Towson University
Department of Economics
Working Paper Series

Working Paper No. 2023-09

Uncertainty of Household Inflation Expectations: Reconciling Point and Density Forecasts

By Yongchen Zhao

December 2023

Uncertainty of Household Inflation Expectations: Reconciling Point and Density Forecasts

Yongchen Zhao

Towson University, Towson, MD 21252, USA

Dec, 2023

Abstract

We examine the uncertainty of household inflation expectations using matched point and density forecasts from the New York Fed's Survey of Consumer Expectations. We argue that using information from both types of forecasts allows for better estimates of uncertainty. Since the two types of forecasts may be inconsistent, we propose to reconcile them by matching the mean (or the median) of individual density forecasts and the corresponding point forecasts using exponential tilting. The reconciled densities provide uncertainty measures that are strictly consistent with the point forecasts by construction. We compare the uncertainty of inflation expectations derived from the reconciled densities with that derived from the original densities. Our results suggest that, at the micro-level, the uncertainty of consistent forecasts tends to be lower after reconciliation, while that of inconsistent forecasts tends to be higher. Aggregate uncertainty measured by averaging individual uncertainty is likely underestimated when using the survey responses directly, without reconciliation. This study contributes to the literature on the measurement of uncertainty and provides insights into the interplay of matched point and density forecasts in this context.

Keywords: Uncertainty measurement, Exponential tilting, Household survey, Consumer sentiment
JEL Codes: C53, E31, D12, C83, D84

Email address: `yzhao@towson.edu` (Yongchen Zhao)

1. Introduction

Accurate estimation of the uncertainty of household inflation expectations holds paramount significance in empirical research and policy formulation. Various methodologies exist for estimating forecast uncertainty, including those not contingent on the availability of density forecasts (e.g., Binder (2017), Rich and Tracy (2021), and Zhao (2022b)). Nevertheless, extracting measures of uncertainty from density forecasts remains the most direct and reliable approach preferred by researchers, especially when working with professional forecasts. However, since households face greater challenges in furnishing reliable density forecasts compared to professional forecasters, using densities alone may not always yield satisfactory results. At the same time, methods that rely exclusively on point forecasts, such as using disagreement as a proxy for uncertainty, are also potentially deficient in several important ways (see, for example, discussions and references in Rich and Tracy (2010), Manski (2018) and Topa (2019)).

In this paper, we consider an approach that uses both types of forecasts jointly. Our data on household expectations come from the New York Fed's Survey of Consumer Expectations (SCE), which provides individual-level matched point and density forecasts. Given the relative ease with which households formulate point forecasts compared to density forecasts, we argue that researchers can capitalize on the information from point forecasts to obtain better estimates of uncertainty. That is, we argue that the uncertainty measures derived from the densities are potentially more reliable if we "reconcile" the two types of forecasts.

The need for reconciliation arises from the potential inconsistency between point and density forecasts reported by the same survey respondent (e.g., Clements (2009, 2010)): Typical household surveys like the SCE lack any inherent mechanism to ensure that the point forecast equals the mean (or the median) of the matching density forecasts. Using the same non-parametric bounds analysis previously used to study professional forecasts (e.g., Engelberg et al. (2009) and Clements (2009)), Zhao (2022a) examined the consistency of the SCE forecasts in detail and reported that approximately 20% of the forecasts are inconsistent even when consistency is loosely defined.[1] The author also noted that, in over half of the inconsistent pairs of point and density forecasts, the point forecast exceeds the bounds by more than 4%. Additionally, when compared with inconsistent ones, consistent forecasts exhibit higher accuracy, lower levels of uncertainty, and are more likely to be

[1] The author considered several alternative definitions of consistency. The least stringent requires that the point forecast falls inside the upper and lower bounds on any of the three measures of central tendency of the matching density forecast: the mean, the median, or the mode.

reported by individuals with higher levels of education and financial literacy. Clearly, consistency is a desirable property of a pair of matching point and density forecasts.

