions of income. To proxy for the likelihood of changing jobs, we stratify respondents based on whether they reported receiving at least one job offer over the previous four months.³¹

Figure 7 reveals that forecasts from low-income respondents who reported receiving a job offer in the past four months can only be rationalized as mode forecasts; rationality for all other centrality measures is rejected. We find that respondents who did not receive a job offer, and high-income respondents who did, provide forecasts that can be rationalized as the mean, mode, and combinations thereof. These results reveal that low-income respondents with more job uncertainty differ meaningfully from the other three subgroups.

5.4 Forecast rationality and survey experience

As economic agents gather experience, their actions and expectations change, see [Van Nieuwerburgh and Veldkamp \(2006\)](#) and [Malmendier and Nagel \(2011\)](#) for example. In recent work, [Kim and Binder \(2023\)](#) show that this effect is true in economic surveys as well: forecasts of inflation show

³⁰Note that a large confidence set for θ does not imply that the respondents are “more rational” than if the confident set were small; a respondent can be fully rational at just a single value of θ , and if the predictive distribution is sufficiently skewed then the confidence set will, in the limit, contain just that value of θ . A large confidence set may reflect a predictive distribution that does not allow for the identification of a single value of θ , or low finite-sample power.

³¹We find very similar results when we stratify respondents using their estimated probability of receiving a job offer in the next four months, or by their estimated probability of staying in the same job for the next four months.

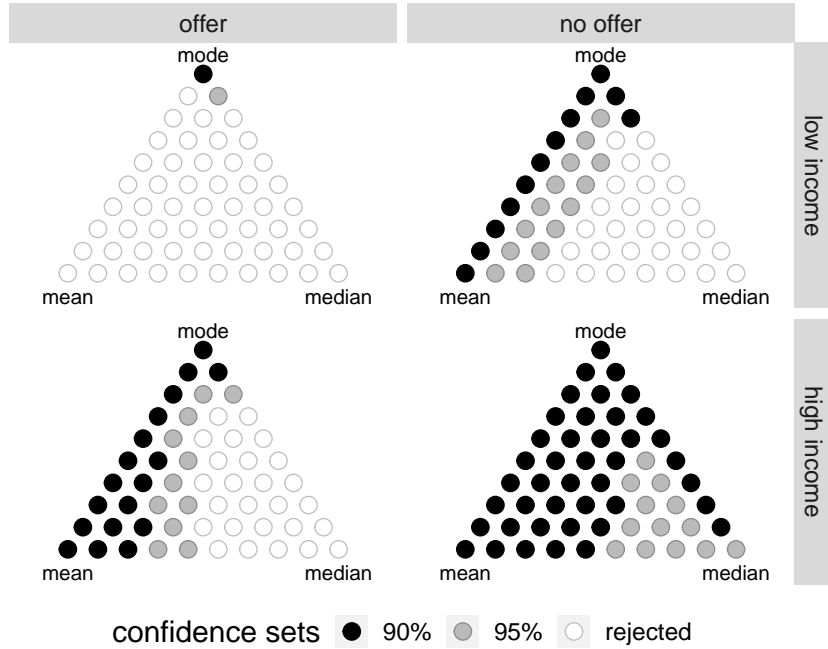


Figure 7: Confidence sets for income survey forecasts, stratified by job offer status and income. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts, for low- and high-income respondents that received a job offer or not. Groups are formed using the median lagged reported income, and whether or not the respondent reported receiving at least one job offer in the past four months. 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.

less uncertainty as respondents garner experience responding to the survey. We next investigate whether having survey experience leads to more rational forecasts. The FRB NY survey includes participants for a maximum of twelve months, and within that period they are asked to predict future income and report current income three times, providing us with up to two matched pairs of forecasts and realizations for each respondent. We have a total of 1,288 respondents with two matched pairs, and we study the rationality of the first and second of these forecasts separately.

Figure 8 reveals a stark difference in the impact of the round of elicitation between low- and high-income respondents. Forecasts from low-income respondents can be rationalized as mode forecasts, and the results change only slightly from the first to the second round: in the first round some “near mode” functionals are also rationalizable, and in the second round the mode functional is only rationalizable at the 95% significance level. For high-income respondents, in contrast, only

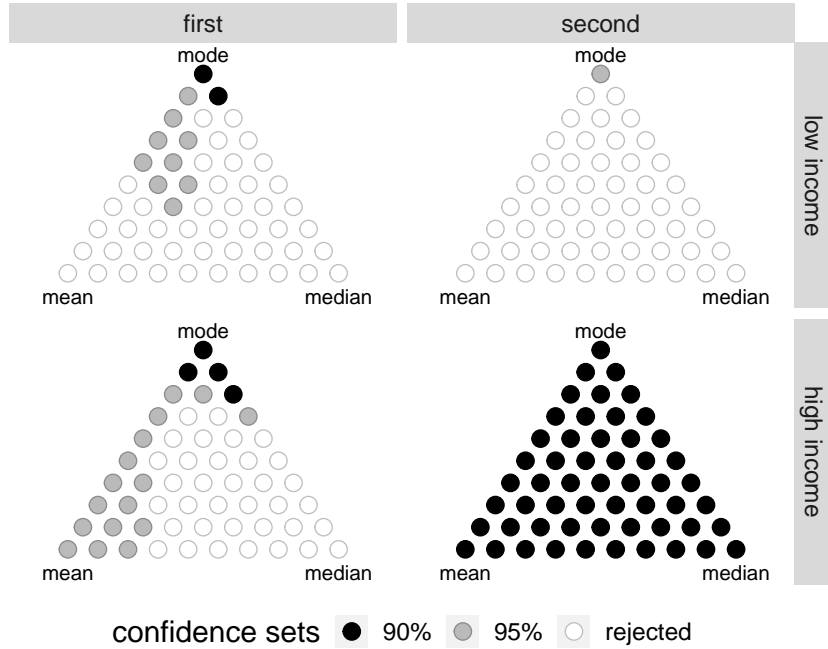


Figure 8: Confidence sets for income survey forecasts, stratified by elicitation round and income. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts, for low- and high-income respondents in their first and second panel round. Black dots 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.

