Journal of Forecasting, *J. Forecast.* **34**, 405–426 (2015)
Published online 25 March 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2340

Measuring Disagreement in Qualitative Expectations

FRIEDER MOKINSKI,[1] XUGUANG (SIMON) SHENG,[2]* AND JINGYUN YANG[3]

[1] *Deutsche Bundesbank, Frankfurt, Germany*

[2] *Department of Economics, American University, Washington, DC, USA*

[3] *Rush Alzheimer's Disease Center & Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA*

ABSTRACT

We assess how well measures of disagreement in qualitative survey expectations reflect disagreement in corresponding quantitative expectations. We consider a variety of measures, belonging to two categories: measures of dispersion in nominal and ordinal variables and measures based on the probability approach of Carlson and Parkin (*Economica*, 1975; **42**, 123–138). Using data from two household surveys that collect both qualitative and quantitative inflation expectations, we find that the probability approaches with time-varying categorization thresholds and either a piecewise uniform or t distribution perform best and the resulting disagreement estimates are highly correlated with the benchmark. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS consumer survey; inflation forecast; qualitative data; quantification method

INTRODUCTION

Forecast disagreement matters. Recent work on macroeconomics has emphasized the role of disagreement in signaling upcoming structural and temporal changes in the economy (Mankiw and Reis, 2002). Similarly, disagreement as a proxy for uncertainty is closely related to business cycle fluctuations (Lahiri and Sheng, 2010). Empirical measures of disagreement can be extracted from either point or qualitative forecasts. Most studies that consider point forecasts rely on measures of cross-sectional heterogeneity such as the standard deviation or the interquartile range. For qualitative forecasts, which are widespread among surveys of both consumers and businesses, however, there are many competing approaches and a lack of consensus among researchers about the appropriate measurement concept.[1] While some studies have used quantification approaches (Dasgupta and Lahiri, 1993; Mankiw *et al.*, 2004), others have employed measures for variation in ordinal variables (Ehrmann *et al.*, 2012; Bachmann *et al.*, 2013) or nominal variables (Maag, 2009; Lamla and Maag, 2012). Given the important role of expectations in the decision making of firms and consumers, a careful examination of the various approaches to measuring disagreement in qualitative expectations cannot be overemphasized.

Drawing on relevant studies from economics, statistics, sociology and psychology, we provide a detailed review of the literature on measuring disagreement in qualitative survey data. The review is primarily concerned with estimating disagreement in three-category qualitative responses. This focus reflects that in most surveys the percentages of respondents who expect a variable to 'increase', 'stay the same' or 'decrease' are the only aggregate statistics available. More specifically, we consider several variants of the probability approach (Carlson and Parkin, 1975). These variants differ in their assumptions about the distribution of the unobserved point forecasts (normal/t/piecewise uniform) and the conversion from quantitative to qualitative forecasts (constant or time-varying thresholds). Moreover, we consider a set of statistical measures of dispersion in nominal or ordinal variables. Earlier work by Batchelor (1986) and Maag (2009) has studied a subset of these disagreement measures. We extend their research by providing a comprehensive and critical survey of measures of disagreement in qualitative expectations.

To assess the performance of various disagreement measures, we employ data from two household surveys that collect both qualitative and quantitative responses: the University of Michigan Survey of Consumers and the Swedish Consumer Tendency Survey. We choose the Michigan survey since household expectations in this survey are often used in the macroeconomics literature (e.g. Carroll, 2003; Ang *et al.*, 2007; Dräger and Lamla, 2012). Through the use of the Swedish survey, we can consistently compare our results with those of Maag (2009). By taking advantage of these two datasets, we are able to directly evaluate the different measures of disagreement in qualitative expectations against a standard benchmark measure of disagreement in point forecasts (cross-sectional standard deviation). We find that (i) statistical measures of nominal or ordinal variation display only weak correlations with the benchmark,

* Correspondence to: Xuguang (SIMON) Sheng, Department of Economics, American University, Washington, DC, USA. E-mail: sheng@american.edu

[1] Curtin (2007) lists 45 countries that conducted qualitative household surveys as of 2007. There are many recent additions in Eastern Europe and Asia.

Copyright © 2015 John Wiley & Sons, Ltd

(ii) the classical probability approach produces a measure that is often moderately correlated with the benchmark, (iii) variants of the probability approach that use non-normal distributions—such as the piecewise uniform distribution proposed in this paper or the t distribution—typically perform better, and (iv) allowing for time-varying categorization thresholds in the probability approach leads to further improvement.

The paper proceeds as follows. In the next section we review the existing methods for measuring disagreement in qualitative expectations and propose alternative approaches. In the third section we empirically assess the performance of various disagreement measures using two household survey datasets that elicit both qualitative and quantitative responses. The fourth section concludes. The Appendix presents additional robustness checks of the main findings.

MEASURING DISAGREEMENT IN QUALITATIVE SURVEY DATA

In this section, we outline the commonly used approaches to measuring disagreement in qualitative survey data. We classify these approaches into two groups: measures of nominal or ordinal variation and probability approaches. Measures of nominal or ordinal variation apply concepts of cross-sectional dispersion to qualitative forecasts. In contrast, the probability approach aims to estimate a measure of cross-sectional variation for the latent distribution of point forecasts.

Measures of nominal or ordinal variation

Below, we briefly summarize five nominal or ordinal variation approaches to estimating disagreement in qualitative expectations based on the raw aggregate share of responses.

The first measure is an index of qualitative variation (IQV), suggested by Gibbs and Poston (1975, p. 472):

$$\mathrm{IQV} = \frac{K}{K-1}\left(1 - \sum_{i=1}^{K} s_i^2\right) \tag{1}$$

where K is the number of categories in the survey and s_i is the percentage of responses in category i. The scaling factor $\frac{K}{K-1}$ ensures that $0 \leq \mathrm{IQV} \leq 1$, with the upper limit being reached when the responses are uniformly distributed across the categories. As a measure of disagreement in qualitative responses, the IQV has been criticized by Blair and Lacy (2000) for wasting information by treating ordinal responses as nominal. However, in Maag (2009)'s recent empirical applications the IQV performs well with five-category qualitative response data.

The remaining alternatives explicitly acknowledge that the qualitative survey responses are measured on an ordinal scale. The first two are based on the same concentration (i.e. absence of dispersion) metric l^2:

$$l^2 = \frac{\sum_{i=1}^{K-1}\left(F_i - \frac{1}{2}\right)^2}{(K-1)/4} \tag{2}$$

where F_i is the cumulative response shares in category i, e.g. $F_2 = s_1 + s_2$. Blair and Lacy's (2000) measure of ordinal variation is $1-l^2$, and Kvalseth's (1995) coefficient of ordinal variation (COV) is $1-l$. These two alternatives measure the distance of an observed distribution from the point of maximum concentration by different metrics. Nonetheless, both of them range between 0 and 1, obtaining their minimum when all responses fall into a single category and their maximum when the distribution of responses is polarized, i.e. $s_1 = s_3 = 0.5$. These measures have, for example, been employed by Ehrmann *et al.* (2012) to measure dispersion in qualitative inflation forecasts from the EC's consumer surveys.[2]

Another alternative is Reardon's (2009) entropy-based measure of ordinal variation:

$$v = \frac{1}{K-1}\sum_{i=1}^{K-1}\left(F_i \log_2 \frac{1}{F_i} + (1-F_i)\log_2 \frac{1}{1-F_i}\right) \tag{3}$$

Similar to COV and $1-l^2$, the v statistic ranges between 0 and 1, obtaining its minimum when all the responses fall into a single category and its maximum when the distribution of the responses is polarized.

Bachmann *et al.* (2013) propose another measure of disagreement in qualitative survey data. Specifically, they code the response 'go up' as 1, 'stay the same' as 0 and 'go down' as -1. Their (unweighted) disagreement measure is defined as the standard deviation of the coded responses:

$$\mathrm{BES} = \sqrt{U_t + D_t - (U_t - D_t)^2} \tag{4}$$

[2] Ehrmann *et al.* (2012) find that enhanced central bank transparency does not affect disagreement among the general public. There is also a large literature that examines the disagreement among professional forecasters; see Lahiri and Sheng (2008), Dovern *et al.* (2012) and Hubert (2014), among others.

Copyright © 2015 John Wiley & Sons, Ltd

where U_t and D_t are the observed shares of 'up' and 'down' responses, respectively. BES ranges from 0 to 1, reaching its minimum when all responses fall into a single category and its maximum when $U = D = \frac{1}{2}$. Note that BES is equivalent to the square root of the disconformity coefficient of Theil (1955).

To sum up, among the five approaches IQV measures the variability in nominal variables and ignores the ordered nature of survey responses. By contrast, BL, COV and Reardon measure the dispersion in ordered data by applying different distance metrics. By coding each category with a cardinal number, BES computes the dispersion as the standard deviation of the coded responses, which conceptually resembles the measure of disagreement for point forecasts.

Probability approach

The probability approach aims at estimating the shape of the cross-sectional distribution of the unobserved point forecasts from observed qualitative responses. Most previous studies have focused on the mean of the unobserved distribution, as evidenced by the two comprehensive surveys of Nardo (2003) and Pesaran and Weale (2006). We distinguish from these studies by looking at forecast disagreement, measured as the cross-sectional standard deviation of the unobserved point forecasts. Theil (1955) presented the first formalization of the probability approach, which has been reinvented by Carlson and Parkin (1975) and named the Carlson–Parkin approach (CP hereafter) in the literature.

The probability approach assumes that survey respondents convert unobserved point forecasts to observed qualitative expectations by a deterministic categorization scheme. If the point forecast f_{it} of respondent i at time t is larger than the threshold $\tau_{\mathrm{up},t}$, the respondent will report 'go up'; if f_{it} is between $\tau_{\mathrm{down},t}$ and $\tau_{\mathrm{up},t}$, the respondent will report 'stay the same'; if f_{it} is below $\tau_{\mathrm{down},t}$, the respondent will report 'go down'.[3] Variants of the probability approach differ in their assumptions about the distribution of the unobserved point forecasts (normal/t/piecewise uniform) and the conversion from quantitative to qualitative forecasts (constant or time-varying thresholds).