To reconcile the inconsistent forecasts, we exponentially tilt the densities to match their mean or median to the point forecasts.[2] Exponential tilting is a standard method for imposing such restrictions, and it has been successfully applied in the context of economic forecasting. For example, Giacomini and Ragusa (2014) used this method to impose moment restrictions implied by an Euler equation on forecasts from Bayesian VAR models; Clements (2016) used exponential tilting to impose consistency between long-run survey forecasts and theory-based steady-state values of consumption, investment, and output growth; Galvão et al. (2021) used the method to improve statistical models' density forecasts using information from professional forecasters' densities.

Although we argue that the measures derived from the reconciled densities hold greater conceptual appeal, we do not seek to prove that they are more accurate than those from the original densities, as the true value of uncertainty is unobservable. Instead, in our empirical exercises, we focus on documenting and discussing their differences. Below, we introduce our data set and explain the reconciliation procedure, followed by results and additional remarks.

2. Data and Methodology

The SCE is a comprehensive survey designed to gauge the beliefs and expectations of U.S. consumers regarding various economic factors. The survey covers a broad range of topics, including inflation, labor market expectations, household finance, and access to credit. It also collects data on respondents' demographic characteristics (e.g., age, income, and education level) and provides measures of respondents' numeracy and financial literacy. We are primarily interested in information on inflation expectations, which is collected using the following questions:

> Question Q8v2part2: What do you expect the rate of inflation/deflation to be over the next 12 months?
> Question Q9b: In your view, what would you say is the percent chance that, over the next 12 months... the rate of inflation will be 12% or higher (bin 1); the rate of inflation will be between 8% and 12% (bin 2)...

The first question asks for a point forecast and the second question provides a forecast density in

[2]Note that we conduct separate exercises to match the mean and the median, and we do not require the mean and the median of a density to be the same.

the form of a histogram with 10 bins.[3] These two questions collect short-run expectations, i.e., over the next 12 months, while two additional similarly worded questions (Q9bv2part2 and Q9c) collect long-run expectations, i.e., over the 12-month period starting 24 months from now. We use the publicly available version of the survey, which does not include data from the most recent months. Our data set contains 149,412 observations from June 2013 to Jan 2023.

For the purpose of reconciliation, we require the point forecast to equal the mean or the median of the matching density forecast. Consider respondent i's responses to month t's survey.[4] Let y_{it} be the point forecast and $p_{it,j}$ be the probability assigned by the respondent to bin $j \in [1, 10]$ of the histogram. Let the left (lower) and right (upper) threshold of bin j be θ_{j+1} and θ_j, respectively, where $\theta_j > \theta_{j+1} \forall j, j+1$. Consistent with Armantier et al. (2017) and Zhao (2022a), we let $\theta_1 = 38\%$, $\theta_{11} = -38\%$, and we Winsorize the point forecasts so all of them fall into the interval $[\theta_{11}, \theta_1]$.[5] Let $h_{it}(y)$ be the forecast density obtained by assuming a uniform distribution in each bin of the histogram. Exponentially tilting $h_{it}(y)$ means to find a new density $f_{it}(y)$ so that its mean (or median) equals the point forecast y_{it}, while minimizing the difference between $h_{it}(y)$ and $f_{it}(y)$ measured by the Kullback-Leibler divergence $\int \log \frac{f_{it}(y)}{h_{it}(y)} f_{it}(y) dy$. We use the same numerical approximation procedure described in Giacomini and Ragusa (2014) when constructing the reconciled density.

For a pair of forecasts to be included in our sample, we require a well-defined $h_{it}(y)$, where $p_{it,j} \geq 0 \forall j$ and $\sum_{j=1}^{10} p_{it,j} = 1$. That is, all ten bins of a histogram must have non-zero probabilities assigned, and that the probabilities sum to unity. We also require a basic level of consistency, where, according to the forecast density, the point forecast occurs with a positive probability, i.e., $h_{it}(y_{it}) \geq 0$. This is ensured by the Winsorization step discussed above.[6] Finally, all our empirical exercises are conducted separately for individual and aggregate forecast densities,[7] separately for

[3]The 10 bins are: 12% or higher, 8% to 12%, 4% to 8%, 2% to 4%, 0% to 2%, -2% to 0%, -4% to -2%, -8% to -4%, -12% to -8%, and -12% or lower.