the mode and mean functionals (and convex combinations thereof) are rationalizable in the first round, while in the second round, four months later, all functionals are rationalizable. Assuming that the *true* conditional distributions of future income are unaffected by survey participation, and from the second round results we can infer these are (approximately) symmetric, this suggests that high-income respondents’ forecasts are more accurate in the second round, and as a result can be (correctly) rationalized via more functionals.³² Our results for high-income respondents are consistent with the findings of [Kim and Binder \(2023\)](#) for inflation forecasts, while the similarity across survey rounds for low-income respondents differs from that study.

³²Recall that a large confidence set does not imply “more rational,” merely that the forecast is rationalizable by a wider variety of measures of centrality.

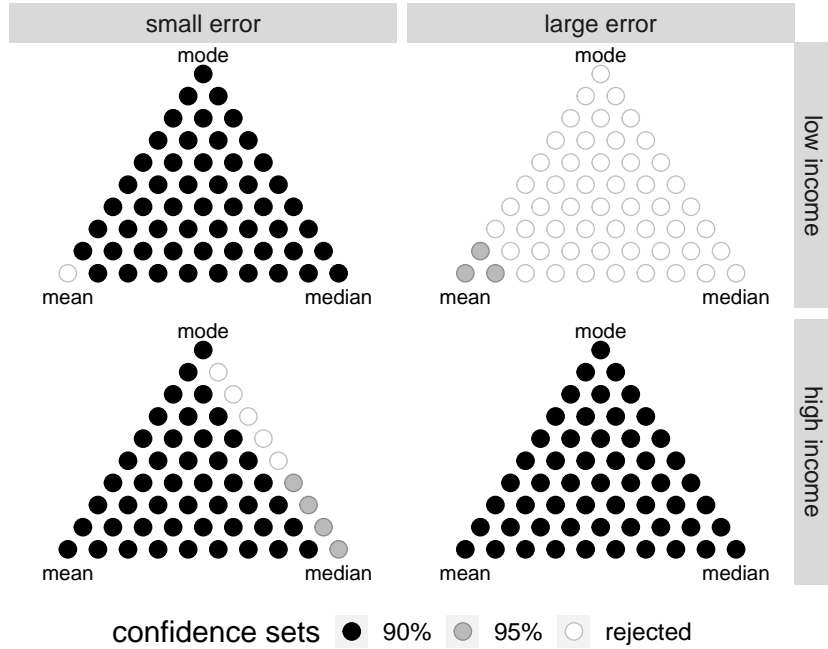


Figure 9: Confidence sets for income survey forecasts, stratified by forecast error size and income. This figure shows the measures of centrality that “rationalize” the New York Federal Reserve income survey forecasts, for low- and high-income respondents with small and large past forecast errors. Black 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.

5.5 Forecast rationality and past forecast accuracy

Finally, we investigate the relationship between past forecast accuracy and the rationality of subsequent forecasts. [D’Acunto et al. \(2019\)](#) find absolute forecast errors to be negatively correlated with cognitive ability measures for inflation expectations, which may suggest that respondents with larger past forecast errors are more likely to issue irrational forecasts. In contrast, [Van Nieuwerburgh and Veldkamp \(2006\)](#) propose that economic agents learn more about the economy when more signals are available, and a large forecast error may signal to the respondent that they need to pay more attention to their forecast. We address this question using the set of 1,288 respondents to the FRBNY survey for which we have two matched pairs of predictions and realizations. We classify respondents as having “small” or “large” past forecast errors using the cross-respondent median of absolute percentage forecast errors in the first round as a cutoff. In addition to the (relative) size of their past forecast errors, we continue to stratify by income.

Figure 9 shows that respondents who had a small forecast error in the first round provide second-round forecasts that can be rationalized using many different measures of centrality, and there are few differences across low- and high-income respondents: for low-income respondents, only the mean leads to a rejection, while for high-income respondents the median and measures between the median and the mode are rejected. For respondents with a large forecast error in the first round, the rationality test results are starkly different for low and high income respondents: high-income respondents can be rationalized using any measure of centrality, while low-income respondents can be rationalized only as mean, or near-mean, forecasters, but only at the 95% confidence level; at the 90% confidence level rationality is rejected for all centrality measures. An alternative interpretation of Figure 9 is that high-income respondents produce forecasts that are rationalizable for almost all measures of centrality, regardless of past forecast accuracy, while the rationality of low-income respondents' forecasts varies greatly with past forecast errors: those with large past forecast errors are not rationalizable as any measure of centrality at the 90% confidence level, while those with small forecast errors are rationalizable as almost any measure of centrality.

5.6 Irrational or optimistic?

The premise of the rationality test proposed in this paper is that survey respondents use a measure of central tendency when providing a point forecast but the specific centrality measure used is unknown. An alternative framework is to model the respondent as reporting either a centrality measure *or* some other functional of their predictive distribution. For example, rather than using the median they may use some other quantile, capturing optimism or pessimism. Elliott et al. (2005) (EKT) consider such a case, and model respondents as using either “lin-lin” loss, which elicits a quantile of the predictive distribution, or “quad-quad” loss, which elicits an expectile of the predictive distribution. In both cases, the loss function contains an asymmetry parameter, and the special case of symmetry corresponds to the respondent using the median or the mean.