Constant thresholds

Among the variants of the probability approach, the classical CP approach is arguably the most popular one. It assumes that (i) the unobserved point forecasts $\{\mathrm{f}_{it}\}_{(i=1,\ldots,N_t)}$ in period t are independent and identically distributed normal with mean μ_t and standard deviation σ_t, and (ii) the thresholds are constant and symmetric around zero, i.e. $\tau_{\mathrm{up},t} = -\tau_{\mathrm{down},t} = \tau$. Under these assumptions, Carlson and Parkin (1975) obtain

$$\mu_t = \tau\left(\Phi^{-1}(p_{d,t}) + \Phi^{-1}(1-p_{u,t})\right) / \left(\Phi^{-1}(p_{d,t}) - \Phi^{-1}(1-p_{u,t})\right) \tag{5}$$

$$\sigma_t = 2\tau / \left(\Phi^{-1}(1-p_{u,t}) - \Phi^{-1}(p_{d,t})\right) \tag{6}$$

where Φ is the cumulative distribution function (cdf) of a standard normal random variable, and $p_{u,t}$ and $p_{d,t}$ are the population probabilities of observing an 'up' and 'down' response, respectively. Imputing the observed shares U_t and D_t for the population probabilities, equation (6) determines σ_t up to the scaling constant τ. Since we evaluate the quantified disagreement, σ_t, using a scale-free criterion, our results are invariant to τ.[4]

Despite its popularity, the classical CP approach has been criticized for assuming normality of the latent distribution of point forecasts. Recent evidence against the normality assumption comes from Maag (2009) in his study of quantitative inflation expectations and Breitung and Schmeling (2013) in their study of stock market forecasts. Following Dasgupta and Lahiri (1992), we also experiment with the scaled t distribution to accommodate the excess kurtosis often found in quantitative expectations. Assuming constant thresholds that are symmetric around zero, the mean and standard deviation of the distribution of the unobserved point forecasts are

$$\mu_t = \tau\left(F_n^{-1}(p_{d,t}) + F_n^{-1}(1-p_{u,t})\right) / \left(F_n^{-1}(p_{d,t}) - F_n^{-1}(1-p_{u,t})\right) \tag{7}$$

$$\sigma_t = 2\tau\,(n/(n-1))^{(1/2)} / \left(F_n^{-1}(1-p_{u,t}) - F_n^{-1}(p_{d,t})\right) \tag{8}$$

where F_n is the cdf of a random variable with a t distribution having n degrees of freedom.

Alternatively, we assume a piecewise uniform distribution for the latent point forecasts with the following probability density function (pdf):

$$g(\mathrm{f}_{it}) = p_d\, u_{[b_d,\,-\tau]}(\mathrm{f}_{it}) + p_s\, u_{[-\tau,\,\tau]}(\mathrm{f}_{it}) + p_u\, u_{[\tau,\,b_u]}(\mathrm{f}_{it}) \tag{9}$$

[3] Following Carlson and Parkin (1975), we ignore 'no assessment' responses, since the share of respondents reporting 'no assessment' is typically very small and stable over time.

[4] Pesaran and Weale (2006, p. 741) outline several methods to estimate τ, including imposing a long-run unbiasedness assumption or relying on retrospective survey questions for identification.

Copyright © 2015 John Wiley & Sons, Ltd

In equation (9), p_d, p_s and p_u are the probabilities of 'down', 'same' and 'up' responses, respectively. $u_{[a,\,b]}(x)$ is the pdf of a uniform distribution on the interval $[a, b]$. $b_d = -2\tau - k\tau p_d^2$ and $b_u = 2\tau + k\tau p_u^2$ are the lower and upper bounds of the support for the latent distribution, respectively. τ and k are constants.

Some aspects are worth noting about this specification. First, for the latent point forecasts in each of the three categories, we assume a separate distribution that is invariant to the distribution of the responses across the other two categories. For example, the lower bound of the distribution in the 'down' category is a function of p_d, but not of p_s and p_u. Second, the upper and lower bounds of the support, b_u and b_d, shift outwards as the shares of 'up' and 'down' responses rise, respectively. For instance, the width of the interval of the uniform distribution in the 'up' category, $b_u - \tau$, obtains its minimum τ when $p_u = 0$ and its maximum $(1 + k)\tau$ when $p_u = 1$. The specification is consistent with the intuition that a greater proportion of 'up' ('down') responses implies latent point forecasts farther from the threshold value τ $(-\tau)$. Finally, k determines the width of the 'up' and 'down' intervals relative to the 'stay the same' interval, and is intentionally left as a tuning parameter that permits adapting the method to the dataset at hand. We would, for example, expect a higher chance of observing extreme forecasts in surveys of consumers than professional forecasters, suggesting a higher value of k in the former.

From the model specification in equation (9), we obtain the following expressions for the mean and standard deviation of the latent distribution:

$$\mu_t = \frac{\tau}{2}\left[p_u(1 + m_u) - p_d(1 + m_d)\right] \tag{10}$$

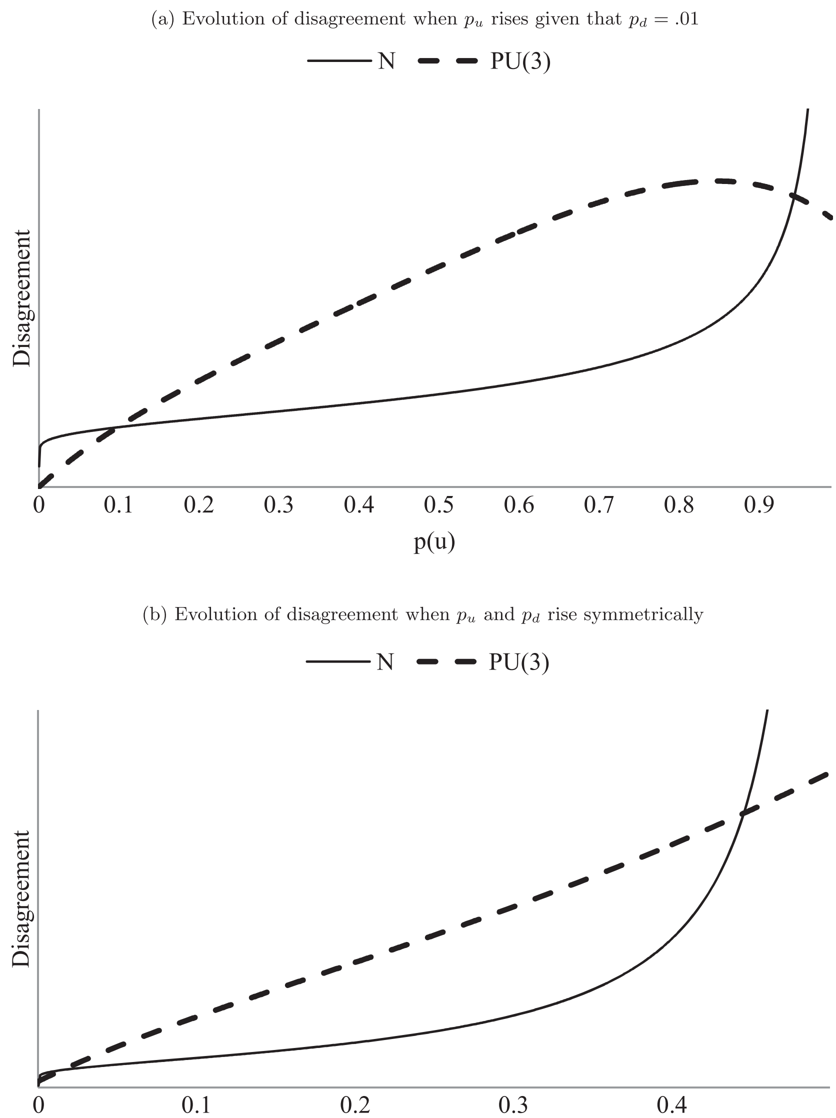

Figure 1. Simulated disagreement from the probability approach: Gaussian versus piecewise uniform distribution

Copyright © 2015 John Wiley & Sons, Ltd

$$\sigma_t = \sqrt{\frac{\tau^2}{3}\left[1 + p_u(m_u + m_u^2) + p_d(m_d + m_d^2)\right] - \mu_t^2} \tag{11}$$

where $m_u = 2 + k p_u^2$ and $m_d = 2 + k p_d^2$. Given k, we can estimate the disagreement up to a scaling factor τ by imputing U_t for p_u and D_t for p_d in equation (11). To estimate the tuning parameter, k, we experiment with the following values $\{1, 2, 3, 4\}$ and evaluate their performance in terms of the correlation with the benchmark.

To illustrate how the different distributional assumptions affect the disagreement estimate, we have undertaken two simulation experiments.[5] Figure 1(a) shows how estimated disagreement changes as a function of the share of ‘up’ responses, while holding the share of ‘down’ responses constant at 1%. The disagreement estimate using the normal distribution *increases monotonically* when the proportion of ‘up’ responses rises. This is counter-intuitive because we would expect that when the share of ‘up’ responses becomes dominant adding more responses to the ‘up’ category would *decrease* the disagreement estimate by making the distribution of responses more concentrated in a single category. Consistent with this intuition, the piecewise uniform distribution gives a disagreement measure that reaches the maximum when the proportion of ‘up’ responses is around 80% and decreases thereafter. As is also illustrated in the graph, when the proportion of ‘up’ responses is either very large or very small, the curve using the normal distribution becomes very steep. Therefore, the classical CP approach to estimating disagreement may be very sensitive to measurement errors in such states (cf. Löffler, 1999). By contrast, the disagreement estimate based on the piecewise uniform distribution is moderately sloped on its entire domain.

Figure 1(b) plots the estimates when the shares of ‘up’ and ‘down’ responses increase symmetrically. Notably, the disagreement estimate based on the normal distribution starts to explode quickly as the share of responses in the two categories increases. The reason is that the normal distribution cannot accommodate bimodality. To generate the low share of the ‘stay the same’ category, the distribution requires extremely high variance. By contrast, the disagreement estimate from our pragmatic alternative distribution rises roughly linearly. These simple simulations suggest that our modification is less prone to measurement errors and better at accommodating potential bimodality in survey responses. Admittedly, the piecewise uniform distribution is an approximation of the true distribution of the point forecasts, but due to the properties mentioned above it might better capture the crucial changes in the latent distribution over time.

Allowing for time-varying thresholds

The assumption of constant threshold values in the probability approach has often been questioned in the literature. Recent findings from Breitung and Schmeling (2013) and Lahiri and Zhao (2015) suggest that the constant threshold assumption explains much of the poor performance of the classical CP procedure. In the spirit of Smith and McAleer (1995), we extend the probability approach to allow for variation in thresholds over time. Specifically, we assume that the threshold parameter τ evolves as a random walk:

$$\tau_t = \tau_{t-1} + \nu_t \tag{12}$$

where ν_t is a Gaussian disturbance with mean 0 and variance σ_ν^2. With time-varying thresholds and normally distributed point forecasts, the mean and standard deviation of the latent distribution are

$$\mu_t = \tau_t \left(\Phi^{-1}(D_t) + \Phi^{-1}(1 - U_t)\right) / \left(\Phi^{-1}(D_t) - \Phi^{-1}(1 - U_t)\right) \tag{13}$$

$$\sigma_t = 2\tau_t / \left(\Phi^{-1}(1 - U_t) - \Phi^{-1}(D_t)\right) \tag{14}$$

Equivalent expressions can be obtained for the t distribution and the piecewise uniform distribution by replacing τ in equations (7–8) and (10–11) with τ_t. Our discussion below focuses on estimation of the disagreement measure for the normal distribution, but the treatment of the two alternative distributions is completely analogous.