[4]To avoid complicating the notation, we do not distinguish the short-run forecasts from the long-run forecasts here. Our empirical exercises are carried out separately for different horizons.

[5]About 3% of the point forecasts are Winsorized. The narrower the interval, the more observations we must Winsorize, since we are unable to reconcile point forecasts that fall outside this range. On the other hand, the wider the interval, the higher the uncertainty (when bins 1 and/or 10 are assigned positive probabilities). We checked the sensitivity of our results to this choice by considering alternative values of +/-16%, +/-50%, and +/-100%. Our main conclusions remain unchanged. We caution against using intervals that are too narrow, especially in high-inflation environments.

[6]To confirm that our results are not driven by the Winsorization, we also considered a stricter version of this second requirement, namely, that the point forecast does not fall into the first or the last bin, i.e., $\theta_{10} \leq y_{it} \leq \theta_2$.

[7]The aggregate forecast density is simply the average of the individual ones, where the probability assigned to each bin j is $p_{t,j} = n_t^{-1} \sum_{i=1}^{n_t} p_{it,j}$, where n_t is the sample size for period t.

the short-run and the long-run expectations, and separately for the two reconciliation requirements that the mean or the median of $f_{it}(y)$ equals y_{it}.

We use the IQR of the individual forecast densities to measure individual-level uncertainty. Consistent with this choice, we use the IQR of the aggregate forecast density to measure aggregate uncertainty.[8] Although our conclusions remain unchanged when we use the standard deviations instead of the IQRs, we prefer the latter given how the alternative is more sensitive to the choice of the width of the first and last bin, and that the New York Fed publishes on the survey website uncertainty measures based on IQR.[9] These individual and aggregate uncertainty measures are reported and discussed in the next section.[10]

3. Results and Discussions

Figure 1 compares the aggregate uncertainty measures derived from the original densities and the reconciled densities for short-run expectations. The figure shows that, using the original densities without reconciliation would underestimate uncertainty (by about 8.6% overall).[11] There are also potentially important differences in the dynamics of the series. For example, the measure based on the original densities exaggerated the jump in uncertainty during the 2020 recession and subsequently understated the impact of rapidly increasing inflation during most of 2021. The measure based on the reconciled densities shows that aggregate uncertainty peaked in Oct 2021 when the CPI inflation rate exceeded 6% for the first time in decades. This is some five months before the peak of the measure based on the original densities. Results on individual-level uncertainty are reported in Figure 2, in which we plot the median of the individual IQRs along with the 10th, the 25th, the 75th, and the 90th percentiles. The top plot shows these statistics computed using the original densities, and the lower plot shows the results from the reconciled densities. The changes

[8]Using uncertainty measures based on standard deviations instead of IQRs do not change our conclusions: Individual-level uncertainty is measured using the standard deviation of the individual forecast density $h_{it}(y)$, denoted by $\sigma_{it,h}$. Aggregate uncertainty is measured using $(n_t^{-1}\sum_{i=1}^{n_t}\sigma_{it,h}^2)^{1/2}$. This measure of aggregate uncertainty is based on the decomposition of the standard deviation of the aggregate forecast density $\sigma_{t,h} \equiv (n_t^{-1}\sum_{i=1}^{n_t}\sigma_{it,h}^2 + d_t^2)^{1/2}$, where d_t is the forecast disagreement. See Zhao (2022a) for additional discussions about these measures.

[9]See https://www.newyorkfed.org/microeconomics/sce#/influncert-1. Note that their measure of aggregate uncertainty is the median of the individual IQRs, which are not based directly on the survey responses but are from generalized beta distributions fitted to the reported histograms.

[10]For brevity, we report only the results from the reconciliation of the short-run expectations, where we require the mean of the reconciled density to equal the corresponding point forecast. Omitted results are available upon request.