Table 4 presents tests of rationality in the EKT framework, as well as the results for the three measures of centrality considered in this paper. These are presented for the full sample, the three income subsamples, and the 2×2 stratifications based on age and income. Results for the other

Table 4: Summary of p -values for rationality tests in different samples

	n	Mean	Median	Mode	Quantiles	Expectiles
Full sample	3916	0.01	0.00	0.77	0.00 [0.61, 0.64]	0.01 [0.49, 0.56]
Low income	1263	0.03	0.00	0.89	0.00 [0.56, 0.61]	0.01 [0.45, 0.57]
Middle income	1263	0.13	0.00	0.92	0.00 [0.62, 0.66]	0.06 [0.40, 0.54]
High income	1263	0.15	0.02	0.46	0.05 [0.62, 0.67]	0.06 [0.47, 0.57]
Under 40, Low income	665	0.00	0.00	0.00	0.00 [0.56, 0.62]	0.00 [0.63, 0.80]
Over 40, Low income	1236	0.05	0.00	0.61	0.00 [0.59, 0.64]	0.02 [0.47, 0.59]
Under 40, High income	667	0.35	0.03	0.83	0.01 [0.61, 0.68]	0.15 [0.42, 0.57]
Over 40, High income	1221	0.15	0.09	0.98	0.26 [0.62, 0.67]	0.05 [0.44, 0.56]

Notes: This table presents 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 panels present p -values from tests of rationality when interpreting the point forecasts as quantiles or expectiles following Elliott et al. (2005), as well as 90% confidence intervals for the asymmetry parameter, given in square brackets.

stratifications considered above are presented in Supplementary Section S.7. As in our previous analysis, we use a constant and the forecast as instruments for the EKT tests.

For the full sample of respondents, in the top line of Table 4, we see that rationality is rejected in the EKT framework. When interpreted as a quantile forecast, the respondents' estimated quantile 90% confidence interval is 0.61 to 0.64, indicating optimism, but the p -value for the test of rationality is less than 0.005, rejecting rationality.³³ When interpreted as an expectile forecast, the shape parameter straddles 0.5 (the case of symmetry, where the expectile equals the mean) but the test again rejects rationality. The only functional for which rationality is not rejected is the mode, using the new mode forecast rationality test proposed in this paper.

The results of the rationality tests stratified by income, presented in the second panel of Table 4, show that middle- and high-income respondents can be rationalized as expectile forecasters, and in both cases the confidence interval for the estimated shape parameter includes the point of symmetry, consistent when the mean functional not being rejected for these subsamples. For low-income respondents, we see that allowing for optimism or pessimism through the EKT framework does not rationalize these forecasts: the p -values of the EKT rationality tests for these cases are both less than or equal to 0.01.

³³The estimated shape parameter is interpretable as the one that makes the observed forecasts as close as possible to rational, and the corresponding p -value determines whether these forecasts are consistent or inconsistent with rationality.

Finally, we stratify survey respondents by both age and income, and present the results in the lower panel of Table 4. Recall that above we found that the younger, low-income subsample could not be rationalized as *any* measure of centrality. When using the EKT framework, we find that the estimated shape parameters for the lin-lin and quad-quad loss functions are both greater than 0.5, the point of symmetry, indicating optimistic forecasts. However the p -values in both cases are less than 0.005, strongly rejecting rationality. Thus, forecasts from younger low-income respondents cannot be rationalized as centrality measures, nor as quantiles or expectiles.

6 Conclusion

Reasonable people can interpret a request for their prediction of a random variable in a variety of ways. Some, including perhaps most economists, will report their expectation of the value of the variable (i.e., the mean of their predictive distribution), others might report the value such that the observed outcome is equally likely to be above or below it (i.e., the median), and others may report the value most likely to be observed (i.e., the mode). Still others might solve a loss minimization problem and report a forecast that is not a measure of central tendency. Economic surveys generally request a *point* forecast, despite calls for surveys to solicit distributional forecasts, see [Manski \(2004\)](#) for example, and the specific type of point forecast (mean, median, etc.) to be reported is generally not made explicit in the survey.

This paper proposes new methods to test the rationality of forecasts of some unknown measure of central tendency. Similar to [Elliott et al. \(2005\)](#), we propose a testing framework that nests the mean forecast as a special case, but unlike that paper we allow for alternative forecasts within a general class of measures of central tendency, rather than measures that represent other aspects of the predictive distribution (such as non-central quantiles or expectiles). We consider a class of central tendency measures that nests convex combinations of the mean, median and mode, and we overcome an inherent identification problem that arises using the methods of [Stock and Wright \(2000\)](#).

As a building block for the above test, we also present new tests for the rationality of mode forecasts. Some recent work (e.g., [Knüppel and SchulteFrankenfeld, 2012](#); [Reifschneider and Tulip,](#)

2019) suggests that point forecasts from central banks should be interpreted as mode rather than mean forecasts, and experimental studies (e.g., Peterson and Miller, 1964; Kröger and Pierrot, 2019b) report that participants are more likely to use the mode to summarize their predictive distribution than other measures of central tendency. While mode regression has received some attention in the recent literature (Kemp and Silva, 2012 and Kemp et al., 2020), tests for mode forecast rationality similar to those available for the mean and median (e.g., Mincer and Zarnowitz, 1969 and Gaglianone et al., 2011) are lacking. Direct analogs of existing tests are infeasible because the mode is not elicitable (Heinrich, 2014). We introduce the concept of *asymptotic elicitability* and show it applies to the mode by considering a generalized modal midpoint with asymptotically vanishing length. We then present results that allow for tests similar to the famous Mincer-Zarnowitz regression for mean forecasts.