To estimate τ_t, we need additional information beyond the raw aggregate shares of responses. Suppose we have a proxy, called y_t, for the mean of latent distribution, i.e. $y_t = \mu_t + u_t$, where the measurement error u_t is serially uncorrelated, homoscedastic and uncorrelated with ν_t at all leads and lags.[6] Denoting $\left(\Phi^{-1}(D_t) + \Phi^{-1}(1 - U_t)\right) / \left(\Phi^{-1}(D_t) - \Phi^{-1}(1 - U_t)\right)$ by x_t, we have

[5] The results with the t distribution are qualitatively identical to those with the normal distribution and thus not reported here.

[6] An alternative approach to estimating τ_t is available if, in addition to the forecasts f_{it}, we also observe an assessment of the past for which the realization on the quantitative scale is known to the respondents at the time when they make their qualitative assessment. In that case, the whole model can be estimated using the realization as y_t and the shares of the assessments as U_t and D_t. Estimated thresholds are then used to quantify the qualitative forecasts. This approach assumes that the categorization thresholds for the assessment are the same as for the forecast. We find the assumption implausible: it is likely that uncertainty surrounding the unobserved point forecasts drives the width of the interval between $\tau_{\mathrm{down},t}$ and $\tau_{\mathrm{up},t}$ (see, for example, Batchelor and Orr, 1988), but there is no reason for this to be the case regarding the realization. In addition, such qualitative assessments are rarely available in practice.

Copyright © 2015 John Wiley & Sons, Ltd

$$y_t = \tau_t x_t + u_t \tag{15}$$

Equations (12) and (15) specify a state-space model, in which y_t is the measurement variable and τ_t is the state variable. We estimate this model using the Kalman filter to obtain smoothed estimates of the time-varying thresholds (cf. Koopman, 1997). By imputing the estimated thresholds and the shares of responses into equation (14), we get the disagreement measure.

The estimation procedure above requires a valid proxy for the mean of the unobserved distribution. We experiment with several approaches. One approach is to use the realization of the target variable. For example, for inflation expectations formed in February 2012 with a 1-year horizon, we use the inflation rate between February 2012 and February 2013. This approach implicitly assumes that survey respondents have perfect foresight, which is arguably a very strict assumption. Alternatively, we use the forecasts generated by a time series model as a proxy. Following Atkeson and Ohanian (2001), our first method forecasts the annualized inflation rate at the horizon of interest by the average monthly inflation rate of the last 12 months. Our second method is a random walk forecast, which predicts that the inflation rate at the horizon of interest will be the same as it has been over the same horizon to the present date. For example, for inflation expectations formed in February 2012 with a 5-year horizon, we use the inflation rate between February 2007 and February 2012. These two simple methods have been found to produce good inflation forecasts (cf. Faust and Wright, 2013).

To avoid poor threshold estimates, we make the following adjustment when implementing the proxies for the mean of the latent distribution. In particular, we drop all observations of the proxy variable that fail to meet the following criteria: (i) the value of the proxy variable must be positive (negative) if the share of 'up' responses is bigger (smaller) than the share of 'down' responses; and (ii) the sign of the change in the proxy variable must coincide with the sign of the change in the balance statistic, defined as the difference between the share of 'up' and 'down' responses.[7] We thus end up with a proxy variable that has missing values. Fortunately, the state-space modeling approach used for threshold estimation can naturally accommodate these missing values. This adjustment procedure is motivated by the fact that the proxy may work poorly in certain situations. Consider the random walk forecast. If inflation has been high in the recent past, the random walk model will forecast high inflation in the future and vice versa. Since inflation is highly persistent, this approach provides accurate forecasts for most of the time but is unlikely to identify turning points in inflation expectations. In such situations, the threshold estimate without any adjustment is likely to be seriously biased. For this reason, we make the adjustment based on the observed qualitative forecasts when implementing the proxies for the mean of the latent distribution.

EMPIRICAL APPLICATIONS

In this section, we compare the different disagreement measures in two consumer surveys that collect both qualitative and quantitative responses: the University of Michigan Survey of Consumers and the Swedish Consumer Tendency Survey. Taking advantage of these datasets, we are able to directly assess the performance of the estimated disagreement series for the qualitative responses by comparing them with the quantitative benchmark: the cross-sectional standard deviation of the point forecasts.

Data description

University of Michigan Survey of Consumers

The University of Michigan Survey Research Center has conducted regular surveys of consumers in the USA for over five decades. While household inflation expectations in this survey are often used in the macroeconomics literature, to the best of our knowledge no recent study after Batchelor (1986) has examined the quality of the quantified disagreement in comparison to the disagreement in point forecasts. We consider inflation expectations with horizons of 1 year (i.e. short-run) and 5–10 years (i.e. long-run) for which overlapping qualitative and quantitative responses are available.

The qualitative forecast questions read:

1. 'During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?' and
2. 'What about the outlook for prices over the next 5 to 10 years? Do you think prices will be higher, about the same, or lower, 5 to 10 years from now?'

[7] Specifically, criterion (ii) is applied in the following way. Suppose that the proxy at period t meets both of our criteria. Next, we check whether the proxy at period $t+1$ is such that the sign of its change relative to period t coincides with the sign of changes in the balance statistic. If not, we treat the observation at period $t+1$ as missing and proceed with the observation at period $t+2$. We then evaluate whether the sign of changes in the proxy variable relative to period t coincides with the sign of changes in the balance statistic between t and $t+2$. If not, we again treat the observation at period $t+2$ as missing and proceed with evaluating the observation at $t+3$ relative to period t, etc.

Copyright © 2015 John Wiley & Sons, Ltd

Table I. Summary statistics

	Mean	Median	10th percentile	90th percentile
US short-run				
Average_t	3.67	3.65	2.73	4.58
SD_t	4.41	4.36	3.33	5.44
i.q.r._t	4.19	4.00	3.00	5.00
Skewness_t	1.12	1.07	0.16	2.06
Kurtosis_t	8.83	8.20	5.45	12.79
$\text{Share}_{\text{up},t}$	0.80	0.81	0.68	0.89
$\text{Share}_{\text{same},t}$	0.16	0.16	0.09	0.24
$\text{Share}_{\text{down},t}$	0.04	0.03	0.01	0.07
US long-run				
Average_t	3.61	3.39	3.08	4.47
SD_t	3.44	3.19	2.68	4.65
i.q.r._t	3.03	3.00	2.43	4.00
Skewness_t	1.21	1.10	0.32	2.33
Kurtosis_t	9.47	8.33	6.14	14.54
$\text{Share}_{\text{up},t}$	0.93	0.94	0.90	0.96
$\text{Share}_{\text{same},t}$	0.03	0.03	0.02	0.05
$\text{Share}_{\text{down},t}$	0.04	0.03	0.02	0.06
Swedish short-run				
Average_t	2.50	2.48	1.66	3.22
SD_t	4.04	3.97	3.56	4.55
i.q.r._t	3.67	3.50	3.00	5.00
Skewness_t	1.95	1.99	1.28	2.50
Kurtosis_t	14.17	13.80	10.88	17.23
$\text{Share}_{\text{up},t}$	0.65	0.64	0.52	0.78
$\text{Share}_{\text{same},t}$	0.30	0.31	0.20	0.37
$\text{Share}_{\text{down},t}$	0.05	0.04	0.02	0.10

Note: The table shows summary statistics for the point and qualitative forecasts. The statistics from average_t to kurtosis_t refer to point forecasts, whereas $\text{share}_{\text{up},t}$, $\text{share}_{\text{same},t}$, $\text{share}_{\text{down},t}$ are the shares in each category of qualitative responses. We obtain the summary statistics by (i) computing the statistic indicated by the corresponding row for each survey wave separately—e.g. the interquartile range for row i.q.r._t—which gives us a time series of summary statistics at the survey wave frequency, and (ii) summarizing each time series by the statistic indicated in the column—e.g. the column 'Median' gives the median of the survey-wave level interquartile range over time.

If the respondent answers 'go up' or 'go down', point forecasts are requested:

1. 'By about what percent do you expect prices to go up/down on the average, during the next 12 months?' and
2. 'By about what percent per year do you expect prices to go up/down on the average, during the next 5 to 10 years?'

Interviewers record no point forecast if a respondent predicts that prices will stay the same over the forecast horizon. In this case, we follow Curtin (1996) by imputing a zero for the missing point forecast.

For the short-run forecasts, the sample comprises 374 monthly survey waves from March 1982 through April 2013. We discard all previous observations because, according to Curtin (1996, p. 6), before March 1982 there was a problem in the interview design related to the 'stay the same' category, which may introduce severe biases.[8] For the long-run forecasts, we have 277 monthly surveys from April 1990 to April 2013. Before April 1990, these forecasts were only requested sporadically. The number of respondents was 600–700 before October 1987 and about 500 afterwards. To limit the impact of outliers, we discard all point forecasts of inflation outside the $\pm 30\%$ interval. As we do not use the point forecasts during the process of quantification of qualitative responses, we retain all non-missing

[8] Before March 1982, when respondents replied 'stay the same', it was not clear whether they referred to the price level or the inflation rate. As later data suggest, many respondents indeed referred to the inflation rate and in those cases the quantitative forecasts should have been recorded but were not.

Copyright © 2015 John Wiley & Sons, Ltd

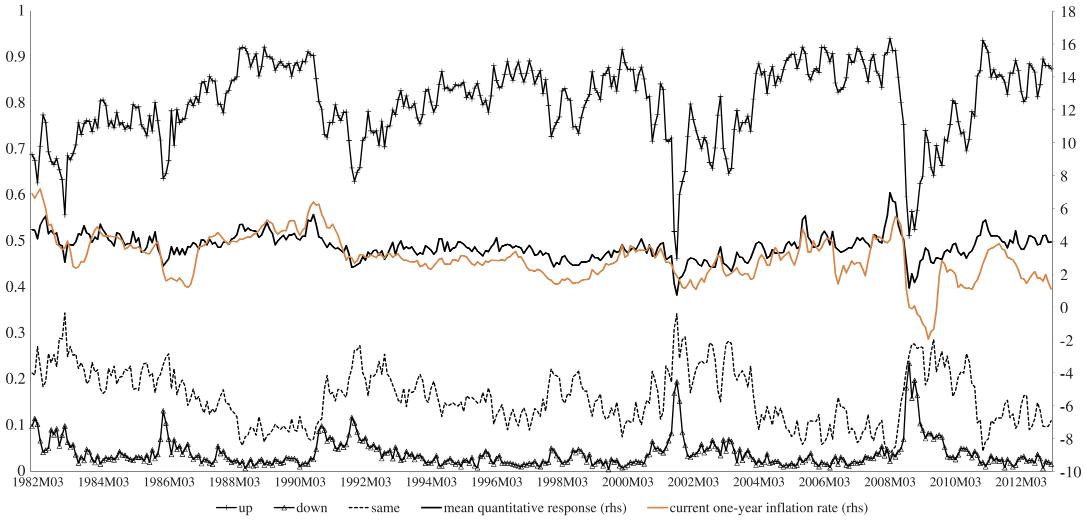

Figure 2. Short-run inflation expectations from the Michigan survey: March 1982–April 2013

qualitative responses, even if the matched point forecast is missing. This occurs in roughly 8% of all cases at the 1-year horizon and 10% for the long-run forecasts.[9]

Table I shows summary statistics for the qualitative and point forecasts. Considering the point forecasts, the average inflation expectations are 3.67% (short-run) and 3.61% (long-run), and disagreement measured by the standard deviation is higher for short-run (4.41% on average) than for long-run forecasts (3.44% on average). Both distributions are skewed to the right and display excess kurtosis, suggesting that the t distribution might provide a better fit to the data than the normal distribution. Differences among the two series are more pronounced in the qualitative responses. In the long-run forecast, there is very little variation in the shares of responses: the difference between the 10th and 90th percentiles of the share of 'up' responses is only 6 percentage points. This small variation suggests that it is difficult for the measurement approaches to replicate the admittedly large disagreement in point forecasts at the distant horizons. In contrast, there is much larger variation in the shares of qualitative responses for the 1-year-ahead forecasts.