[11]Results for the long-run expectations are similar, with an overall level of underestimation of 6.4%.

to the 75th and the 90th percentiles are the most notable. Compared with the series based on the original densities, the series based on the reconciled densities evolve over time much more similarly to the median and the aggregate uncertainty measures shown in Figure 1.

The reconciliation naturally results in a larger change to the densities with means/medians further away from the corresponding point forecasts. To better understand the implications of reconciliation, we separately examine the subset of consistent and inconsistent forecasts. Figure 3 compares the average of individual-level uncertainty (i.e., IQR) before and after reconciliation, along with the percentage of consistent and inconsistent forecasts. Since the figure is based on the forecast densities that are tilted to match their means with the corresponding point forecasts, consistency in this context is defined as the point forecast lying within the non-parametric bounds for the mean, i.e., $y_{it} \in [\sum_{j=1}^{10} \theta_{j+1} p_{it,j}, \sum_{j=1}^{10} \theta_j p_{it,j}]$.[12] The figure shows that, on average, the forecast densities that are consistent with the corresponding point forecasts (depicted in the upper plot) tend to exhibit lower uncertainty after reconciliation, while the opposite is true for the inconsistent densities (depicted in the lower plot). Consequently, if one were to measure aggregate uncertainty using the median of individual-level IQRs, reconciliation would lead to reduced levels of aggregate uncertainty, as long as most individual forecast densities are consistent.[13] Figure 3 also shows that, even the inconsistent densities seem to provide valuable information about uncertainty (as opposed to pure noise) – in the lower plot, average uncertainty increased notably from 2021 throughout the first half of 2022 in a way similar to that in the upper plot.

Finally, we compare the mean and the three quartiles of individual-level uncertainty for various demographic groups before and after reconciliation. The results, which are mostly as expected, are reported in Table 1. The same differences across demographic groups are present in both the original and the reconciled densities. Comparing the means before and after reconciliation, we observe that the procedure resulted in an increase in the vast majority of the cases. For the medians, the opposite is observed. Comparing various demographic groups, we find that survey respondents with higher levels of education, household income, and numeracy tend to have lower uncertainty. White and Asian Americans tend to have lower uncertainty compared to other minority groups. No significant difference is present across groups defined by age or region of residence.

[12] When the reconciliation is based on the median of the forecast density, consistency is defined as the point forecast lying within the bounds for the median, i.e., $y_{it} \in [\theta_{s+1}, \theta_s]$, where $\sum_{j=1}^{s-1} p_{it,j} \leq 0.5$ and $\sum_{j=1}^{s} p_{it,j} > 0.5$.

[13] This is indeed what we observe. It is also one reason we preferred the average of individual-level IQRs. Using the average instead of the median allows the resulting measure of aggregate uncertainty to be more responsive, with potential issues caused by extreme outliers mitigated by the reconciliation procedure.

4. Concluding Remarks

As already stated, we can neither prove nor disprove the claim that the reconciliation procedure provides more accurate measurement of uncertainty. Nevertheless, we find much appeal in reconciliation. First, it produces strictly consistent point and density forecasts at both the individual and aggregate levels. This eliminates the need to choose one or the other when constructing uncertainty measures. Second, the procedure provides a natural and intuitive way to mitigate the effects of outliers by limiting the support of the forecast densities and Winsorizing the point forecasts correspondingly. Furthermore, we believe that joint use of information from matched point and density forecasts allows us to construct uncertainty measures that better capture various signals of uncertainty, such as the spread of point forecasts (i.e., forecast disagreement), the shape of reported forecast densities, and the degree of consistency between point and density forecasts.[14] Therefore, we advocate the use of the reconciliation procedure or its derivative, such as constructing aggregate uncertainty measures using a selected subset of reconciled individual densities, where the selection may be based on the degree of consistency, thus avoiding reconciling point and density forecasts that are too different. Additional research on this topic may prove fruitful.

References

Armantier, O., Topa, G., van der Klaauw, W., Zafar, B., 2017. An overview of the survey of consumer expectations. Economic Policy Review , 51–72.