We apply our tests to individual income expectation survey data collected by the Federal Reserve Bank of New York. We reject forecast rationality when interpreting responses as mean or median forecasts, however we cannot reject rationality when interpreting them as mode forecasts. We also find evidence of heterogeneity across respondents: for example, forecasts from younger, low-income respondents are not rationalizable using any measure of centrality, while forecasts from high-income respondents, regardless of their age, are rationalizable for many, though not all, measures of centrality. Further, we find that the behavior of high-income survey participants changes as they gain experience in the survey, consistent with the learning suggested in Kim and Binder (2023).

Previous analyses of economic surveys have typically assumed that responses are mean forecasts, and deviations from that benchmark have motivated learning and attention models which can account for such irrationality (e.g., Bordalo et al., 2020; Coibion and Gorodnichenko, 2015; Kohlhas and Walther, 2021). This paper proposes an intriguingly simple alternative to mean rationality, and one which is supported by our empirical work: point forecasts may not reflect the statistical average, but simply the most likely value, the mode.

A Proofs

Proof of Theorem 2.4. For the proof of statement (a), let $Y \sim P \in \mathcal{P}$ with density f , define $\tilde{K}_\delta(e) = \frac{1}{\delta}K\left(\frac{e}{\delta}\right)$ and introduce the notation $\bar{L}_\delta^K(x, P) = \mathbb{E}_{Y \sim P} [L_\delta^K(x, Y)]$. Then, it holds that

$$\bar{L}_\delta^K(x, P) = - \int \frac{1}{\delta}K\left(\frac{x-y}{\delta}\right) f(y) dy = - \int \tilde{K}_\delta(x-y) f(y) dy = -(f * \tilde{K}_\delta)(x),$$

where $f * \tilde{K}_\delta$ denotes the convolution of the functions f and \tilde{K}_δ . Ibragimov (1956) shows that for any log-concave density, its convolution with any other unimodal distribution function is again unimodal.³⁴ Thus, $\bar{L}_\delta^K(x, P)$ exhibits a unique minima which shows that Γ_δ^K is well-defined by (2.6).

We continue with statement (b) and show that $\Gamma_\delta(P) \rightarrow \text{Mode}(P)$ for all $P \in \mathcal{P}$. Notice that

$$\bar{L}_\delta^K(x, P) = - \int \frac{1}{\delta}K\left(\frac{x-y}{\delta}\right) f(y) dy = - \int f(x+u\delta)K(u) du = - \int f(x+u\delta) d\mathcal{K}(u), \quad (\text{A.1})$$

by applying integration by substitution and by interpreting the kernel $K(\cdot)$ as the density of the probability measure \mathcal{K} . It holds that

$$\sup_{x \in \mathbb{R}} |\bar{L}_\delta^K(x, P) - (-f(x))| = \sup_{x \in \mathbb{R}} \left| f(x) - \int f(x+u\delta) d\mathcal{K}(u) \right| \quad (\text{A.2})$$

$$\leq \sup_{x \in \mathbb{R}} \int |f(x) - f(x+u\delta)| d\mathcal{K}(u) \leq \sup_{x \in \mathbb{R}} \int |c\delta u| d\mathcal{K}(u) = c\delta \int |u|K(u) du \rightarrow 0 \quad (\text{A.3})$$

as $\delta \rightarrow 0$ as f is Lipschitz continuous with constant $c \geq 0$ and $\int |u|K(u) du < \infty$. Hence, $\bar{L}_\delta^K(x, P)$ converges uniformly (for all $x \in \mathbb{R}$) to $-f(x)$ as $\delta \rightarrow 0$ and it holds that $\arg \min_x \bar{L}_\delta^K(x, P) \rightarrow \arg \min_x (-f(x))$ as $\delta \rightarrow 0$. Consequently,

$$\lim_{\delta \rightarrow 0} \Gamma_\delta(P) = \lim_{\delta \rightarrow 0} \left(\arg \min_{x \in \mathbb{R}} \bar{L}_\delta^K(x, P) \right) = \arg \min_{x \in \mathbb{R}} \left(\lim_{\delta \rightarrow 0} \bar{L}_\delta^K(x, P) \right) = \arg \min_{x \in \mathbb{R}} (-f(x)),$$

which equals the mode for distributions with continuous Lebesgue density by definition.

For the proof of statement (c), we consider a fixed $\delta > 0$ and define $\bar{V}_\delta^K(x, P) := -\mathbb{E}_{Y \sim P} [V_\delta(x, Y)] =$

³⁴Ibragimov (1956) calls densities satisfying this property *strongly unimodal*. It is important to note that his notion of strong unimodality is different from ours introduced in Definition 2.1.

$\frac{1}{\delta^2} \int K' \left(\frac{x-y}{\delta} \right) f(y) dy$ and $\tilde{K}'_\delta(e) = \frac{1}{\delta^2} K' \left(\frac{e}{\delta} \right)$. Then,

$$\bar{V}_\delta^K(x, P) = \int \tilde{K}'_\delta(x-y) f(y) dy = (\tilde{K}'_\delta * f)(x) = (\tilde{K}_\delta * f')(x) = \int \tilde{K}_\delta(x-y) f'(y) dy \quad (\text{A.4})$$

$$= \int \tilde{K}_\delta(x + \text{Mode}(P) - y) f'(y - \text{Mode}(P)) dy \quad (\text{A.5})$$

$$= \int \tilde{K}_\delta(x + \text{Mode}(P) - y) g'(y) dy \quad (\text{A.6})$$

for some shifted density g with mode at zero. As the kernel \tilde{K}_δ is log-concave it has a monotone likelihood ratio (Proposition 2.3 (b) in [Saumard and Wellner, 2014](#)), i.e. for any $a \leq b$ and $y \geq 0$, it holds that $\frac{\tilde{K}_\delta(a-y)}{\tilde{K}_\delta(a)} \leq \frac{\tilde{K}_\delta(b-y)}{\tilde{K}_\delta(b)}$ and as $g'(y) \leq 0$ for $y \geq 0$, this implies that

$$\frac{\tilde{K}_\delta(a-y)}{\tilde{K}_\delta(b-y)} g'(y) \geq \frac{\tilde{K}_\delta(a)}{\tilde{K}_\delta(b)} g'(y). \quad (\text{A.7})$$