Figures 2 and 3 visually compare the qualitative response shares, the average point forecast and the current inflation rate. The current inflation rate is measured as the year-over-year percentage change in the consumer price index (CPI) in the short run and as the 5-year annualized percentage change in the long run. In both figures, the average point forecast co-moves closely with the current inflation rate, suggesting that respondents adjust their forecasts upward in times of rising inflation, and vice versa. The respective correlation coefficients are 0.66 for the short-run and 0.86 for the long-run forecasts. However, this does not imply that the respondents make good inflation forecasts on average. The predictive correlations between the average forecasts and their realizations are rather low: 0.02 for the short run and 0.18 for the long run. This observation is consistent with recent imperfect information theories: when forming expectations, consumers face a tremendous amount of frictions and limitations in acquiring and processing of information (cf. Coibion and Gorodnichenko, 2012; Andrade and Le-Bihan, 2013). Turning to the qualitative short-run forecasts in Figure 2, the share of 'go up' responses mostly stays above 70%, even during periods with low or declining inflation. The share of 'go down' responses shows similar persistence and remains below 10% most of the time. Of particular interest are the two episodes around October 2001 and November 2008. Shortly after the September 11 attacks, in October 2001, the share of 'up' responses dropped from 72% to 52%: the largest single decline over the entire sample period. Similarly, the share of 'up' responses fell by 15 percentage points from October to November 2008, when the recent financial crisis peaked. By contrast, the share of 'up' responses in the long-run forecasts (Figure 3) consistently stays above 90% and shows very little variation over time.

Swedish consumer tendency survey

The Swedish Consumer Tendency Survey also elicits both qualitative and quantitative inflation expectations that are recorded in a two-step procedure. In the first step, each respondent is asked to report expected inflation on a category ordinal scale. The question reads:

> 'Compared to the situation today, do you think that in the next 12 months prices in general will increase faster/increase at the same rate/increase at a slower rate/stay about the same/fall slightly/don't know?'

[9] We have repeated our analysis by using only the data where both qualitative and quantitative responses are available. Our conclusions remain the same.

Copyright © 2015 John Wiley & Sons, Ltd

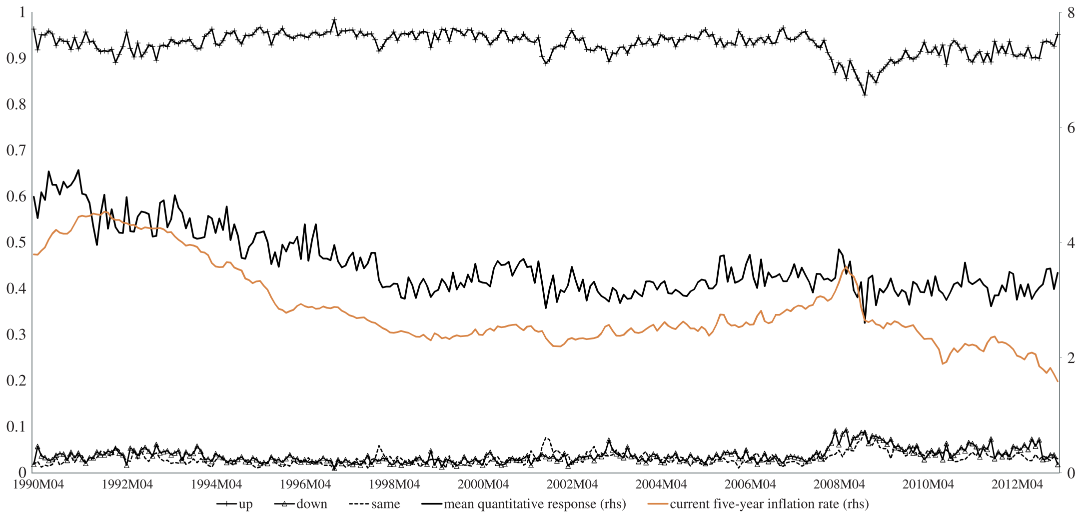

Figure 3. Long-run inflation expectations from the Michigan survey: April 1990–April 2013

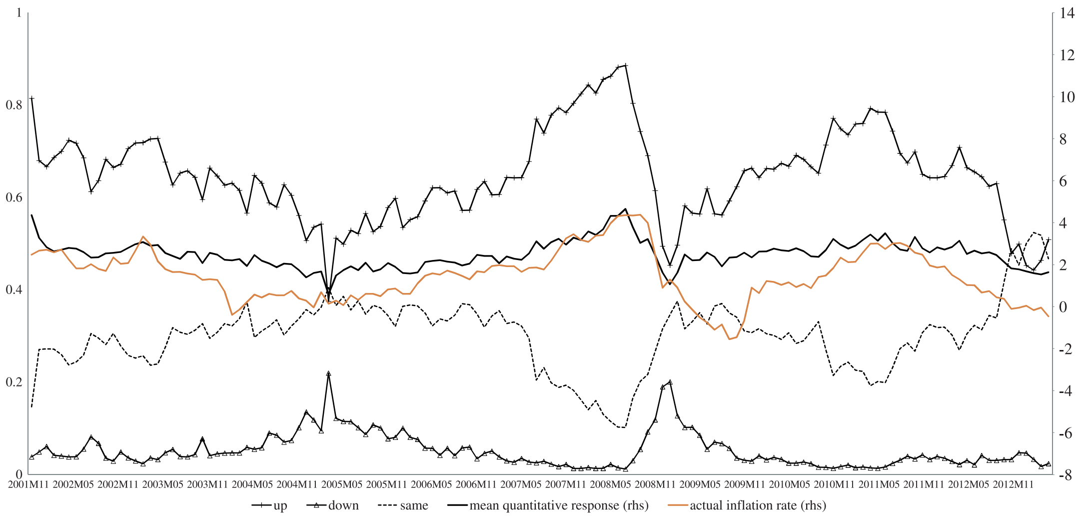

Figure 4. Short-run inflation expectations from the Swedish survey: November 2001–April 2013

A quantitative forecast is requested in the second step by asking:

> 'Compared with today, how much in percent do you think that prices will go up/down (i.e., the rate of inflation 12 months from now)?'

The sample used in our study is based on 138 monthly surveys of roughly 1500 households from November 2001 to April 2013. We discard all previous observations because, according to Maag (2009), there is a potential structural break due to a change in the surveying institution from Statistics Sweden to GfK Sweden in November 2001. To adjust for outliers, we consistently apply the methodology of discarding point forecasts outside the $\pm30\%$ interval, resulting in a 0.3% reduction in the sample size. In our sample, we retain all non-missing qualitative responses, even if the matched quantitative response is unavailable, which occurs in roughly 12% of all cases. Since our study focuses on three-category quantification approaches, we combine the qualitative responses that prices in general will increase 'faster,' 'at the same rate', and 'at a slower rate' into a single 'increase' category.

Table I presents summary statistics for Swedish consumer inflation expectations. Similar to the University of Michigan survey, Swedish point forecasts are skewed to the right and feature excess kurtosis. Notably, the corresponding qualitative expectations display more variation over time and have a less dominant share of 'up' responses than the short-run qualitative expectations for the USA. Figure 4 plots qualitative and quantitative expectations with the actual inflation rate measured as the year-over-year percentage change in the CPI. The average point forecast generally follows the pattern of actual inflation: the contemporaneous correlation between the two series is 0.73. The mean forecast, however, persistently overestimates the level of actual inflation. This overestimation is particularly pronounced during the most recent recession in 2009, when inflation expectations exceed actual inflation by roughly 3 percentage points on average. The large discrepancy between inflation expectations and actual inflation is

Copyright © 2015 John Wiley & Sons, Ltd

also evidenced by a negative predictive correlation between the two (correlation of −0.13). Considering the qualitative forecasts, the proportion of ‘go up’ responses exhibits high volatility, ranging from 38% to 89%, and is strongly correlated with the average point forecast (correlation of 0.94). By contrast, the share of ‘go down’ responses remains below 10% over the entire sample period, except for two spikes around March 2005 and January 2009.

Evaluation of alternative disagreement measures

We use the correlation coefficient to compare the alternative disagreement measures based on qualitative responses with the benchmark measure of disagreement obtained from quantitative responses (cross-sectional standard deviation). The correlation coefficient is a scale-free criterion and allows us to directly compare the involved series of different scales.[10] Given that the augmented Dickey–Fuller test rejects the unit root null hypothesis for the benchmark and most of estimated disagreement series, our discussion centers around the level of the series.[11] Table II summarizes the main results.

Panel A of Table II shows correlations between the benchmark measure of disagreement and the measures of nominal or ordinal variation. Although these measures give good disagreement estimates for the Michigan survey at the 1-year horizon (correlations of 0.43–0.54), their estimates are only weakly correlated with the benchmark for the other two series (correlations of 0.01–0.19). Among the measures of nominal or ordinal variation, the IQV performs the worst, possibly reflecting that it ignores the ordered nature of the qualitative responses. This poor performance in measuring disagreement of Swedish short-run inflation expectations is in sharp contrast to the recent result of Maag (2009) that the IQV dominates the classical CP approach. To understand the differences, note that Maag calculates the IQV based on the five-category qualitative responses, while we use the ‘transformed’ three-category responses. Indeed, when using the five-category responses, the correlation of the IQV with the benchmark disagreement becomes 0.36, which is very similar to what Maag (2009) finds (correlation of 0.33 in his Table A.II). The slight differences between Maag’s (2009) and our results are due to different sample periods: our sample ends in April 2013 rather than October 2008. Note also that in our sample the BL, COV and Reardon disagreement measures based on the five-category responses are all negatively related to the benchmark disagreement (correlations of −0.31, −0.34 and −0.30, respectively).[12] These findings suggest that the measures of nominal or ordinal variation are sensitive to the number of categories used in qualitative surveys. In addition, Maag (2009) finds that the probability approaches yield a slightly better disagreement estimate with three-category (correlation of 0.42) than with five-category response (correlation of 0.30). Future research is warranted to assess the performance of disagreement measures in the function of the number of categories.

