Online Appendixes to Anchoring Inflation Expectations in Unconventional Times: Micro Evidence for the Euro Area

Jonas Dovern[a] and Geoff Kenny[b]
[a]FAU Erlangen-Nürnberg and CESifo
[b]European Central Bank

Appendix A. SPF Density Forecasts and Other Data Sources

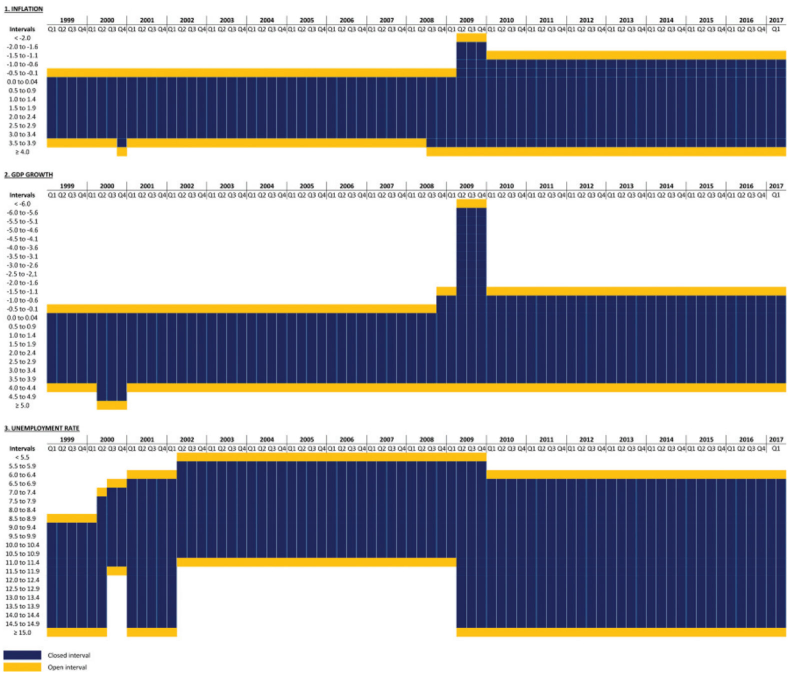
A.1 Survey Data

The Survey of Professional Forecasters (SPF) is a quarterly survey of macroeconomic expectations, conducted by the European Central Bank (ECB). The SPF panel of point and probability forecasts for euro-area gross domestic product (GDP), HICP inflation, and the unemployment rate at long, medium, and short horizons were obtained from the ECB website at http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html.

The survey asks for expectations about inflation, real GDP growth, and the unemployment rate in the euro area. The group of institutions which takes part in the survey consists of large commercial banks, insurance companies, other financial and nonfinancial firms, and research institutes from the European Union. Panelists are identified by unique (though anonymous) IDs that allow tracking of panelists over time. The SPF was launched in 1999:Q1. One part of the survey provides information about density forecasts which constitute a quantitative assessment of the subjective uncertainty surrounding the point forecasts of panelists.

The survey of the density forecasts is more complex than a survey of point forecasts only. In fact, the SPF provides only discrete approximations to the subjective predictive densities of the forecasters. Forecasters are asked to assign probabilities to a range of intervals (frequently called "bins") that indicate how likely they believe it is that the future realization of a variable

Figure A.1. Range of Intervals in the SPF over Time



Notes: A probability assigned to an open interval indicates the expected probability that the variable's outcome will be larger (or smaller) than the lower (upper) interval boundary.

falls into a certain interval. A sample questionnaire showing how this is done in practice can be found using the following link: https://www.ecb.europa.eu/stats/pdf/spfquestionnaire.pdf.

The range of intervals used in the SPF has changed across time for all three variables considered in the survey. Figure A.1 provides an overview about the changes. Most of the adjustments were made in response to the Great Recession after 2008. In addition, there was some more modest adjustment of the appropriate range of intervals during the early phase of the survey, especially in relation to the unemployment rate. In contrast, the width of the inner intervals (0.5 percentage point) has remained constant across time for all variables. Note that the outer intervals at the extremities of the surveyed ranges are "open intervals," i.e., upper/lower bounds are not defined.

A.2 ECB Balance Sheet Data

Monthly data were obtained from Datastream. We include the most recent change of the total volume of the ECB's assets and liabilities over a period of three months that was known to the forecasters at the time they submitted their forecasts to the SPF.

A.3 Inflation Rate

We use inflation rates based on the Harmonised Index of Consumer Prices (HICP). Monthly data were obtained from Datastream. In the regressions, we use the most recent change of the annual inflation rate over a period of three months that was known to the forecasters at the time they submitted their forecasts to the SPF.