Analogously, the same inequality follows for any $a \leq b$ and $y \leq 0$, where the monotone likelihood ratio above holds with the reverse inequality but at the same time $g'(y) \geq 0$. Thus, for any $x > z$,

$$\bar{V}_\delta^K(x, P) = \int \tilde{K}_\delta(x + \text{Mode}(P) - y) g'(y) dy \quad (\text{A.8})$$

$$= \int \frac{\tilde{K}_\delta(x + \text{Mode}(P) - y)}{\tilde{K}_\delta(z + \text{Mode}(P) - y)} \tilde{K}_\delta(z + \text{Mode}(P) - y) g'(y) dy \quad (\text{A.9})$$

$$\geq \frac{\tilde{K}_\delta(x + \text{Mode}(P))}{\tilde{K}_\delta(z + \text{Mode}(P))} \int \tilde{K}_\delta(z + \text{Mode}(P) - y) g'(y) dy \quad (\text{A.10})$$

$$= \frac{\tilde{K}_\delta(x + \text{Mode}(P))}{\tilde{K}_\delta(z + \text{Mode}(P))} \bar{V}_\delta^K(z, P). \quad (\text{A.11})$$

For any root x^* such that $\bar{V}_\delta^K(x^*, P) = 0$, it follows that $\bar{V}_\delta^K(x_u, P) \geq 0$ for all $x_u > x^*$ and $\bar{V}_\delta^K(x_l, P) \leq 0$ for all $x_l < x^*$ as the kernel \tilde{K}_δ has infinite support. If there exist two distinct roots $x_l^* < x_u^*$ of $\bar{V}_\delta^K(\cdot, P)$, the above implies that $\bar{V}_\delta^K(x, P) = 0$ for all $x \in [x_l^*, x_u^*]$, which implies that the roots of $\bar{V}_\delta^K(\cdot, P)$ constitute a closed interval. As $\bar{L}_\delta^K(x, P)$ has a unique maximum for all $P \in \tilde{\mathcal{P}} \subset \mathcal{P}$ by part (a) and $\bar{V}_\delta^K(x, P) = -\frac{\partial}{\partial x} \bar{L}_\delta^K(x, P)$, it follows that $\bar{V}_\delta^K(x, P)$ must have a unique root, which implies that $V_\delta^K(x, Y)$ is a strict identification function for the generalized modal midpoint. \square

Proof of Theorem 2.6. We define

$$g_{t,T} := \delta_T^{3/2} T^{-1/2} \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) = -(T\delta_T)^{-1/2} K' \left(\frac{X_t - Y_{t+1}}{\delta_T} \right) \mathbf{h}_t, \quad (\text{A.12})$$

$$g_{t,T}^e := \mathbb{E}_t[g_{t,T}], \quad \text{and} \quad g_{t,T}^* := g_{t,T} - g_{t,T}^e, \quad (\text{A.13})$$

such that

$$\delta_T^{3/2} T^{-1/2} \sum_{t=1}^T \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) = \sum_{t=1}^T g_{t,T}^e + \sum_{t=1}^T g_{t,T}^*. \quad (\text{A.14})$$

Lemma S.1.1 shows that $\sum_{t=1}^T g_{t,T}^e \xrightarrow{P} 0$. Thus, it remains to show that $\sum_{t=1}^T g_{t,T}^* \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mode}})$. For some arbitrary, but fixed $\lambda \in \mathbb{R}^k$, $\|\lambda\|_2 = 1$, we define

$$z_{t,T} := \lambda^\top g_{t,T}^*, \quad \bar{\omega}_T^2 := \sum_{t=1}^T \text{Var}(z_{t,T}), \quad h_{t,T} := \frac{z_{t,T}}{\bar{\omega}_T}, \quad \text{and} \quad \omega^2 := \lambda^\top \Omega_{\text{Mode}} \lambda, \quad (\text{A.15})$$

and show that a univariate CLT for martingale difference arrays (MDA) holds for $\sum_{t=1}^T h_{t,T}$. It obviously holds that $(g_{t,T}^*, \mathcal{F}_{t+1})$ is a MDA as $g_{s,T}^* \in \mathcal{F}_{t+1}$ for all $s \leq t$ and $\mathbb{E}_t[g_{t,T}^*] = 0$ a.s. by definition. Thus, $(z_{t,T}, \mathcal{F}_{t+1})$ and $(h_{t,T}, \mathcal{F}_{t+1})$ are also MDAs.