Panel B of Table II demonstrates that the probability approach with constant thresholds works well, except for long-run inflation forecasts in the Michigan survey. Given the limited variation in the shares of responses at the distant horizons, the constant thresholds cannot accommodate the large disagreement in the corresponding point forecasts. Considering the short-run forecasts, while the normal distribution performs well only for the Swedish survey (correlation of 0.54), the t distribution with four to eight degrees of freedom (t_4-t_8) and the piecewise uniform distribution with the tuning parameter $k = 3$ (PU_3) perform quite well across both surveys (correlations of 0.44 to 0.56).

Panel C of Table II shows the performance of the probability approach with time-varying thresholds. Using the realization as a proxy for the mean of the latent distribution does not provide consistent improvement compared to constant thresholds, echoing the earlier result of the large discrepancy between the average point forecast and the realization. In contrast, using the model-based forecasts as a proxy (AO and RW) gives disagreement estimates that are highly correlated with the benchmark. For example, the correlations between $\text{PU}_{3,\text{RW}}$, which uses the random walk forecast of the target variable as a proxy, and the benchmark disagreement are 0.67 and 0.82 for the short- and long-run inflation forecasts in the Michigan survey, respectively. In addition, we explore the impact of keeping only the reasonable observations for the proxy variable, as outlined above (‘Allowing for time-varying thresholds’). Table A.II reports the results with and without making this adjustment. The adjustment increases correlations of the quantified disagreement measures with the benchmark by 8 percentage points on average and by 47 percentage points in the best-case scenario.

In Panel D of Table II, we consider combinations of the individual approaches. Specifically, we average the standardized individual disagreement series in each panel (panel A, B and C). These intra-panel averaged disagreement estimates perform well, closely tracking the best individual method within each panel.

To conclude, our general findings can be summarized as follows:

[10] The results using Spearman’s rank correlation are very similar to those using the simple correlation coefficients and are thus not reported here.

[11] We conduct the analysis based on the first difference of all disagreement series. Table A.I shows that the correlations tend to be lower than among the variables in levels and the probability approaches with constant thresholds seem to perform better than other measures. We have to interpret these results with caution, bearing in mind that there is a problem of over-differencing since the benchmark and most of the estimated disagreement series are stationary.

[12] Another possible explanation is that the share of ‘stay the same’ responses is much larger and display more variation over time in the Swedish survey than the Michigan survey. The size of the ‘stay the same’ category might also matter for the correspondence between the quantitative and qualitative disagreement measures.

Copyright © 2015 John Wiley & Sons, Ltd

Table II. Accuracy of alternative measures for disagreement in qualitative expectations

	US short-run	US long-run	Swedish short-run
(A) *Measures of nominal or ordinal variation*			
IQV	0.43	0.04	0.01
BL	0.50	0.09	0.11
COV	0.51	0.09	0.13
Reardon	0.51	0.10	0.14
BES	0.54	0.13	0.19
(B) *Probability approaches with constant thresholds*			
N	0.29	0.21	0.54
t_2	0.61	0.24	0.37
t_4	0.54	0.24	0.46
t_6	0.48	0.23	0.50
t_8	0.44	0.23	0.52
PU_1	0.54	0.13	0.25
PU_2	0.57	0.13	0.42
PU_3	0.56	0.14	0.52
PU_4	0.19	0.16	0.41
(C) *Probability approaches with time-varying thresholds*			
N, realiz.	0.35	0.43	0.06
t_4, realiz.	0.42	0.39	0.10
PU_3, realiz.	0.35	0.47	0.00
N, AO	0.70	0.54	0.49
t_4, AO	0.72	0.51	0.54
PU_3, AO	0.69	0.62	0.49
N, RW	0.68	0.28	0.42
t_4, RW	0.70	0.35	0.48
PU_3, RW	0.67	0.82	0.41
(D) *Combination of alternative approaches*			
Panel A	0.50	0.09	0.12
Panel B	0.55	0.24	0.51
Panel C	0.66	0.73	0.33

Note: For each measure of disagreement in qualitative forecasts, this table shows its correlation with the standard deviation of the corresponding point forecasts. Panel A includes the measures of nominal or ordinal variation: the index of qualitative variation (IQV); Blair and Lacy's (2000) measure of ordinal variation (BL); the coefficient of ordinal variation (COV); Reardon's (2009) entropy-based measure of ordinal variation (Reardon); Bachmann *et al.*'s (2013) measure of disagreement (BES). Panel B considers the probability approach with time-constant thresholds for a range of alternative distributional assumptions: N uses the normal distribution, t_{dof} uses the t distribution with dof degrees of freedom, PU_k uses the piecewise uniform distribution with tuning parameter k. Panel C analyzes the probability approach with time-varying thresholds for a range of different distributions and alternative proxy variables for the mean of the latent distribution of point forecasts. Specifically, 'realiz.' uses the realization of the target variable, whereas AO and RW use model-based forecasts of the target variable. Panel D shows the intra-group combinations: average across the measures in panel A, B and C, respectively.

- The measures of nominal or ordinal variation fail to robustly correlate with the benchmark measures of disagreement.
- The probability approach with constant thresholds and the t or the piecewise uniform distribution performs well for short-run forecasts.
- The probability approach with time-varying thresholds performs best when using model-based forecasts as a proxy for the mean of the latent distribution.
- The combination of alternative approaches insures against selecting an inferior individual approach to measuring disagreement in qualitative expectations.

We conduct additional analyses to check the robustness of our results with respect to: (i) an alternative benchmark disagreement measure in the point forecasts; (ii) an alternative truncation scheme for the point forecasts; (iii) a different sample period excluding the recent financial crisis; and (iv) the correlation between disagreement and the business cycle.

First, we use the interquartile range as a benchmark measure of disagreement in point forecasts. The results in Table A.III show that the measures of nominal or ordinal variation (panel A) and the probability approaches with

Copyright © 2015 John Wiley & Sons, Ltd

constant thresholds (panel B) display lower correlations with this alternative benchmark than with the standard deviation of point forecasts. Except for a few cases, the probability approach with time-varying threshold (panel C) performs similarly as before.

Second, we experiment with a different truncation scheme for the point forecasts, truncating at −5%/+30% as in Pfajfar and Santoro (2013); cf. Table A.IV. The probability approaches with both constant (panel B) and time-varying thresholds (panel C) perform well and similarly to those with a truncation at −30%/+30%. In contrast, the measures of nominal or ordinal variation (panel A) become less correlated with the benchmark, suggesting a lack of robustness of these measures due to ignoring the distribution of point forecasts.

Third, we redo the estimation based on a short sample ending in December 2006 by excluding the recent financial crisis period; cf. Table A.V. For both series in the Michigan survey, most disagreement measures show a slight improvement in terms of the correlation with the benchmark. For the Swedish survey, however, we see severe deteriorations in their correlations with the benchmark for almost all disagreement measures. When interpreting the latter result, we have to keep in mind the small sample size—only 5 years of data.

Finally, we explore the relation between disagreement and the business cycle. Figure 5 plots the quantified disagreement measure $t_{4,\mathrm{RW}}$ and Figures A.1–A.3 plot the averaged disagreement measure from each panel. Evidently, disagreement tends to rise during the recessions and $t_{4,\mathrm{RW}}$ tracks the benchmark very closely. To further illustrate the association between disagreement and macroeconomic variables, we report their correlations in Table A.VI. The benchmark disagreement measures, using either standard deviation or interquartile range, tend to (i) rise with the level of inflation, (ii) rise with the variability of inflation and (iii) decline with industrial production at least for short-run forecasts. As for quantified disagreement measures, the probability approach with time-varying thresholds (panel C) behaves similarly to the benchmark when using model-based forecasts as a proxy for the mean of the latent distribution (AO and RW). The disagreement measures based on constant thresholds (panel B) and nominal or ordinal variation (panel A), however, display weakly positive or even negative correlations with the level and variability of inflation. The latter result might reflect that these measures cannot accommodate the time-varying nature of the underlying inflation rate, again casting doubt on their reliability to capture the actual disagreement among consumers.

CONCLUDING REMARKS

During the past decade, many central banks and academic researchers have begun to explore ways to better analyze consumer and firm expectations from survey data. However, most consumer and business surveys only provide qualitative expectations. Our paper contributes to this ongoing literature by providing a detailed review and assessment of the measurement of disagreement in qualitative survey data. We consider about two dozen individual approaches, which may be categorized as measures of dispersion in nominal and ordinal variables and the probability approach with alternative distribution and threshold assumptions. Using data from two household surveys that collect both qualitative and quantitative inflation expectations, we find that the accuracy of these approaches varies greatly and depends on the forecast horizon and the distribution of responses in the survey. In terms of the correlation with established measures of disagreement in point forecasts (cross-sectional standard deviation or interquartile range), the probability approach performs better than the measures of dispersion in nominal and ordinal variables. Within the probability approach, the variants that use time-varying thresholds tend to outperform those with constant thresholds, and the ones with the t or piecewise uniform distribution tend to dominate those with the normal distribution. Our results provide guidance to the growing field of applied macroeconomics research that studies the disagreement among non-professional forecasters, who typically report qualitative expectations.

Admittedly, the quantified disagreement measures are not perfect. Since our results suggest that respondents use time-varying categorization thresholds, one possible solution is to explicitly state the thresholds in the questionnaire. As an example, one could ask whether the respondent expects prices to go down by more than 1%/not change by more than 1% in either direction/rise by more than 1% over the next 12 months. Another alternative is to ask respondents to report their thresholds directly. The ultimate solution is, of course, to elicit point forecasts, or even subjective probability distributions as advised in Manski (2004). However, the practice of asking qualitative questions in many surveys, in particular those with large panels of non-professional forecasters, will probably not be discontinued in the near future. Consistent with this sentiment, the probability approaches for the analysis of qualitative expectations may provide useful tools to characterize disagreement for macroeconomic modeling and forecasting.

ACKNOWLEDGEMENTS

This paper was presented at the 19th Federal Forecasters Conference and Society of Government Economists Annual Conference in Washington, DC. We thank the participants in the conferences for helpful comments and suggestions. We also thank Derek Bunn (the editor) and two anonymous referees for comments, which have significantly improved

Copyright © 2015 John Wiley & Sons, Ltd

the paper. The research for this paper was conducted while Frieder Mokinski was a researcher at ZEW Mannheim. Xuguang (Simon) Sheng would like to acknowledge the financial support from ZEW's Heinz Konig Young Scholar Award. Jingyun Yang's research was supported by the National Institutes of Health/National Institute on Aging grant R01AG036042 and the Illinois Department of Public Health. Opinions expressed in this paper do not necessary reflect the views of the Deutsche Bundesbank or its staff.