Appendix B. Robustness with Respect to Approach for Estimating Moments of Predictive Densities

To analyze how our results are affected by the choice of approach for estimating the moments of the individual predictive densities, this appendix compares our baseline results with those obtained from alternative parametric and nonparametric approaches. We start by looking at the raw correlations between different moment estimates, followed by a comparison of break test results discussed in section 3 of the paper. We then turn to the possible impact on the panel regression results discussed in section 4 of the paper.

The alternative approaches that we consider are the following: First, instead of assuming that the probability mass is condensed at the midpoint of each bin (_d), we assume that it is uniformly distributed within each bin (_u). Second, we fit a normal distribution to the probabilities attached to the discrete bins (_n) by minimizing the squared deviations between the fitted distribution and the empirically observed values of the cumulative distribution function (CDF); note that this restricts the skewness and the excess kurtosis to 0 by construction. Finally, we fit a beta distribution to the

Table B.1. Correlations between Different Moment Estimates

Mean					Inflation Uncertainty				
	_d	_u	_n	_b		_d	_u	_n	_b
_d	1.00				_d	1.00			
_u	1.00	1.00			_u	1.00	1.00		
_n	1.00	1.00	1.00		_n	0.98	0.98	1.00	
_b	1.85	0.85	0.86	1.00	_b	0.99	0.99	1.00	1.00
Skewness					**Kurtosis**				
	_d	_u	_n	_b		_d	_u	_n	_b
_d	1.00				_d	1.00			
_u	0.97	1.00			_u	0.94	1.00		
_n	—	—	—		_n	—	—	—	
_b	0.60	0.69	—	1.00	_b	0.57	0.68	—	1.00

Notes: Correlations are computed using the full sample of individual long-term inflation expectations. The four methods used to estimate the moments of the individual predictive densities are described in the text above.

probabilities attached to the discrete bins (_b), again by minimizing the squared deviations between the fitted distribution and the empirically observed values of the CDF.[1]

Table B.1 shows the raw correlations between the different moment estimates for the first four moments of the individual predictive densities. For the mean, the estimate using the beta function has a correlation of roughly 0.85 with the other methods which among themselves are perfectly correlated (in the case of _d and _u by construction). For the standard deviation, we find correlations of close to 1 for all pairwise correlations including the beta function. For the skewness and the kurtosis, also both discrete approximations (_d and _u) are highly correlated while the correlations with the estimates produced by the beta function tend to be lower (between 0.57 and 0.69). These lower correlations suggest that a key possible source of measurement error may link to the assumption of a discrete versus

[1] Following Engelberg, Manski, and Williams (2009), we fit a triangular distribution in those cases in which a panelist attaches positive probability mass to only two of the bins.

a continuous distribution function. We have opted for the discrete approach because it is in line with the discrete nature of the survey design (i.e., the discrete bins used in the survey questionnaire) and does not attribute a specific continuous functional form to the respondent's subjective density. Nonetheless, should a respondent's true subjective density be a continuous function like the beta distribution, the analysis in table B.1 points to a potentially important role for measurement error in influencing, in particular, the higher (third and fourth) moments. To investigate further this issue, we consider below whether the assumption of a beta function changes the main conclusions of our breakpoint analysis and the reduced-form panel regression analysis.

Next, we look at the results of the breakpoint tests discussed in section 3 for the different moment estimates. Table B.2 lists the breaks that we detect for the different types of moment estimates obtained by the Bai-Perron test (LWZ statistic) and the Andrews-Ploberger (AP) test, respectively. The results show that, with some exceptions, the conclusions regarding breaks in the moments are very often the same no matter which type of moment estimate is chosen. In particular, across most methods considered, we observe a downward break in the mean, an upward break in the standard deviation, and a break indicating more negative skewness. The exceptions are the following: Based on the beta distribution approach, no break in mean expectations is detected. This result highlights again that the result of a downward break in mean long-term expectations is less robust to the choice of continuous versus discrete distributions. This is also very much in line with our view and conclusions that the break in mean expectations should not be overemphasized given its small size. Interestingly, however, the stability tests on the third and fourth moments using the beta function (for which the correlations in table B.1 were even lower than the mean) point to a high level of robustness compared with all of the discrete approximations. For example, the estimate using the beta distribution also points to a break in skewness in 2010:Q2 (the same data as our baseline _d method). Similarly, the break tests on kurtosis using the beta distribution also suggest no change in tail risk (kurtosis) over our sample, in line with the main results highlighted in the paper. At the same time, while the broad conclusions presented in the main paper tend to be supported, the analysis presented in table B.2 highlights also

