Discussion Papers in Economics

No. 05/2022

Extending the Procedure of Engelberg et al. (2009) to Surveys with Varying Interval-Widths

Christoph Becker
Heidelberg University

Peter Dürsch
University of Mannheim

Thomas Eife
Heidelberg University

Alexander Glas
FAU Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/discussion-papers/

Extending the Procedure of Engelberg et al. (2009) to Surveys with Varying Interval-Widths*

Christoph Becker[1], Peter Duersch[2], Thomas Eife[1] and Alexander Glas[3]

[1]Heidelberg University
[2]University of Mannheim
[3]Friedrich-Alexander-Universität Erlangen-Nürnberg

August 12, 2022

Abstract

The approach by Engelberg, Manski, and Williams (2009) to convert probabilistic survey responses into continuous probability distributions implicitly assumes that the question intervals are equally wide. Almost all recently established household surveys have intervals of varying widths. Applying the standard approach to surveys with varying widths gives implausible and potentially misleading results. This note shows how the approach of Engelberg et al. (2009) can be adjusted to account for intervals of unequal width.

JEL Codes: C18, C82, C83
Keywords: Survey methods, probabilistic questions, density forecasts.

1 Introduction

One of the less expected consequences of the long period of low interest, that started with the financial crisis of 2008/09 in the U.S. and other countries, is a vast supply of new survey data. With one of their main policy tools incapacitated, central banks try to 'guide' households' and firms' expectations about future interest rates and future inflation (forward guidance). An important part of this management of expectations is their measurement, which is typically done via large representative surveys. Following

*We would like to thank Christian Conrad, Timo Dimitriadis, Matthias Hartmann, and Manuel Schick, as well as seminar participants in Heidelberg for helpful comments. E-mail addresses: christoph.becker@awi.uni-heidelberg.de, duersch@xeeron.de, thomas.eife@gmail.com (corresponding author), alexander.glas@fau.de.

the lead of the Survey of Consumer Expectations (SCE), which was established by the Federal Reserve Bank of New York in 2013, numerous other central banks recently initiated surveys that include probabilistic (density) forecasts.[1] In probabilistic surveys, respondents are shown a number of intervals and are then asked to assign probabilities to all intervals that represent their beliefs about the expected outcome of a variable (e.g., inflation rate). It is standard to leave the two outermost intervals open ended. A key advantage of probabilistic forecasts over point forecasts is that they allow survey participants to express their subjective uncertainty about future outcomes (Manski, 2004).

The important contribution of Engelberg, Manski, and Williams (2009, from here on EMW) was to show how to turn the survey data into rigorous measurements of a respondent's subjective probability distribution. The procedure provides a parametric distribution from which important statistics (e.g., subjective measures of location, spread, or tail risk) may be computed. The original methodology implicitly assumes that all intervals are equally wide. EMW analyze inflation and real GDP growth expectations from the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia where this assumption is satisfied. However, most of the recently established household and firm surveys include probabilistic questions with varying interval width, as do some expert surveys such as the ECB-SPF (see the example in Figure 2 below).

When applied to surveys with varying interval widths, EMW's original approach frequently assigns probabilities to events that the respondent explicitly ruled out, leading to inflated measures of a respondent's uncertainty, and to systematically overestimated tail risks. This note shows how to extend EMW's original procedure to surveys with unequal interval widths. In addition, the modified procedure provides more flexibility when respondents assign positive probabilities to one of the two open intervals.

[1] Examples are the forthcoming Consumer Expectations Survey conducted by the European Central Bank (which will survey representative samples of households in all countries of the euro area) and similar surveys conducted by the central banks of Canada, France, Germany, the Netherlands, Ukraine, and the United Kingdom.

2 Modified Procedure for Intervals with Unequal Widths

EMW propose to fit a continuous distribution to the probabilities assigned by the respondents. EMW's choice of the continuous distribution depends on the number of intervals a respondent uses. When a respondent places positive probabilities in three or more intervals, EMW suggest to fit a generalized Beta distribution. When a respondent uses one or two intervals, the suggested distribution is isosceles triangular.

The original procedure requires two modifications when the widths of the intervals vary. First, we need to distinguish between the probability a respondent assigns to an interval and the corresponding density of the interval. Second, when fitting triangular distributions, we need to modify the original procedure because EMW's underlying assumptions are compatible only when all intervals are equally wide. Section 2.1 describes the modifications for the triangular distribution.

When a respondent assigns positive probabilities to three or more intervals, we follow EMW and assume that the subjective distribution takes the form of a generalized Beta distribution. The technical details are described in EMW, but the distinction between probabilities and densities requires some attention since the bar chart of the assigned probabilities does not necessarily describe a proper histogram. Consider the following fictional example where a respondent reports a 0.2 chance that inflation will be in the interval $[1\%, 2\%)$, a 0.35 chance for the interval $[2\%, 4\%)$, and a 0.45 chance for the interval $[4\%, 6\%)$. Following EMW, we can infer these points on the respondent's PDF: $f(0.01) = 0$, $f(0.02) = 0.2$, $f(0.04) = 0.175$, and $f(0.06) = 0.225$; and on the respondent's CDF: $F(0.01) = 0$, $F(0.02) = 0.2$, $F(0.04) = 0.55$, and $F(0.06) = 1$. In this example, the respondent provides "single-peaked" probabilities of 0.2, 0.35, and 0.45 but the corresponding densities are bimodal (0.2, 0.175, and 0.225). With two types of responses (probabilities and densities), we have to reconsider EMW's requirement that all responses be unimodal. We continue this discussion in Section 2.2 when we discuss non-standard scenarios.