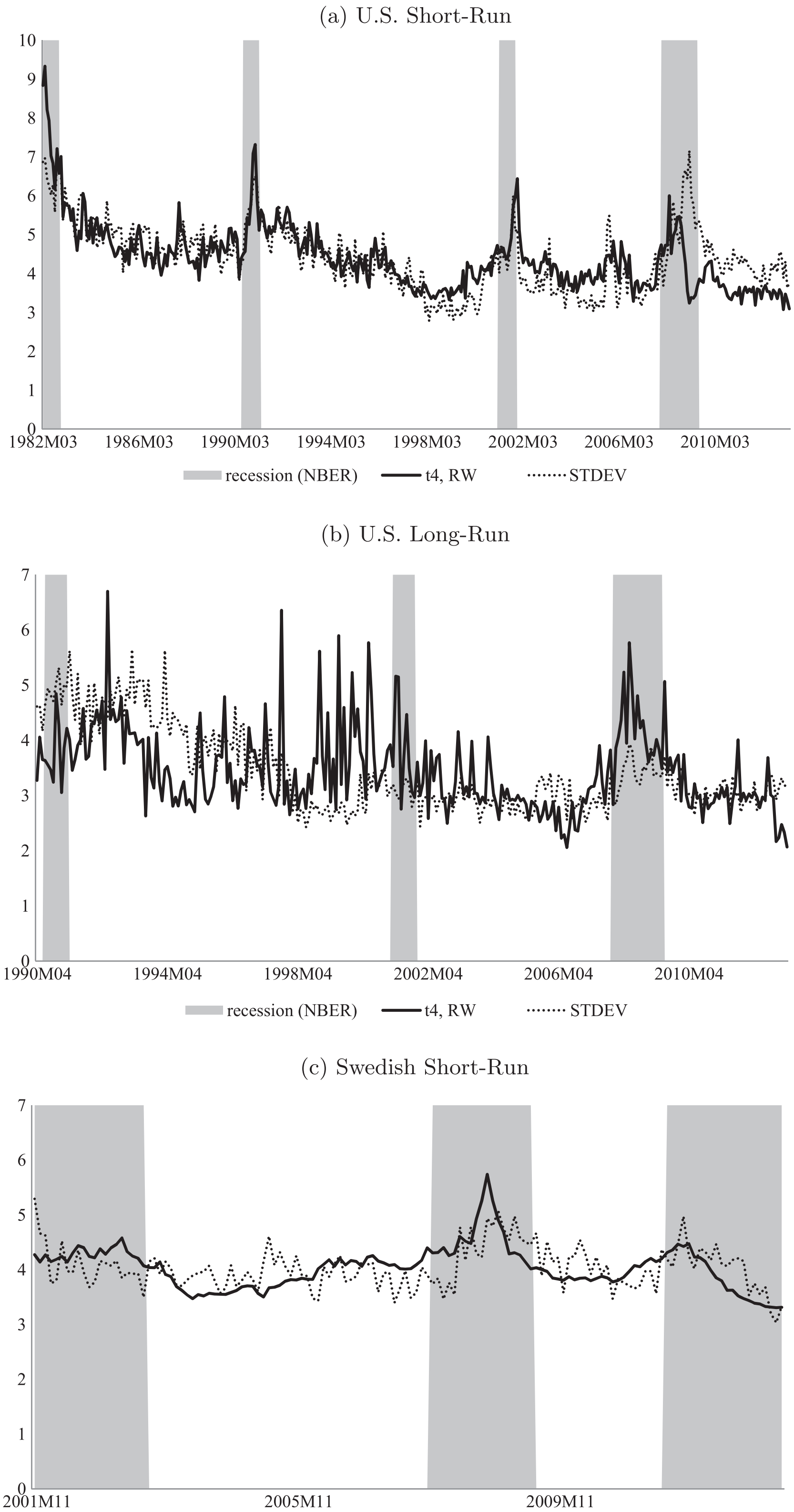

Figure 5. Disagreement and the business cycle

Copyright © 2015 John Wiley & Sons, Ltd

REFERENCES

Andrade P, Le-Bihan H. 2013. Inattentive professional forecasters. *Journal of Monetary Economics* **60**: 967–982.

Ang A, Bekaert G, Wei M. 2007. Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics* **54**: 1163–1212.

Atkeson A, Ohanian L. 2001. Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review* **25**(1): 2–11.

Bachmann R, Elstner S, Sims E. 2013. Uncertainty and economic activity: evidence from business survey data. *American Economic Journal: Macroeconomics* **5**: 217–249.

Batchelor RA. 1986. Quantitative v. qualitative measures of inflation expectations. *Oxford Bulletin of Economics and Statistics* **48**(2): 99–120.

Batchelor RA, Orr AB. 1988. Inflation expectations revisited. *Economica* **55**(219): 317–331.

Blair J, Lacy M. 2000. Statistics of ordinal variation. *Sociological Methods and Research* **28**: 251–280.

Breitung J, Schmeling M. 2013. Quantifying survey expectations: what's wrong with the probability approach? *International Journal of Forecasting* **29**: 142–154.

Carlson JA, Parkin MJ. 1975. Inflation expectations. *Economica* **42**: 123–138.

Carroll CD. 2003. Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of Economics* **118**(1): 269–298.

Coibion O, Gorodnichenko Y. 2012. What can survey forecasts tell us about information rigidities? *Journal of Political Economy* **120**: 116–159.

Curtin R. 1996. Procedure to estimate price expectations. Working paper. University of Michigan Survey Research Center.

Curtin R. 2007. Consumer sentiment surveys: worldwide review and assessment. *Journal of Business Cycle Measurement and Analysis* **3**: 7–42.

Dasgupta S, Lahiri K. 1992. A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data. *Journal of Business and Economic Statistics* **10**(4): 391–400.

Dasgupta S, Lahiri K. 1993. On the use of dispersion measures from NAPM surveys in business cycle forecasting. *Journal of Forecasting* **12**(3–4): 239–253.

Dovern J, Fritsche U, Slacalek J. 2012. Disagreement among forecasters in G7 countries. *Review of Economics and Statistics* **94**(4): 1081–1096.

Dräger L, Lamla M. 2012. Updating inflation expectations: evidence from micro-data. *Economics Letters* **117**(3): 807–810.

Ehrmann M, Eijffinger S, Fratzscher M. 2012. The role of central bank transparency for guiding private sector forecasts. *Scandinavian Journal of Economics* **114**(3): 1018–1052.

Faust J, Wright JH. 2013. Forecasting inflation. In *Handbook of Economic Forecasting*, Vol. 2, Elliott G, Timmermann A (eds). Elsevier: Amsterdam; 2–56.

Gibbs JP, Poston DL. 1975. The division of labor: conceptualization and related measures. *Social Forces* **53**(3): 468–476.

Hubert P. 2014. FOMC forecasts as a focal point for private expectations. *Journal of Money, Credit and Banking* **46**: 1381–1420.

Koopman SJ. 1997. Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* **92**(440): 1630–1638.

Kvalseth T. 1995. Coefficients of variation for nominal and ordinal categorical data. *Perceptual and Motor Skills* **80**: 843–847.

Lahiri K, Sheng X. 2008. Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics* **144**(2): 325–340.

Lahiri K, Sheng X. 2010. Measuring forecast uncertainty by disagreement: the missing link. *Journal of Applied Econometrics* **25**(2): 514–538.

Lahiri K, Zhao Y. 2015. Quantifying survey expectations: a critical review and generalization of the Carlson–Parkin method. *International Journal of Forecasting* **31**: 51–62.

Lamla M, Maag T. 2012. The role of media for inflation forecast disagreement of households and professional forecasters. *Journal of Money, Credit and Banking* **44**: 1325–1350.

Löffler G. 1999. Refining the Carlson–Parkin method. *Economics Letters* **64**(2): 167–171.

Maag T. 2009. On the accuracy of the probability method for quantifying beliefs about inflation. KOF Swiss Economic Institute working paper 230.

Mankiw NG, Reis R. 2002. Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics* **117**: 1295–1328.

Mankiw NG, Reis R, Wolfers J. 2004. Disagreement about inflation expectations. In *NBER Macroeconomics Annual 2003*, Gertler M, Rogoff K (eds). MIT Press: Cambridge, MA; 209–248.

Manski C. 2004. Measuring expectations. *Econometrica* **72**: 1329–1376.

Nardo M. 2003. The quantification of qualitative survey data: a critical assessment. *Journal of Economic Surveys* **17**(5): 645–668.

Pesaran MH, Weale M. 2006. Survey expectations. In *Handbook of Economic Forecasting*, Vol. 1, Elliott G, Granger CW, Timmermann A (eds). North–Holland: Amsterdam; 715–776.

Pfajfar D, Santoro E. 2013. News on inflation and the epidemiology of inflation expectations. *Journal of Money, Credit and Banking* **45**(6): 1045–1067.

Reardon SF. 2009. Measures of ordinal segregation. *Research on Economic Inequality* **17**: 129–155.

Smith J, McAleer M. 1995. Alternative procedures for converting qualitative response data to quantitative expectations: an application to Australian manufacturing. *Journal of Applied Econometrics* **10**(2): 165–185.

Theil H. 1955. Recent experiences with the Munich business test: an expository article. *Econometrica* **23**: 184–192.

Authors' biographies:

Frieder Mokinski is an Economist at Deutsche Bundesbank. His reseach interests include time series econometrics, survey expectations data and economic forecasting.

Copyright © 2015 John Wiley & Sons, Ltd

Xuguang (Simon) Sheng is an Assistant Professor of Economics at American University. His research interests include time series econometrics, economic and financial forecasting. He has published in Journal of Econometrics, Journal of Applied Econometrics, Journal of Accounting and Economics, and Oxford Handbook on Economic Forecasting.

Jingyun Yang is an Assistant Professor of Rush Alzheimer's Disease Center at Rush University Medical Center. His research interests include time Alzheimer's disease, genetic epidemiology and time-varying effect model. He has published more than 40 papers in journals such as Computational Statistics & Data Analysis, Oxford Bulletin of Economics and Statistics, Communications in Statistics, Diabetes and Neurology.

Authors' addresses:

Frieder Mokinski, Deutsche BundesbankP.O. Box 10 06 0260006 Frankfurt am MainGermany.

Xuguang (Simon) Sheng, Department of Economics, American University, 4400 Massachusetts Avenue NW, Washington, DC, 20016, USA.

Jingyun Yang, Rush Alzheimer's Disease Center Rush University Medical Center 600 S. Paulina ST, Suite 1029 Chicago, IL 60612.

APPENDIX

Table A.I. Accuracy of alternative measures based on changes in forecast disagreement

	US short-run	US long-run	Swedish short-run
(A) *Measures of nominal or ordinal variation*			
IQV	0.10	0.17	−0.13
BL	0.22	0.25	−0.04
COV	0.24	0.25	−0.03
Reardon	0.24	0.26	−0.02
BES	0.29	0.28	0.02
(B) *Probability approaches with constant thresholds*			
N	0.34	0.20	0.35
t_2	0.41	0.27	0.18
t_4	0.39	0.24	0.25
t_6	0.38	0.23	0.29
t_8	0.37	0.22	0.31
PU_1	0.31	0.28	0.06
PU_2	0.36	0.29	0.18
PU_3	0.39	0.30	0.31
PU_4	0.34	0.31	0.32
(C) *Probability approaches with time-varying thresholds*			
N, realiz.	0.31	0.24	0.17
t_4, realiz.	0.34	0.27	0.14
PU_3, realiz.	0.03	0.17	−0.08
N, AO	0.36	0.23	0.28
t_4, AO	0.39	0.27	0.26
PU_3, AO	0.12	0.15	0.11
N, RW	0.35	0.18	0.23
t_4, RW	0.38	0.22	0.22
PU_3, RW	0.10	0.22	0.08
(D) *Combination of alternative approaches*			
Panel A	0.23	0.25	−0.04
Panel B	0.40	0.28	0.29
Panel C	0.39	0.31	0.15

Note: This table reproduces Table II considering first differences of the disagreement measures instead of their levels.