Table B.2. Robustness of Breakpoint Tests

	Bai-Perron Test	**Andrews-Ploberger Test**
	Mean	
_d	2013:Q3	2013:Q3
_u	2013:Q3	2013:Q3
_n	2013:Q3	2013:Q3
_b	—	—
	Inflation Uncertainty	
_d	2009:Q3	2009:Q3
_u	2009:Q3	2009:Q3
_n	2009:Q3	2009:Q3
_b	2009:Q3	2009:Q3
	Skewness	
_d	2010:Q2	2010:Q1
_u	—	2010:Q1
_n	Constant by Construction	Constant by Construction
_b	2010:Q2	2010:Q2
	Kurtosis	
_d	—	2006:Q4
_u	—	—
_n	Constant by Construction	Constant by Construction
_b	—	—

Notes: Breaks according to the Bai-Perron test are based on the modified Schwarz criterion (LWZ); breaks according to the Andrews-Ploberger test are determined using a 5 percent significance level. The four methods used to estimate the moments of the individual predictive densities are described in the text above.

some uncertainty as regards the precise timing and occurrence of the breaks in different moments that may be linked to measurement error. For example, while the Bai-Perron test does not find any break in the skewness based on the approach that distributed the probability mass uniformly within bins, the Andrews-Ploberger test does locate the break in the skewness slightly earlier in 2010:Q1 for two methods and in 2010:Q2 for one method. The Andrews-Ploberger test identifies a break in the kurtosis based on our baseline estimates, but not based on the other estimates. Given that this is found for only one of the eight tests presented and the earlier evidence that measurement error may disproportionately affect higher moment estimates, we would not overweight this finding.

Overall, across the four moment estimates considered, the analysis presented in table B.2 tends to support the results presented in the main text and conclusions. In particular, it suggests (i) at most only a small break in the mean around the time of the sovereign debt crisis, (ii) much clearer and robust evidence of a break in long-term inflation uncertainty in the more immediate wake of the Great Recession, (iii) robust evidence of a shift toward a more negatively skewed distribution in the first half of 2010, and (iv) no robust evidence of a change in perceived long-term tail risks (kurtosis).

Finally, we assess whether the regression results from section 4 depend on the moment estimation method. To do this, we run all regressions using the different moment estimates (applying the same estimation approaches also to the predictive densities for GDP growth and the unemployment rate). Table B.3 replicates our baseline results for mean expectations (columns 1 and 5) and adds estimates for the other three moment estimates (columns 2–4 and 6–8). Overall, there are very few substantial differences across moment estimation methods.[2] In particular, as emphasized in the main body of the paper, the results on the importance of short-term inflation expectations and the central bank performance measure for the updating of mean expectations as well as the relative stability of their estimated coefficients are confirmed by all four estimation methods. The negative sign on the central bank balance sheet linked to the first part of the sample is also observed for all four methods. Based on the regressions in table B.3, there are only two changes compared with the results presented in the main body of the paper. First, the coefficient corresponding to the change in the observed inflation rate is not significantly different from 0 in the full-sample

[2] Note that the results for mean expectations are equal by construction for the first two moment estimation methods.

Table B.3. Robustness of Regression for Mean Expectations

	Without Break				With Break in 2007:Q4			
	_d	_u	_n	_b	_d	_u	_n	_b
	Full-Sample Coefficients				Pre-2007:Q4 Coefficients			
$dE_i[\pi(1y)]_t$	0.011	0.011	0.012	0.021	0.006	0.006	0.012	0.023
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)
$dE_i[\pi(2y)]_t$	0.177***	0.177***	0.166***	0.162***	0.205***	0.205***	0.198***	0.211***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)
$(\mathrm{MA}(\pi) - E_i[\pi(5y)])_{t-1}$	0.155***	0.155***	0.151***	0.176***	0.187***	0.187***	0.181***	0.215***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
$\pi - E_i[\pi(1y)]_{t-4}$	−0.006	−0.004	−0.004	0.005	0.003	0.003	0.003	−0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
$\pi - E_i[\pi(2y)]_{t-8}$	0.007	0.006	0.006	−0.004	0.017	0.017	0.010	−0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
$dE_i[\mathrm{GDP}(5y)]_t$	0.007	0.007	0.015	0.039	0.019	0.019	0.033	0.047
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$dE_i[\mathrm{U}(5y)]_t$	−0.013	−0.013	−0.000	0.001	0.000	0.000	−0.000	0.007
	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)	(0.02)	(0.00)	(0.01)
$d\pi_{t-1}$	0.031**	0.030**	0.032**	0.024	0.032	0.032	0.039	0.018
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
dCBBS_{t-1}	−0.160**	−0.164**	−0.161**	−0.150**	−0.882***	−0.882***	−0.911**	−0.978***
	(0.06)	(0.06)	(0.06)	(0.07)	(0.34)	(0.34)	(0.36)	(0.35)
MPA Dummy	0.018	0.019	0.022	0.023				
	(0.02)	(0.02)	(0.02)	(0.02)				
ELB Dummy	0.011	0.011	0.017*	0.005				
	(0.01)	(0.01)	(0.01)	(0.02)				
Constant	−0.000	−0.000	−0.000	0.000	−0.004	−0.004	−0.005	−0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