Binder, C.C., 2017. Measuring uncertainty based on rounding: New method and application to inflation expectations. Journal of Monetary Economics 90, 1–12.

Clements, M.P., 2009. Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. The methodology and practice of econometrics. A festschrift in honour of David F. Hendry , 206–226.

Clements, M.P., 2010. Explanations of the inconsistencies in survey respondents' forecasts. European Economic Review 54, 536–549.

[14]Further research is needed to better understand this last source of uncertainty signals. Figure 3 shows that the degree of consistency (e.g., percent of consistent responses) decreased significantly in early 2020 and late 2022, while our intuition suggests that inflation uncertainty may be elevated during both time periods.

Clements, M.P., 2016. Long-run restrictions and survey forecasts of output, consumption and investment. International Journal of Forecasting 32, 614–628.

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

Galvão, A.B., Garratt, A., Mitchell, J., 2021. Does judgment improve macroeconomic density forecasts? International Journal of Forecasting 37, 1247–1260.

Giacomini, R., Ragusa, G., 2014. Theory-coherent forecasting. Journal of Econometrics 182, 145–155.

Manski, C.F., 2018. Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise. NBER Macroeconomics Annual 32, 411–471.

Rich, R., Tracy, J., 2010. The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts. The Review of Economics and Statistics 92, 200–207.

Rich, R., Tracy, J., 2021. A Closer Look at the Behavior of Uncertainty and Disagreement: Micro Evidence from the Euro Area. Journal of Money, Credit and Banking 53, 233–253.

Topa, G., 2019. Chapter 10 - Consumer expectations and the macroeconomy, in: Haughwout, A., Mandel, B. (Eds.), Handbook of US Consumer Economics. Academic Press, pp. 257–272.

Zhao, Y., 2022a. Internal consistency of household inflation expectations: Point forecasts vs. density forecasts. International Journal of Forecasting , forthcoming.

Zhao, Y., 2022b. Uncertainty and disagreement of inflation expectations: Evidence from household-level qualitative survey responses. Journal of Forecasting 41, 810–828.

Table 1: Mean and quartiles of individual-level uncertainty by demographic groups

This table compares the mean, the 25th percentile, the 50th percentile, and the 75th percentiles of individual-level uncertainty (IQR) before and after reconciliation, separately for each demographic group. The results are based on the short-run expectations.

	Original densities				Reconciled densities			
Group	Mean	25th	50th	75th	Mean	25th	50th	75th
Age								
40 to 60	4.9	1.8	2.8	5.6	5.4	1.8	2.7	5.3
Over 60	4.8	1.7	2.7	5.3	5.2	1.7	2.6	5.0
Under 40	4.8	2.0	3.1	5.4	5.4	2.0	2.9	5.3
Education								
College or Above	3.9	1.7	2.7	4.3	4.2	1.7	2.5	4.0
High School or Below	7.2	2.1	4.4	11.1	8.3	2.1	4.1	11.7
Some College	5.7	1.9	3.2	7.4	6.4	1.9	3.1	7.2
Income								
50k to 100k	4.5	1.8	2.8	5.0	5.0	1.8	2.7	4.9
Over 100k	3.6	1.7	2.5	4.0	3.8	1.7	2.5	3.7
Under 50k	6.3	2.1	3.7	8.7	7.1	2.1	3.5	8.9
Numeracy								
High	4.0	1.8	2.7	4.4	4.3	1.8	2.6	4.2
Low	7.1	2.1	4.2	11.2	8.3	2.1	4.0	11.7
Race								
American Indian/Alaska Native	6.9	4.8	6.4	8.5	7.6	5.2	6.7	9.2
Asian	4.1	1.8	2.6	4.2	4.5	1.8	2.5	4.1
Black/African American	7.8	2.2	4.9	11.8	9.0	2.1	4.7	13.7
Hispanic/Latino/Spanish	6.0	1.9	3.3	8.1	7.0	2.0	3.2	8.3
Hawaiian/Pacific Islander	6.4	5.9	6.4	6.9	6.2	5.7	6.1	6.7
Other	5.4	2.0	3.3	6.7	5.8	1.9	3.2	6.4
White	4.5	1.8	2.8	4.9	4.9	1.8	2.7	4.7
Region								
Midwest	4.7	1.9	2.8	5.2	5.1	1.9	2.7	5.0
Northeast	4.5	1.7	2.7	4.8	5.0	1.7	2.6	4.6
South	5.3	1.9	3.0	6.3	5.9	1.9	2.9	6.1
West	4.6	1.8	2.9	5.0	5.0	1.8	2.7	4.8