In the following, we verify the following three conditions of Theorem 24.3 of Davidson (1994): (a) $\sum_{t=1}^T \text{Var}(h_{t,T}) = 1$, (b) $\sum_{t=1}^T h_{t,T}^2 \xrightarrow{P} 1$, and (c) $\max_{1 \leq t \leq T} |h_{t,T}| \xrightarrow{P} 0$. Lemma S.1.2 shows that $\bar{\omega}_T^2 = \sum_{t=1}^T \text{Var}[z_{t,T}] \rightarrow \lambda^\top \Omega_{\text{Mode}} \lambda = \omega^2$. Thus, as Ω_{Mode} is assumed to be positive definite, $\bar{\omega}_T^2$ is strictly positive for all T sufficiently large and hence, $h_{t,T}$ is well-defined and $\sum_{t=1}^T \text{Var}(h_{t,T}) = 1$, which shows condition (a). Lemma S.1.3 shows that $\sum_{t=1}^T z_{t,T}^2 \xrightarrow{P} \omega^2$ as $T \rightarrow \infty$ which implies condition (b), i.e. $\sum_{t=1}^T h_{t,T}^2 = \sum_{t=1}^T \frac{z_{t,T}^2}{\bar{\omega}_T^2} \xrightarrow{P} 1$. Eventually, Lemma S.1.4 shows condition (c) and we can apply Theorem 24.3 of Davidson (1994) in order to conclude that for all $\lambda \in \mathbb{R}^k$, $\|\lambda\|_2 = 1$, it holds that $\sum_{t=1}^T h_{t,T} \xrightarrow{d} \mathcal{N}(0, 1)$. As $\bar{\omega}_T^2 \rightarrow \omega^2$, Slutsky's theorem implies that $\sum_{t=1}^T z_{t,T} = \sum_{t=1}^T \lambda^\top g_{t,T}^* \xrightarrow{d} \mathcal{N}(0, \omega^2)$, and as this holds for all $\lambda \in \mathbb{R}^k$, $\|\lambda\|_2 = 1$, we apply the Cramér-Wold theorem and get that $\sum_{t=1}^T g_{t,T}^* \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mode}})$, which concludes the proof of this theorem. \square

Proof of Theorem 2.9. We let

$$g_{t,T} := \delta_T^{3/2} T^{-1/2} \psi(Y_{t+1}, X_t, \mathbf{h}_t, \delta_T) = -(T\delta_T)^{-1/2} K' \left(\frac{X_t - Y_{t+1}}{\delta_T} \right) \mathbf{h}_t,$$

$$g_{t,T}^e := \mathbb{E}[g_{t,T}], \quad \text{and} \quad g_{t,T}^* := g_{t,T} - g_{t,T}^e,$$

such that $\sum_{t=1}^T g_{t,T} = \sum_{t=1}^T g_{t,T}^e + \sum_{t=1}^T g_{t,T}^*$ as in the proof of Theorem 2.6. Then, it holds that $\sum_{t=1}^T g_{t,T}^* \xrightarrow{d} \mathcal{N}(0, \Omega_{\text{Mode}})$ as in the proof of Theorem 2.6. Notice for this that the employed Lemmas S.1.2–S.1.4 do not require the null hypothesis in (2.8).

We continue to analyze the term $\sum_{t=1}^T g_{t,T}^e$ as in the proof of Lemma S.1.1. However, the Taylor expansion in (S.1.13) now entails another term as we now impose $\mathbb{H}_{A,\text{loc}}$ in (2.13) instead of using $f'_t(0) = 0$ that resulted from \mathbb{H}_0 in (2.8). Hence, using that $\int K(u)du = 1$, we get

$$\sum_{t=1}^T g_{t,T}^e = T^{-1/2} \delta_T^{3/2} \sum_{t=1}^T f'_t(0) \mathbf{h}_t + o_P(1) = c \cdot \left(a_T T^{1/2} \delta_T^{3/2} + o_P(1) \right) + o_P(1) \quad (\text{A.16})$$

For statement (a), $a_T T^{1/2} \delta_T^{3/2} \xrightarrow{P} 1$, and thus, $\sum_{t=1}^T g_{t,T}^e = c + o_P(1)$. As Theorem 2.7 shows that $\widehat{\Omega}_{T,\text{Mode}} - \Omega_{T,\text{Mode}} \xrightarrow{P} 0$, applying Slutsky's theorem implies $\widehat{\Omega}_{T,\text{Mode}}^{-1/2} \sum_{t=1}^T g_{t,T} \xrightarrow{d} \mathcal{N}(\Omega_{\text{Mode}}^{-1/2} c, I_k)$. Hence, J_T converges to a non-central χ^2 -distribution,

$$J_T = \left(\widehat{\Omega}_{T,\text{Mode}}^{-1/2} \sum_{t=1}^T g_{t,T} \right)^\top \left(\widehat{\Omega}_{T,\text{Mode}}^{-1/2} \sum_{t=1}^T g_{t,T} \right) \xrightarrow{d} \chi_k^2(c^\top \Omega_{\text{Mode}}^{-1} c).$$

For statement (b), $a_T T^{1/2} \delta_T^{3/2} = o_P(1)$ and hence, $\sum_{t=1}^T g_{t,T}^e = o_P(1)$ putting us exactly in the situation of Theorem 2.6.

Finally, for statement (c), it holds that $(a_T T^{1/2} \delta_T^{3/2})^{-1} \xrightarrow{P} 0$, which implies that for any $\bar{c} \in \mathbb{R}$, $\mathbb{P} \left(\left| \sum_{t=1}^T g_{t,T}^e \right| \geq \bar{c} \right) \rightarrow 1$, and consequently also $\mathbb{P} \left(\left| \sum_{t=1}^T g_{t,T} \right| \geq \bar{c} \right) \rightarrow 1$. As furthermore $J_T = \left(\sum_{t=1}^T g_{t,T} \right)^\top \widehat{\Omega}_{T,\text{Mode}}^{-1} \left(\sum_{t=1}^T g_{t,T} \right)$ and $\widehat{\Omega}_{T,\text{Mode}} - \Omega_{T,\text{Mode}} \xrightarrow{P} 0$ by Theorem 2.7, where $\Omega_{T,\text{Mode}}$ is uniformly positive definite for T large enough, the conditions of Theorem 8.13 of White (1994) are satisfied and we can conclude that for any $\bar{c} \in \mathbb{R}$, $\mathbb{P}(|J_T| \geq \bar{c}) \rightarrow 1$, which concludes the proof of this theorem. \square