(continued)

Table B.3. (Continued)

	Without Break				With Break in 2007:Q4			
	_d	_u	_n	_b	_d	_u	_n	_b
					Change of Coefficients after 2007:Q4			
$dE_i[\pi(1y)]_t$					0.010	0.010	0.002	0.001
					(0.04)	(0.04)	(0.04)	(0.04)
$dE_i[\pi(2y)]_t$					−0.052	−0.052	−0.056	−0.093*
					(0.06)	(0.06)	(0.06)	(0.06)
$(\text{MA}(\pi) - E_i[\pi(5y)])_{t-1}$					−0.026	−0.026	−0.028	−0.054
					(0.05)	(0.05)	(0.05)	(0.05)
$\pi - E_i[\pi(1y)]_{t-4}$					−0.010	−0.010	−0.009	0.003
					(0.03)	(0.03)	(0.02)	(0.03)
$\pi - E_i[\pi(2y)]_{t-8}$					−0.013	−0.013	−0.005	0.002
					(0.03)	(0.03)	(0.03)	(0.03)
$dE_i[\text{GDP}(5y)]_t$					−0.017	−0.017	−0.035	−0.036
					(0.06)	(0.06)	(0.06)	(0.05)
$dE_i[\text{U}(5y)]_t$					−0.019	−0.019	0.000	−0.014
					(0.02)	(0.02)	(0.00)	(0.02)
$d\pi_{t-1}$					0.001	0.001	−0.004	0.012
					(0.03)	(0.03)	(0.03)	(0.04)
dCBBS_{t-1}					0.764**	0.764**	0.788**	0.868**
					(0.34)	(0.34)	(0.37)	(0.36)
MPA Dummy					0.016	0.016	0.021	0.019
					(0.02)	(0.02)	(0.02)	(0.02)
ELB Dummy					0.012	0.012	0.018*	0.003
					(0.01)	(0.01)	(0.01)	(0.01)
Observations	1,180	1,180	1,165	1,165	1,180	1,180	1,165	1,165
R^2	0.195	0.195	0.187	0.182	0.204	0.204	0.195	0.193

Notes: Dependent variable is the change in long-term inflation expectations. All models include fixed effects for each forecaster. The constant is identified by restricting the average of the fixed effects to equal 0. We report the within R^2. Standard errors are computed using the method of Driscoll and Kraay (1998) and are robust against general forms of spatial and temporal dependence. *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively. The different methods for estimating the mean expectations (_d, _u, _n, and _b) are described at the beginning of this appendix.

regression based on mean expectations computed using the assumption of a beta distribution, while it is in the three other cases. Second, the effective lower bound (ELB) dummy is significant in the regression based on mean expectations computed using the assumption of a normal distribution (p-value 0.063), while it is not in the three other cases. The estimated effect is, however, found to be positive, implying that the hitting of the ELB generated an upward revision in mean expectations. Given that this effect is not found for the other three estimates, and also given that the assumption of a Gaussian distribution appears the least easy to justify in the context of the SPF, we have not emphasized this result in the main paper.

Table B.4 replicates our baseline results for inflation uncertainty (columns 1 and 5) and adds estimates for the other three moment estimates (columns 2–4 and 6–8). Again, the reported results are very much in line with the main findings reported in the paper. In particular, across all four methods presented, we observe an effect of short-term inflation volatility, long-term GDP volatility, and a downward impact from the absolute change in the central bank's balance sheet volume. Also, in line with the results reported in the main body of the paper, we observe for each of the moment estimates an upward impact effect of nonstandard monetary policy announcements on long-term inflation uncertainty. There are, however, also some notable cases of a change in significance levels compared with the baseline moment estimate approach. In particular, the coefficient corresponding to the uncertainty about long-term unemployment is not significantly different from 0 in the case of inflation uncertainty based on the normal or beta distribution assumption, while it is for the other two methods. Overall, however, the results reported in table B.4 suggest that our main results and conclusions appear to be reasonably robust: In particular, most of the significant co-movement appears similar across different moment estimation methods.