Figure 1: Comparing aggregate uncertainty measures

This figure compares the aggregate uncertainty (IQR of the aggregate forecast density) derived from the original densities and the reconciled densities. Results are based on short-run expectations. Shaded areas correspond to NBER recessions.

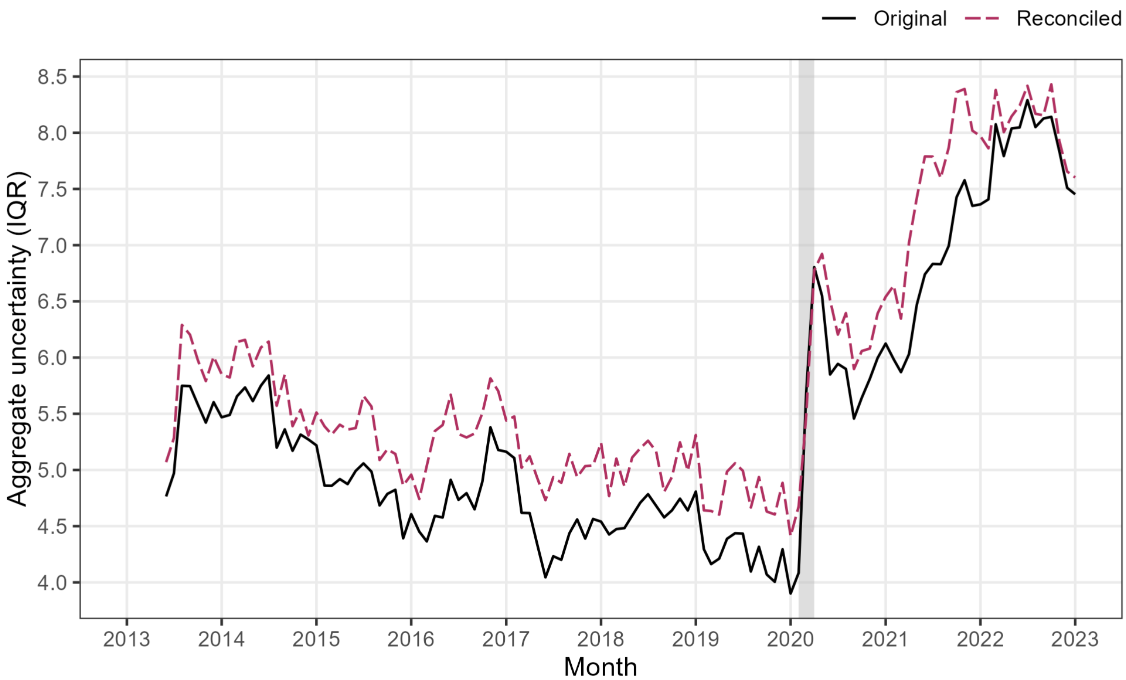

Figure 2: Selected quantiles of individual-level uncertainty measures

This figure shows the median of the individual IQRs, along with the 10th, the 25th, the 75th, and the 90th percentiles. The top plot depicts the original densities, and the bottom plot depicts the reconciled densities. Results are based on short-run expectations. Shaded areas correspond to NBER recessions.