Copyright © 2015 John Wiley & Sons, Ltd

Table A.II. The impact of variable adjustment on the estimated disagreement measures

	US short-run		US long-run		Swedish short-run	
	Level	Change	Level	Change	Level	Change
N, realiz.	0.35	0.23	−0.04	0.20	−0.12	−0.03
Adjusted	0.35	0.31	0.43	0.24	0.06	0.17
t_4, realiz.	0.42	0.28	0.02	0.24	−0.10	−0.02
Adjusted	0.42	0.34	0.39	0.27	0.10	0.14
PU_3, realiz.	0.36	0.07	0.15	0.16	−0.07	−0.07
Adjusted	0.35	0.03	0.47	0.17	0.00	−0.08
N, AO	0.57	0.32	0.37	0.21	0.27	0.07
Adjusted	0.70	0.36	0.54	0.23	0.49	0.28
t_4, AO	0.60	0.36	0.36	0.25	0.31	0.09
Adjusted	0.72	0.39	0.51	0.27	0.54	0.26
PU_3, AO	0.61	0.17	0.40	0.18	0.33	0.06
Adjusted	0.69	0.12	0.62	0.15	0.49	0.11
N, RW	0.52	0.30	0.79	0.30	0.23	0.06
Adjusted	0.68	0.35	0.28	0.18	0.42	0.23
t_4, RW	0.55	0.34	0.74	0.30	0.26	0.08
Adjusted	0.70	0.38	0.35	0.22	0.48	0.22
PU_3, RW	0.60	0.14	0.81	0.24	0.31	0.06
Adjusted	0.67	0.10	0.82	0.22	0.41	0.08

Note: This table shows the impact of adjusting the proxy variable for the mean of the latent distribution of point forecasts (see the end of section on 'Allowing for time-varying thresholds for detailed discussion on the adjustment. For each measure of disagreement, the table shows its correlation with the standard deviation of the corresponding point forecasts. For each block, the first row refers to the disagreement measure without any adjustment and the second row ('Adjusted') refers to the disagreement measure with the adjustment. Columns 'Level' consider the correlation among the levels of the disagreement measures, whereas columns 'Change' consider the correlation based on changes in disagreement estimates. The approaches displayed here are the probability approach with time-varying thresholds for a range of different distributions and alternative proxy variables for the mean of the latent distribution of point forecasts. Specifically, N uses the normal distribution, t_4 uses the t distribution with four degrees of freedom, PU_3 uses the piecewise uniform distribution with tuning parameter $k = 3$. 'realiz.' uses the realization of the target variable as a proxy, whereas AO and RW use model-based forecasts of the target variable as a proxy.

Table A.III. Accuracy of alternative measures for disagreement in qualitative expectations using IQR as an alternative benchmark measure

	US short-run	US long-run	Swedish short-run
(A) *Measures of nominal or ordinal variation*			
IQV	0.54	−0.09	−0.30
BL	0.56	−0.07	−0.30
COV	0.55	−0.07	−0.31
Reardon	0.57	−0.06	−0.03
BES	0.57	−0.04	−0.01
(B) *Probability approaches with constant thresholds*			
N	0.13	0.14	0.04
t_2	0.51	0.08	−0.20
t_4	0.41	0.12	−0.12
t_6	0.33	0.13	−0.08
t_8	0.28	0.13	−0.05

Copyright © 2015 John Wiley & Sons, Ltd

Table A.III. (Continued)

	US short-run	US long-run	Swedish short-run
PU_1	0.57	−0.04	−0.21
PU_2	0.56	−0.04	−0.02
PU_3	0.46	−0.03	0.29
PU_4	−0.03	−0.02	0.42
(C) *Probability approaches with time-varying thresholds*			
N, realiz.	0.36	0.42	−0.06
t_4, realiz.	0.42	0.40	−0.09
PU_3, realiz.	0.40	0.46	−0.06
N, AO	0.69	0.49	0.43
t_4, AO	0.72	0.47	0.40
PU_3, AO	0.75	0.59	0.46
N, RW	0.67	0.24	0.43
t_4, RW	0.70	0.28	0.42
PU_3, RW	0.73	0.49	0.46
(D) *Combination of alternative approaches*			
Panel A	0.56	−0.07	−0.29
Panel B	0.41	0.07	0.01
Panel C	0.65	0.62	0.29

Note: This table reproduces Table II using the interquartile range as an alternative measure of disagreement in the point forecasts.

Table A.IV. Accuracy of alternative measures for disagreement in qualitative expectations using an alternative truncation scheme for the quantitative forecasts

	US short-run	US long-run	Swedish short-run
(A) *Measures of nominal or ordinal variation*			
IQV	0.26	−0.01	−0.15
BL	0.30	0.03	−0.07
COV	0.31	0.03	−0.06
Reardon	0.31	0.04	−0.04
BES	0.34	0.07	0.00
(B) *Probability approaches with constant thresholds*			
N	0.26	0.21	0.43
t_2	0.42	0.20	0.18
t_4	0.41	0.22	0.28
t_6	0.38	0.22	0.33
t_8	0.35	0.22	0.36
PU_1	0.35	0.07	0.06
PU_2	0.37	0.07	0.23
PU_3	0.39	0.08	0.43
PU_4	0.18	0.10	0.42
(C) *Probability approaches with time-varying thresholds*			
N, realiz.	0.37	0.42	0.02
t_4, realiz.	0.40	0.37	0.03
PU_3, realiz.	0.40	0.48	−0.02
N, AO	0.70	0.53	0.50
t_4, AO	0.69	0.49	0.52
PU_3, AO	0.70	0.63	0.50
N, RW	0.69	0.27	0.45
t_4, RW	0.68	0.32	0.48
PU_3, RW	0.69	0.82	0.45

Copyright © 2015 John Wiley & Sons, Ltd

Table A.IV. (Continued)

	US short-run	US long-run	Swedish short-run
(D) *Combination of alternative approaches*			
Panel A	0.31	0.03	−0.06
Panel B	0.41	0.20	0.35
Panel C	0.65	0.70	0.32

Note: This table reproduces Table II using an alternative truncation scheme for point forecasts. Specifically, instead of truncating at −30%/30%, we truncate the point forecasts at −5%/30% as in Pfajfar and Santoro (2013).

Table A.V. Accuracy of alternative measures for disagreement in qualitative expectations on an alternative sample extending only until 2006:M12

	US short-run	US long-run	Swedish short-run
(A) *Measures of nominal or ordinal variation*			
IQV	0.38	0.16	−0.35
BL	0.44	0.25	−0.27
COV	0.45	0.25	−0.25
Reardon	0.45	0.26	−0.25
BES	0.48	0.30	−0.21
(B) *Probability approaches with constant thresholds*			
N	0.32	0.22	0.39
t_2	0.56	0.32	−0.03
t_4	0.53	0.28	0.11
t_6	0.49	0.27	0.19
t_8	0.46	0.26	0.24
PU_1	0.49	0.30	−0.17
PU_2	0.51	0.31	−0.02
PU_3	0.51	0.32	0.34
PU_4	0.19	0.34	0.44
(C) *Probability approaches with time-varying thresholds*			
N, realiz.	0.64	0.40	0.10
t_4, realiz.	0.66	0.34	0.09
PU_3, realiz.	0.61	0.58	0.08
N, AO	0.84	0.58	0.24
t_4, AO	0.83	0.55	0.26
PU_3, AO	0.80	0.71	0.20
N, RW	0.84	0.23	0.24
t_4, RW	0.83	0.32	0.26
PU_3, RW	0.80	0.85	0.21
(D) *Combination of alternative approaches*			
Panel A	0.44	0.25	−0.27
Panel B	0.52	0.35	0.21
Panel C	0.83	0.73	0.28

Note: This table reproduces Table II by excluding the recent financial crisis period.

Copyright © 2015 John Wiley & Sons, Ltd

Table A.VI. Disagreement and macroeconomic variables: simple correlations

	(1) Level of inflation			(2) Squared changes in inflation			(3) Industrial production		
	US SR	US LR	Sw. SR	US SR	US LR	Sw. SR	US SR	US LR	Sw. SR
Benchmark disagreement measures based on point forecasts									
STDEV	0.33	0.84	0.13	0.42	0.11	0.20	−0.43	−0.01	−0.45
IQR	0.25	0.58	0.33	0.49	0.13	0.07	−0.40	0.10	−0.05
$STDEV_{[-5,30]}$	0.45	0.83	0.24	0.33	0.14	0.19	−0.29	0.04	−0.37
(A) *Measures of nominal or ordinal variation*									
IQV	−0.33	−0.01	−0.74	0.41	−0.31	0.07	−0.43	−0.60	−0.30
BL	−0.30	0.02	−0.66	0.44	−0.31	0.05	−0.51	−0.61	−0.36
COV	−0.30	0.02	−0.64	0.44	−0.30	0.05	−0.54	−0.60	−0.37
Reardon	−0.29	0.03	−0.64	0.43	−0.32	0.04	−0.52	−0.62	−0.37
BES	−0.25	0.05	−0.59	0.44	−0.32	0.03	−0.54	−0.62	−0.38
(B) *Probability approaches with constant thresholds*									
N	0.30	0.15	0.32	0.05	0.09	−0.07	−0.29	0.06	−0.28
t_2	−0.04	0.14	−0.31	0.39	−0.11	−0.01	−0.64	−0.31	−0.42
t_4	0.11	0.16	−0.13	0.28	−0.02	−0.03	−0.55	−0.15	−0.41
t_6	0.18	0.16	−0.01	0.21	0.01	−0.04	−0.48	−0.08	−0.39
t_8	0.21	0.16	0.06	0.17	0.03	−0.05	−0.44	−0.05	−0.37
PU_1	−0.25	0.05	−0.51	0.44	−0.32	0.01	−0.55	−0.62	−0.38
PU_2	−0.20	0.05	−0.22	0.42	−0.31	−0.04	−0.57	−0.62	−0.34
PU_3	−0.05	0.06	0.37	0.34	−0.31	−0.12	−0.55	−0.61	−0.13
PU_4	0.28	0.07	0.69	−0.04	−0.30	−0.13	−0.17	−0.61	0.08
(C) *Probability approaches with time-varying thresholds*									
N, realiz.	0.38	0.46	0.34	0.00	−0.06	−0.03	0.13	0.14	0.13
t_4, realiz.	0.33	0.41	0.24	0.09	−0.08	−0.03	−0.01	0.10	0.07
PU_3, realiz.	0.35	0.62	0.31	0.01	0.08	−0.02	0.18	0.46	0.16
N, AO	0.66	0.57	0.70	0.21	−0.16	0.02	−0.25	0.05	−0.22
t_4, AO	0.60	0.53	0.62	0.24	−0.18	0.03	−0.31	−0.01	−0.29
PU_3, AO	0.59	0.69	0.64	0.28	−0.09	0.11	−0.27	0.14	−0.28
N, RW	0.68	0.32	0.78	0.17	0.26	−0.05	−0.21	0.39	−0.11
t_4, RW	0.62	0.39	0.72	0.21	0.14	−0.04	−0.27	0.21	−0.18
PU_3, RW	0.61	0.91	0.76	0.25	0.28	0.00	−0.23	0.25	−0.14
(D) *Combination of alternative approaches*									
Panel A	−0.30	0.02	−0.66	0.43	−0.32	0.05	−0.51	−0.61	−0.36
Panel B	0.07	0.14	0.03	0.30	−0.17	−0.06	−0.56	−0.41	−0.34
Panel C	0.59	0.72	0.66	0.18	0.03	0.00	−0.15	0.26	−0.11

Note: This table shows the correlation between disagreement and macroeconomic variables. STDEV is the cross-sectional standard deviation in point forecasts truncating at −30%/30%. IQR is the interquartile range in point forecasts. $STDEV_{[-5,30]}$ is the cross-sectional standard deviation in point forecasts truncating at −5%/30%. For definitions of other variables, see Table II.