Proof of Theorem 3.3. For all fixed $\lambda \in \mathbb{R}^k$ such that $\|\lambda\|_2 = 1$, we define

$$\begin{aligned} \sigma_T^2 := \lambda^\top \Sigma_T(\theta_0) \lambda &= \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\theta_{10}^2 \left(\mathbf{h}_t^\top w_{\text{Mean}} \lambda \right)^2 \varepsilon_t^2 + \theta_{20}^2 \left(\mathbf{h}_t^\top w_{\text{Med}} \lambda \right)^2 \left(\mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}} \right)^2 \right. \\ &\quad + \theta_{30}^2 \left(\mathbf{h}_t^\top w_{\text{Mode}} \lambda \right)^2 f_t(0) \int K'(u)^2 du \\ &\quad \left. + 2\theta_{10}\theta_{20} \left(\mathbf{h}_t^\top w_{\text{Mean}} \lambda \right) \left(\mathbf{h}_t^\top w_{\text{Med}} \lambda \right) \varepsilon_t \left(\mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}} \right) \right]. \end{aligned} \quad (\text{A.17})$$

Lemma S.1.5 shows that $\sum_{t=1}^T \text{Var} \left(T^{-1/2} \phi_{t,T}^*(\theta_0) \lambda \right) - \sigma_T^2 \rightarrow 0$. As $\sigma^2 := \lambda^\top \Sigma(\theta_0) \lambda$, which is strictly positive by assumption, is the limit of σ_T^2 , the latter must be strictly positive for T large enough and consequently, $\sigma_T^{-1} T^{-1/2} \sum_{t=1}^T \phi_{t,T}^*(\theta_0) \lambda$ is well-defined. In the following, we first show that $\sigma_T^{-1} T^{-1/2} \sum_{t=1}^T \phi_{t,T}^*(\theta_0) \lambda \xrightarrow{d} \mathcal{N}(0, 1)$ by applying Theorem 24.3 in Davidson (1994) and by verifying that the respective conditions hold (with $X_{t,T} = \sigma_T^{-1} T^{-1/2} \phi_{t,T}^*(\theta_0) \lambda$).

Lemma S.1.6 shows that $T^{-1} \sum_{t=1}^T (\phi_{t,T}^*(\theta_0) \lambda)^2 - \sigma_T^2 \xrightarrow{P} 0$, which implies condition (a) of Theorem 24.3 of Davidson (1994), i.e. $\sigma_T^{-2} T^{-1} \sum_{t=1}^T (\phi_{t,T}^*(\theta_0) \lambda)^2 \xrightarrow{P} 1$. Lemma S.1.7 shows condition (b), i.e. $\max_{t=1, \dots, T} |\sigma_T^{-1} T^{-1/2} \phi_{t,T}^*(\theta_0) \lambda| \xrightarrow{P} 0$. Thus, we can apply Theorem 24.3 of Davidson (1994) and conclude that $\sigma_T^{-1} T^{-1/2} \sum_{t=1}^T \phi_{t,T}^*(\theta_0) \lambda \xrightarrow{d} \mathcal{N}(0, 1)$.

As σ_T^2 has the limit σ^2 , Slutsky's theorem implies that $T^{-1/2} \sum_{t=1}^T \phi_{t,T}^*(\theta_0) \lambda \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and as this holds for all $\lambda \in \mathbb{R}^k$ such that $\|\lambda\|_2 = 1$, we can conclude that $T^{-1/2} \sum_{t=1}^T \phi_{t,T}^*(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta_0))$. Furthermore, $\|T^{-1/2} \sum_{t=1}^T (\phi_{t,T}^*(\theta_0) - \phi_{t,T}(\theta_0))\| = \|T^{-1/2} \sum_{t=1}^T u_{t,T}(\theta_0)\| \xrightarrow{P} 0$ by Assumption 3.2 (D) and

$$T^{-1/2} \sum_{t=1}^T \left(\hat{\phi}_{t,T}(\theta_0) - \phi_{t,T}(\theta_0) \right) = T^{-1/2} \sum_{t=1}^T \theta_0 \cdot \begin{pmatrix} \mathbf{h}_t^\top (\hat{w}_{T,\text{Mean}} - w_{\text{Mean}}) \varepsilon_t \\ \mathbf{h}_t^\top (\hat{w}_{T,\text{Med}} - w_{\text{Med}}) \left(\mathbb{1}_{\{\varepsilon_t > 0\}} - \mathbb{1}_{\{\varepsilon_t < 0\}} \right) \\ \mathbf{h}_t^\top (\hat{w}_{T,\text{Mode}} - w_{\text{Mode}}) \delta_T^{-1/2} K' \left(\frac{-\varepsilon_t}{\delta_T} \right) \end{pmatrix} \xrightarrow{P} 0,$$

as it holds that $\hat{w}_{T,\text{Mean}} \xrightarrow{P} w_{\text{Mean}}$, $\hat{w}_{T,\text{Med}} \xrightarrow{P} w_{\text{Med}}$ and $\hat{w}_{T,\text{Mode}} \xrightarrow{P} w_{\text{Mode}}$ by assumption. Hence we can conclude that $T^{-1/2} \sum_{t=1}^T \hat{\phi}_{t,T}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta_0))$, which concludes the proof of this theorem. \square