Table B.4. Robustness of Regression for Inflation Uncertainty

	Without Break				**With Break in 2007:Q4**			
	_d	**_u**	**_n**	**_b**	**_d**	**_u**	**_n**	**_b**
	Full-Sample Coefficients				**Pre-2007:Q4 Coefficients**			
$dV_i[\pi(1y)]_t$	0.067	0.073	0.048	0.066	−0.012	−0.001	−0.063	−0.015
	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.10)	(0.10)
$dV_i[\pi(2y)]_t$	0.265***	0.261***	0.281***	0.297***	0.417***	0.418***	0.401***	0.468***
	(0.05)	(0.04)	(0.06)	(0.05)	(0.08)	(0.08)	(0.11)	(0.10)
$\lvert(\text{MA}(\pi) - E_i[\pi(5y)])_{t-1}\rvert$	0.005	0.006	0.003	0.020	−0.014	−0.011	−0.008	0.020
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)
$\lvert\pi - E_i[\pi(1y)]_{t-4}\rvert$	0.003	0.004	0.005	0.004	0.004	0.009	0.006	0.005
	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
$\lvert\pi - E_i[\pi(2y)]_{t-8}\rvert$	−0.003	−0.003	−0.003	−0.003	−0.005	−0.009*	−0.008	−0.008
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
$dV_i[GDP(5y)]_t$	0.250***	0.241***	0.277***	0.269***	0.282***	0.267***	0.286***	0.262***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)	(0.03)	(0.03)
$dV_i[U(5y)]_t$	0.101***	0.104***	−0.000	0.001	0.074**	0.078**	−0.000***	−0.003
	(0.02)	(0.02)	(0.00)	(0.01)	(0.03)	(0.03)	(0.00)	(0.01)
$\lvert d\pi_{t-1}\rvert$	0.021***	0.020***	0.027***	0.023***	0.032**	0.031**	0.062***	0.047***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
$\lvert\text{dCBBS}_{t-1}\rvert$	−0.082*	−0.080*	−0.087*	−0.073*	−0.323**	−0.347***	−0.447***	−0.293**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.16)	(0.13)	(0.17)	(0.14)
MPA Dummy	0.020***	0.019***	0.019***	0.021***				
	(0.01)	(0.01)	(0.01)	(0.01)				
ELB Dummy	−0.004	−0.003	−0.008	0.002				
	(0.01)	(0.01)	(0.01)	(0.01)				
Constant	−0.000	0.000	−0.001	−0.001	−0.003	−0.003	−0.004	−0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

(continued)

Table B.4. (Continued)

	Without Break				With Break in 2007:Q4			
	_d	_u	_n	_b	_d	_u	_n	_b
					Change of Coefficients after 2007:Q4			
$dV_i[\pi(1y)]_t$					0.134	0.125	0.185	0.147
					(0.09)	(0.10)	(0.12)	(0.11)
$dV_i[\pi(2y)]_t$					-0.231^{**}	-0.234^{**}	-0.198^{*}	-0.272^{***}
					(0.09)	(0.09)	(0.12)	(0.10)
$\lvert(\mathrm{MA}(\pi) - E_i[\pi(5y)])_{t-1}\rvert$					0.026	0.024	0.016	−0.005
					(0.02)	(0.02)	(0.03)	(0.03)
$\lvert\pi - E_i[\pi(1y)]_{t-4}\rvert$					0.001	−0.005	−0.001	−0.000
					(0.02)	(0.01)	(0.01)	(0.01)
$\lvert\pi - E_i[\pi(2y)]_{t-8}\rvert$					0.004	0.010	0.009	0.008
					(0.02)	(0.01)	(0.01)	(0.01)
$dV_i[\mathrm{GDP}(5y)]_t$					−0.077	−0.062	−0.029	−0.021
					(0.07)	(0.07)	(0.06)	(0.06)
$dV_i[\mathrm{U}(5y)]_t$					0.048	0.044	0.000	0.030^{*}
					(0.04)	(0.04)	(0.00)	(0.02)
$\lvert d\pi_{t-1}\rvert$					−0.015	−0.014	-0.046^{**}	-0.033^{**}
					(0.01)	(0.02)	(0.02)	(0.02)
$\lvert\mathrm{dCBBS}_{t-1}\rvert$					0.296^{*}	0.317^{**}	0.432^{**}	0.271^{*}
					(0.17)	(0.14)	(0.18)	(0.15)
MPA Dummy					0.021^{***}	0.020^{***}	0.019^{**}	0.021^{***}
					(0.01)	(0.01)	(0.01)	(0.01)
ELB Dummy					−0.001	−0.000	−0.004	0.004
					(0.01)	(0.01)	(0.01)	(0.01)
Observations	1,180	1,180	1,165	1,165	1,180	1,180	1,165	1,165
R^2	0.388	0.389	0.307	0.328	0.402	0.403	0.320	0.346

Notes: Dependent variable is the change in long-term inflation uncertainty. All models include fixed effects for each forecaster. The constant is identified by restricting the average of the fixed effects to equal 0. We report the within R^2. Standard errors are computed using the method of Driscoll and Kraay (1998) and are robust against general forms of spatial and temporal dependence. *, **, and *** denote significance at the 10 percent, 5 percent, and 1 percent level, respectively. The different methods for estimating the mean expectations (_d, _u, _n, and _b) are described at the beginning of this appendix.