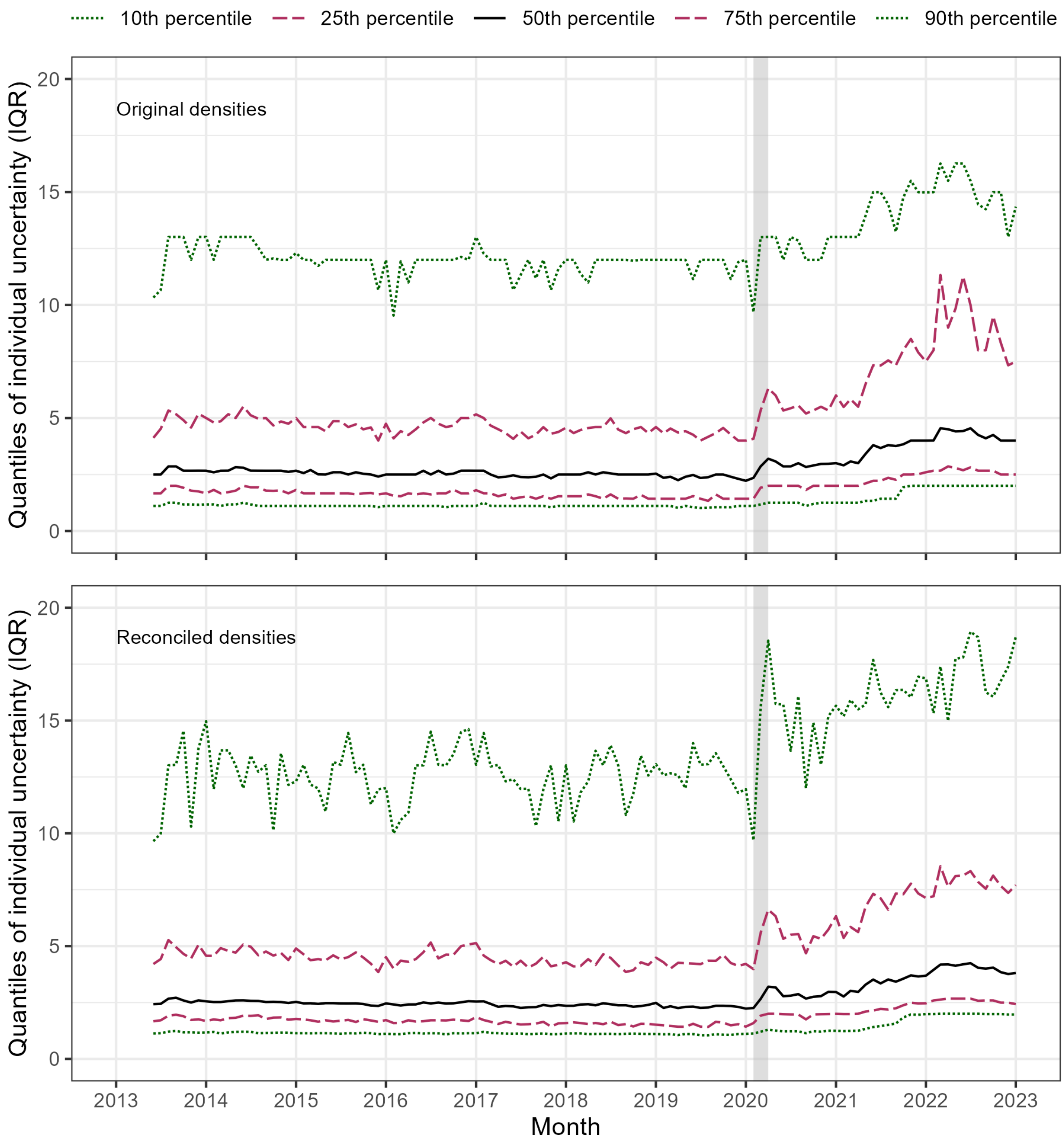

Figure 3: Average of individual-level uncertainty measures: consistent vs. inconsistent forecasts

This figure compares the average of individual-level uncertainty derived from the original and the reconciled densities. The top plot shows the results obtained using the consistent forecasts, and the bottom plot shows the results obtained using the inconsistent forecasts. In each plot, "sample size (%)" is the percentage of each month's observations that are consistent/inconsistent. Consistency is defined as the point forecast lying within the non-parametric bounds for the mean. Results are based on short-run expectations. Shaded areas correspond to NBER recessions.