References

- Abernethy, J. D. and Frongillo, R. M. (2012). A characterization of scoring rules for linear properties. In *Conference on Learning Theory*, pages 27–1. JMLR Workshop and Conference Proceedings.
- Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–89.
- Bank of England (2019). *Inflation Report, August 2019*. Available at <https://www.bankofengland.co.uk/inflation-report/2019/august-2019>.
- Bekaert, G. and Popov, A. (2019). On the link between the volatility and skewness of growth. *IMF Economic Review*, 67(4):746–790.
- Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5):725–733.
- Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.
- Bierens, H. B. (1982). Consistent model specification tests. *Journal of Econometrics*, 20:105–134.
- Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–82.
- 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 models. *Econometrica*, 75(5):1243–1284.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–62.
- Christoffersen, P. F. and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, 13(06):808–817.
- Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78.
- Corradi, V. and Swanson, N. R. (2002). A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics*, 110(2):353–381.
- D’Acunto, F., Hoang, D., Paloviita, M., and Weber, M. (2019). Cognitive abilities and inflation expectations. In *AEA Papers and Proceedings*, volume 109, pages 562–66.
- Das, M., Dominitz, J., and Van Soest, A. (1999). Comparing predictions and outcomes: Theory and application to income changes. *Journal of the American Statistical Association*, 94(445):75–85.
- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. OUP Oxford.
- Dearborn, K. and Frongillo, R. (2020). On the indirect elicibility of the mode and modal interval. *Annals of the Institute of Statistical Mathematics*, 72:1095–1108.
- Eddy, W. F. (1980). Optimum kernel estimators of the mode. *Annals of Statistics*, 8(4):870–882.
- Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, 72(4):1107–1125.
- Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–157.
- Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.
- 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.
- Eyting, M. and Schmidt, P. (2021). Belief elicitation with multiple point predictions. *European Economic Review*, 135.
- Fissler, T. and Hoga, Y. (2023). Backtesting systemic risk forecasts using multi-objective elicibility. *Journal of Business & Economic Statistics*. Forthcoming.
- Fissler, T. and Ziegel, J. F. (2016). Higher order elicibility and Osband’s principle. *Annals of Statistics*, 44(4):1680–1707.
- Foster, G. (1986). *Financial Statement Analysis, 2/e*. Pearson Education India.
- Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, 29(1):150–160.
- Garcia, J. A. and Manzanares, A. (2007). What can probability forecasts tell us about inflation risks? ECB Working Paper No. 825, Available at http://ssrn.com/abstract_id=1020964.
- Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.
- Givoly, D. and Hayn, C. (2000). The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become more conservative? *Journal of Accounting and Economics*, 29(3):287–320.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society*, 20:199–207.
- Gu, Z. and Wu, J. S. (2003). Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics*, 35(1):5–29.
- Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, 101(1):245–251.
- Heinrich-Mertsching, C. and Fissler, T. (2022). Is the mode elicitable relative to unimodal distributions? *Biometrika*, 109(4):1157–1164.
- Howard, R. C. C., Hardisty, D. J., Sussman, A. B., and Lukas, M. F. (2022). Understanding and neutralizing the expense prediction bias: The role of accessibility, typicality, and skewness. *Journal of Marketing Research*, 59(2):435–452.
- Ibragimov, I. A. (1956). On the composition of unimodal distributions. *Theory of Probability and Its Applications*, 1:255–260.
- Kahneman, D., Slovic, S. P., Slovic, P., and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kemp, G. C., Parente, P. M., and Silva, J. S. (2020). Dynamic vector mode regression. *Journal of Business & Economic Statistics*, 38(3):647–661.
- Kemp, G. C. and Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, 170(1):92 – 101.
- Kim, G. and Binder, C. (2023). Learning-through-survey in inflation expectations. *American Economic Journal: Macroeconomics*, 15(2):254–78.
- Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123.
- Knüppel, M. and Schultefrankenfeld, G. (2012). How informative are central bank assessments of macroeconomic risks? *International Journal of Central Banking*, 8:87–139.

- Kohlhas, A. N. and Walther, A. (2021). Asymmetric attention. *American Economic Review*, 111(9):2879–2925.
- Kröger, S. and Pierrot, T. (2019a). Comparison of different question formats eliciting point predictions. Working Paper, Available at <https://bibliothek.wzb.eu/pdf/2019/ii19-213.pdf>.
- Kröger, S. and Pierrot, T. (2019b). What point of a distribution summarises point predictions? Working Paper, Available at <https://bibliothek.wzb.eu/pdf/2019/ii19-212.pdf>.
- Li, Q. and Racine, J. S. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Madeira, C. (2018). Testing the rationality of expectations of qualitative outcomes. *Journal of Applied Econometrics*, 33(6):837–852.
- Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, 126(1):373–416.
- Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.
- McClave, J. T., Benson, P. G., and Sincich, T. (2017). *Statistics for Business and Economics, 13th Edition*. Pearson, New York.
- 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.
- Mincer, J. and Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pages 3–46. National Bureau of Economic Research, Inc.
- Mueller, A. I. and Spinnewijn, J. (2022). Expectations data, labor market and job search. *Handbook of Economic Expectations*.
- Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, 11(4):1833–1874.
- Patton, A. J. (2020). Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, 38(4):796–809.
- Patton, A. J. and Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480):1172–1184.
- Peterson, C. and Miller, A. (1964). Mode, median, and mean as optimal strategies. *Journal of Experimental Psychology*, 68(4):363–367.
- Reifschneider, D. and Tulip, P. (2019). Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve’s approach. *International Journal of Forecasting*, 35:1564–1582.
- Romer, C. D. and Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 90(3):429–457.
- Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4):1063–1119.
- Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8:45–114.
- Schmidt, P., Katzfuss, M., and Gneiting, T. (2021). Interpretation of point forecasts with unknown directive. *Journal of Applied Econometrics*, 36(6):728–743.
- Steinwart, I., Pasin, C., Williamson, R. C., and Zhang, S. (2014). Elicitation and identification of properties. *Journal of Machine Learning Research*, 35:1–45.
- Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5):1055–1096.

- Van Nieuwerburgh, S. and Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, 53(4):753–772.
- von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, 36(1):299–303.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1084.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press.
- White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, San Diego.
- Zhao, Y. (2022). Internal consistency of household inflation expectations: Point forecasts vs. density forecasts. *International Journal of Forecasting*.