Copyright © 2015 John Wiley & Sons, Ltd

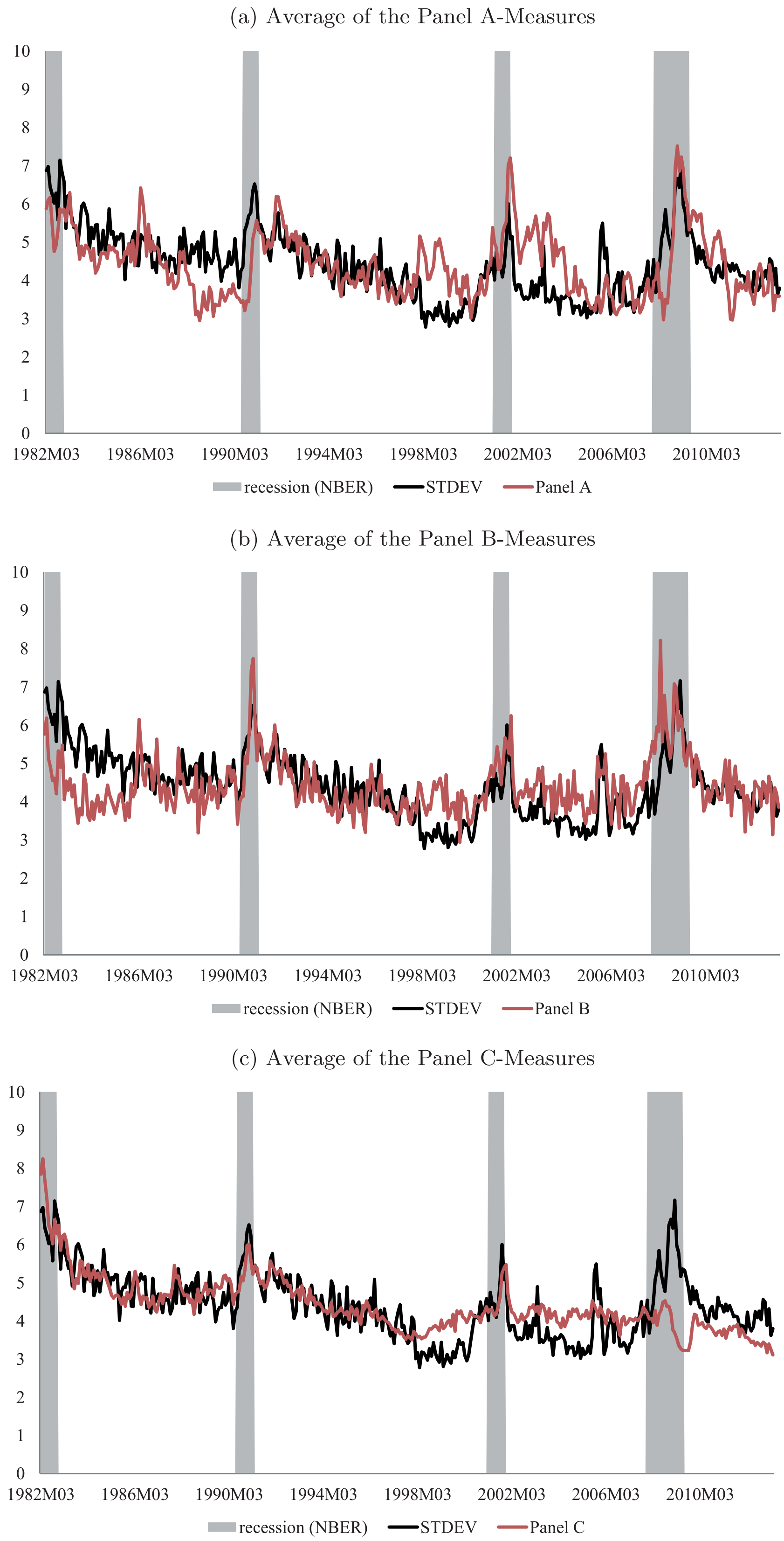

Figure A.1. Disagreement and the business cycle: US short-run inflation expectations

Copyright © 2015 John Wiley & Sons, Ltd

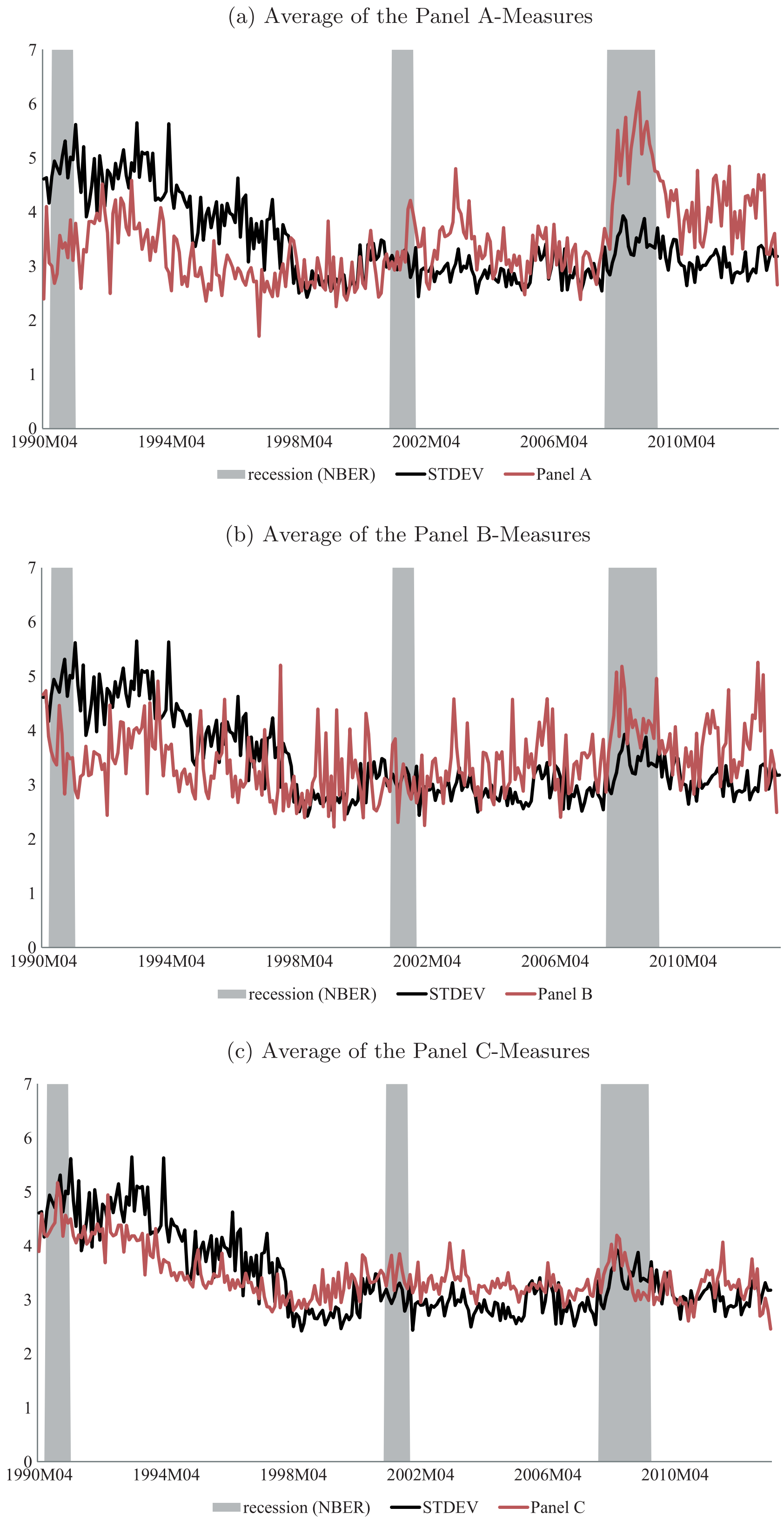

Figure A.2. Disagreement and the business cycle: US long-run inflation expectations

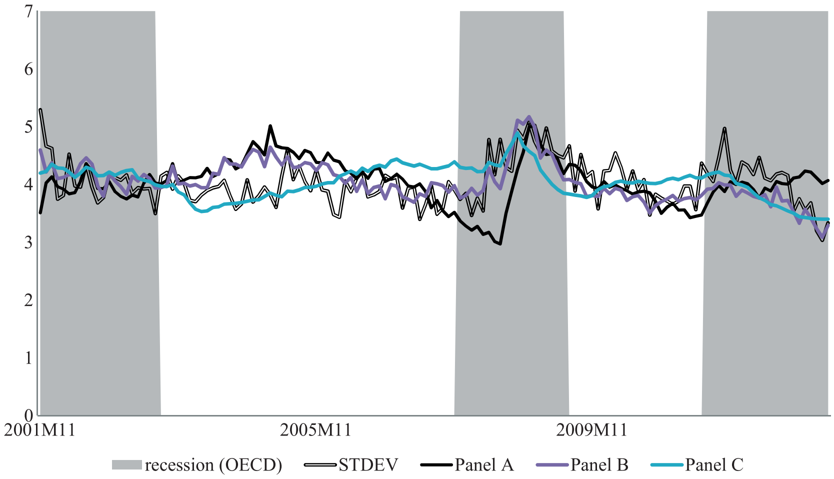

Figure A.3. Disagreement and the business cycle: Swedish short-run inflation expectations

Copyright © 2015 John Wiley & Sons, Ltd