Appendix C. Assessment of Effects of Measurement Error in the Moment Estimations

In this appendix, we present the results of two simulation experiments aimed at assessing the impact of measurement error on our results.

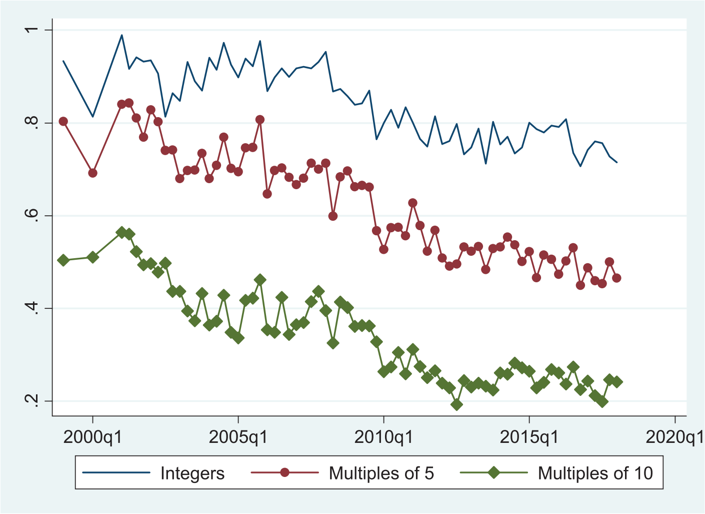
We first assess how much measurement error can be expected due to our baseline method for estimating the moments of the predictive densities. To do so, we assume a simulation environment with 30 individual forecasters, each of whose true predictive density is given by a beta distribution defined over the support [–2; 4.5] to approximate the ranges used in the actual SPF survey rounds. For each individual forecaster the parameters a and b of the assumed beta distribution are drawn randomly from a normal distribution with an assumed mean of 2 and a standard deviation of 0.5.[3] Furthermore, we assume that the moments of this distribution have to be estimated (using our baseline approach) given information similar to information provided in the SPF responses. In particular, we assume that we observe the probabilities for 13 intervals with fixed length of 0.5.

In reality, there is some degree of rounding in the survey responses. Manski and Molinari (2010) find strong evidence for rounding in probabilistic expectations in the Health and Retirement Study. For short- and medium-term forecasts from both the European SPF and the U.S. counterpart, Glas and Hartmann (2017) find that typically three-quarters of forecasters round *most of the stated probabilities* to multiples of 5 percent. For our sample of long-term inflation density forecasts, we find that about 81.5 percent of reported interval probabilities are integers, 63 percent are multiples of 5, and 36.5 percent are multiples of 10. We also note that there is a tendency during recent years to use less rounding (figure C.1), which suggests that potential bias induced by rounding behavior has diminished recently. The slow decline in rounding behavior suggests that this is a very low frequency phenomenon which is unlikely to influence any higher-frequency results.

Against this background, we allow for rounding in our simulations. Specifically, we consider simulation variants in which we round

[3] In line with empirical observation, this calibration allows for a relatively high degree of heterogeneity in observed density characteristics.

Figure C.1. Share of Rounded Interval Probabilities over Time



Notes: The lines show for each survey wave the fraction of all (nonzero) interval probabilities that are integers, multiples of 5, and multiples of 10, respectively.

all interval probabilities to either integers, multiples of 5, or multiples of 10, respectively.

The simulation contains the following steps:

1. For each forecaster draw a and b from a normal distribution with mean 2 and a standard deviation of 0.5.

2. Calculate the true first four moments of the beta distribution for each forecaster and also the interval probabilities implied by the distribution.

3. Allow for rounding of the interval probabilities at the individual level. We consider four cases:[4]

[4] In the cases 3.b., 3.c., and 3.d. we apply the following rounding algorithm. We start by looking at the outer intervals and round the contained probabilities downward. The difference in probability mass is shifted to the next-most inner interval. Then we do the same for the next intervals, and so on, until we reach the middle interval.

a. No rounding

b. Rounding to integer values

c. Rounding to multiples of 5

d. Rounding to multiples of 10

4. Use the interval probabilities to estimate the first four moments of the individual distributions by assuming that the probability mass of each interval is located at its midpoint.

5. Compute the average estimated and true moments across the sample of forecasters.

6. Repeat steps 1–5 1,000 times, saving all true and estimated moments.

7. Use the 1,000 sets of moments to calculate the root mean squared error (RMSE) and bias for each of the first four moments.

The results of this simulation exercise are reported in table C.1. In interpreting the results of the simulation, it is important to distinguish between the case where we assume no rounding behavior at the individual level and cases where we allow for rounding. In particular, when we do not allow for rounding, we can isolate only the effects of the different estimation strategies on the measurement of the moments.

The upper panel shows that the RMSEs are very small when we do not allow for rounding. Not surprisingly, it is the fourth moment measuring the fluctuations in tail risks that is most affected by measurement error. For the other three moments, the approximation error is very small (relative to the assumed true moments) and always below 0.01 under the no-rounding scenario. In the lower panel, we observe that the approximation errors for the first and third moment are not linked to bias, while bias plays a role in the case of the second and fourth moment. In the no-rounding scenario our baseline method, by construction, overestimates these two moments. Overall, this is very much in line with the results presented earlier

Table C.1. Approximation Errors for Baseline Approach

		Root Mean Squared Error			
	Avg. True Moments	**No Rounding**	**Round to 1**	**Round to 5**	**Round to 10**
Mean	1.250	0.001	0.001	0.005	0.011
SD	1.453	0.005	0.026	0.138	0.265
Skewness	0.000	0.002	0.007	0.027	0.048
Kurtosis	−0.857	0.029	0.067	0.175	0.243
		Bias			
	Avg. True Moments	**No Rounding**	**Round to 1**	**Round to 5**	**Round to 10**
Mean	1.250	0.000	0.000	0.000	0.000
SD	1.453	0.005	−0.026	−0.138	−0.265
Skewness	0.000	0.000	0.000	0.000	0.000
Kurtosis	−0.857	0.015	−0.034	−0.136	−0.207

Notes: Simulations are based on 1,000 iterations. This set of simulations assume that average moments are calculated as the mean of moments from N = 30 different panelists. For each of those panelists we draw the beta distribution like in the original simulation and also compute the discrete approximation and its moments in the same way.

in appendix B highlighting only a limited impact of the different moment estimation strategies on the baseline results as presented in the main body of the paper.

The simulation analysis under the different rounding scenarios highlights a potentially more important role for approximation error in effecting the moments. However, even in the case in which we round to multiples of 5, the RMSE for the mean (0.027) and for the standard deviation (0.142) remain relatively small compared with the size of the true moments of 1.25 and 1.453, respectively. The table also reveals how the approximation errors for the second and fourth moment are most affected by rounding behavior. Our assumption about how rounding is implemented causes the estimates to systematically underestimate the true variance and kurtosis. Given that rounding behavior is observed in the SPF in practice, these results highlight the need for caution when interpreting the estimated "level" of the standard deviation and the kurtosis.[5] However, as nearly all of the approximation error for the second and fourth moments is due to rounding bias, the change in these moments will be less affected because rounding behavior itself is quite persistent. As a result, we would expect the bias to be relatively constant over time and, therefore, the dynamics of the moments over time to be only moderately affected. This can partly explain why in appendix B our regression and breakpoint test results are relatively stable.

In a second simulation, we analyze how the size and power of the breakpoint tests we employ might be affected by the presence of measurement error. We simulate sequences of length 100 of densities using the approach from the first simulation above. To investigate the size of the tests, we keep the parameters of the setting that we use to draw the densities for each period constant for the entire sample. To analyze the power of the tests, we assume that there is a single structural break at the 50th observation. We consider the following breaks:

- *A shift in mean.* We simulate this by shifting the lower bound and the upper bound of the support by 0.5, holding the number and length of the intervals constant.
- *A small shift in all moments.* We simulate this by changing the hyperdistribution for parameter b from $N(2.0,0.5)$ before the break to $N(1.9,0.5)$ after the break.
- *A shift in the variance.* We simulate this by increasing the support by one unit. This is done by adding two intervals of the same size in both tails.

Table C.2 contains the results. The numbers shown indicate the frequency by which the null hypothesis of no structural break is rejected for the time series of estimated moments. The top panel shows that, even though there is measurement error, the tests seem to have the correct size of 10 percent. The three lower panels show that the tests also have reasonable power to detect breaks in those

[5] In this context, the presented simulation results should be interpreted as some kind of upper bound because our assumptions about how rounding is implemented are deliberately chosen to be very extreme.

Table C.2. Size and Power of Breakpoint Tests in the Presence of Measurement Error

Size (No Break)					
		No Rounding	**Round to 1**	**Round to 5**	**Round to 10**
BP	Mean	0.071	0.072	0.074	0.076
	Variance	0.075	0.064	0.066	0.099
	Skewness	0.070	0.072	0.090	0.091
	Kurtosis	0.074	0.062	0.088	0.088
AP	Mean	0.010	0.096	0.098	0.099
	Variance	0.092	0.089	0.088	0.118
	Skewness	0.099	0.099	0.112	0.096
	Kurtosis	0.097	0.086	0.098	0.093
Shift in Mean (at t = 50)*					
		No Rounding	**Round to 1**	**Round to 5**	**Round to 10**
BP	Mean	1.000	1.000	1.000	1.000
	Variance	0.082	0.072	0.095	0.087
	Skewness	0.072	0.068	0.094	0.089
	Kurtosis	0.078	0.068	0.094	0.084
AP	Mean	1.000	1.000	1.000	1.000
	Variance	0.098	0.095	0.077	0.080
	Skewness	0.082	0.094	0.090	0.104
	Kurtosis	0.091	0.094	0.073	0.090
Small Shift in Moments (at t = 50)**					
		No Rounding	**Round to 1**	**Round to 5**	**Round to 10**
BP	Mean	0.513	0.537	0.503	0.480
	Variance	0.503	0.440	0.227	0.118
	Skewness	0.516	0.530	0.330	0.159
	Kurtosis	0.349	0.249	0.119	0.083
AP	Mean	0.637	0.660	0.658	0.590
	Variance	0.627	0.597	0.324	0.156
	Skewness	0.637	0.652	0.458	0.229
	Kurtosis	0.459	0.332	0.166	0.095

(*continued*)

Table C.2. (Continued)

Shift in Variance (at t = 50)* **					
		No Rounding	**Round to 1**	**Round to 5**	**Round to 10**
BP	Mean	0.092	0.086	0.104	0.086
	Variance	1.000	1.000	1.000	1.000
	Skewness	0.093	0.088	0.087	0.146
	Kurtosis	0.089	0.369	0.421	0.192
AP	Mean	0.118	0.099	0.117	0.084
	Variance	1.000	1.000	1.000	1.000
	Skewness	0.116	0.105	0.113	0.113
	Kurtosis	0.117	0.490	0.522	0.215

Notes: All simulations are made with T = 100, N = 30, and using 500 replications. The numbers shown are rejection frequencies. BP results based on modified Schwarz criterion. AP results based on a nominal size of 0.1. *Shift of mean from 1.25 to 1.75. **Shift of mean from 1.25 to 1.33, standard deviation from 1.45 to 1.48, skewness from 0 to −0.04, and kurtosis from −0.86 to −0.87. ***Shift of standard deviation from 1.45 to 1.68.

moments that are affected by the changes that we model in the three scenarios. The breaks in mean and variance generated by changing the support of the distribution are easily detected by both tests even in the presence of measurement error and heavy rounding. In the case of an increase in variance, rounding leads both tests to spuriously detect a break in kurtosis quite frequently. This is because the increased number of bins causes a change in the average probability mass allocated to the outer bins. Consequently, the likelihood that, for instance, the probability assigned to the outer bins is rounded to 0 changes.[6] This result indicates that it is very likely that the very modest evidence of a break in kurtosis that we find (AP test based on our baseline method) is driven by changes in rounding behavior and should not be interpreted as evidence for a genuine change of the true forecast distribution. Also, the power of both tests to

[6] Note that the rejection frequency for the kurtosis is highest for moderate rounding because here the difference is largest while under heavy rounding the probabilities assigned to the outer bins are rounded to 0 in most cases before *and* after the break.

detect the more subtle changes in the second break scenario is reasonably high for all four moments (0.349 to 0.637) with measurement error and no rounding. Once we add rounding, the power for detecting the small break in the mean stays relatively constant, while it deteriorates substantially for the other moments—especially for the variance and the kurtosis. This comes as no surprise because, as seen above, the measurement errors corresponding to these moments are most heavily affected by rounding.

Yet overall, both simulations reported in this appendix suggest that the results emphasized in the main body of the paper are very unlikely to be driven solely by the moment estimation step and its underlying assumptions.

References

Driscoll, J., and A. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (4): 549–60.

Engelberg, J., C. F. Manski, and J. Williams. 2009. "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters." *Journal of Business and Economic Statistics* 27 (1): 30–41.

Glas, A., and M. Hartmann. 2018. "Overconfidence versus Rounding in Survey-based Density Forecasts." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3202810.

Manski, C. F., and F. Molinari. 2010. "Rounding Probabilistic Expectations in Surveys." *Journal of Business and Economic Statistics* 28 (2): 219–31.