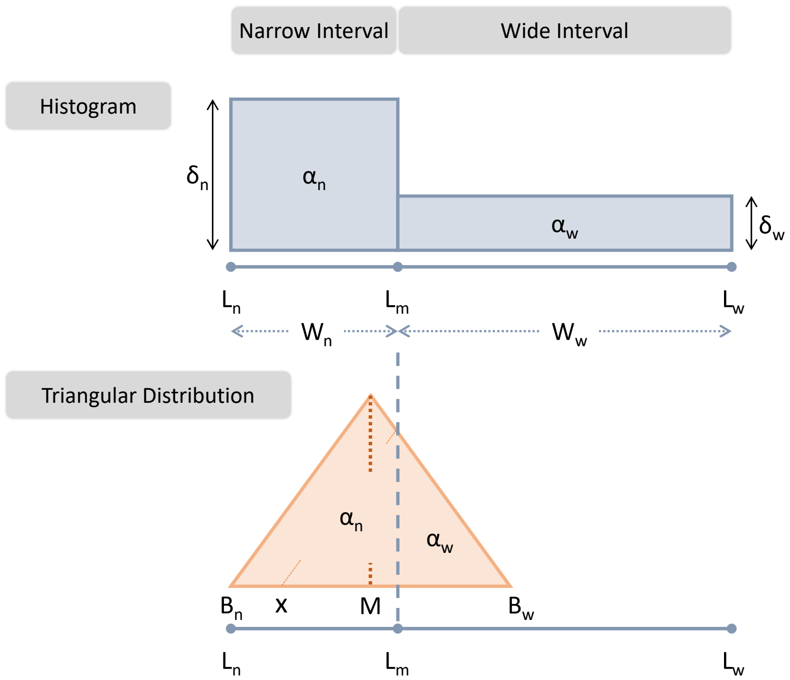

Figure 1: Constructing a triangular subjective probability distribution from survey responses when the intervals have different widths. The figure assumes that $\alpha_n > \alpha_w$ and $\delta_n > \delta_w$ corresponding to Column 4 in Table 1.

2.1 Modified Procedure For One or Two Intervals

EMW use the following four assumptions when a respondent assigns positive probabilities to one or two (adjacent) intervals.

1. The distribution takes the form of an *isosceles* triangle.
2. The support of the subjective distribution contains the entirety of the interval with the higher probability.
3. The triangle's support does not extent outward beyond the limits of the intervals.[2]
4. The border between the two intervals splits the triangle into two parts whose areas correspond to the probability masses assigned by the respondent.

When faced with intervals of unequal width, we first have to replace the term 'probability' in assumption 2 by 'density'.[3] Moreover, the four assumptions are compatible with each other only if we assume intervals of equal width. For surveys with unequal interval widths, one of the assumptions has to be relaxed. Here we suggest to relax the first assumption and to drop the requirement that the triangle is isosceles. EMW's

[2]This assumption is only implicit in EMW but follows directly from their equations when the intervals are equally wide.

[3]When the intervals are equally wide, as in EMW, both formulations are equivalent.

(1)	(2)	(3)	(4)
$\boldsymbol{\alpha_n \in [0, 2s_n^2]}$	$\boldsymbol{\alpha_n \in [2s_n^2, s_n]}$	$\boldsymbol{\alpha_n \in [s_n, 1/2]}$	$\boldsymbol{\alpha_n \in [1/2, 1]}$
$B_n = L_m - \frac{\sqrt{\frac{\alpha_n}{2}}}{1-\sqrt{\frac{\alpha_n}{2}}} (L_w - L_m)$	$B_n = L_n$	$B_n = L_n$	$B_n = L_n$
$M = \frac{B_w+B_n}{2}$	$M = L_n + \frac{(L_n-L_m)^2}{(L_w-L_n)\alpha_n}$	$M = L_m$	$M = \frac{B_w+B_n}{2}$
$B_w = L_w$	$B_w = L_w$	$B_w = L_m + (L_m - L_n) \frac{1-\alpha_n}{\alpha_n}$	$B_w = L_m + (L_m - L_n) \frac{\sqrt{\frac{1-\alpha_n}{2}}}{1-\sqrt{\frac{1-\alpha_n}{2}}}$
$\delta_n \leq \delta_w$	$\delta_n \leq \delta_w$	$\delta_n \geq \delta_w$	$\delta_n \geq \delta_w$

Table 1: The modified procedure for fitting a triangular distribution when a respondent assigns probabilities to one or two (adjacent) intervals. The probability of the narrow interval, α_n, may fall into four regions whose size depends on s_n, the relative size of the narrow interval. When both intervals are equally wide, $2s_n^2 = s_n = 1/2$ and the procedure reduces to the original procedure proposed in EMW.

procedure is a special case of the modified procedure when the intervals are equally wide.

The triangular distribution is completely described by its mode and its support. Consider two adjacent intervals of possibly unequal width. A subscript n refers to the narrow interval and a subscript w to the wide interval. Let L_n (L_w) denote the outer limit of the narrow (wide) interval and L_m the common limit of the two intervals. Figure 1 depicts the situation where the left interval is narrow, i.e., $L_n < L_m < L_w$. With α_n (α_w) denoting the probability assigned to the narrow (wide) interval, we can express the densities as

$$\delta_n = \frac{\alpha_n}{W_n} \text{ and } \delta_w = \frac{\alpha_w}{W_w}, \tag{1}$$

where $W_n = |L_n - L_m|$ and $W_w = |L_w - L_m|$.

Let M be the mode of the triangular distribution and B_n and B_w the limits of the triangle's support. Since the triangle's area equals one, the points B_n, M, and B_w fully describe the shape of the distribution and define mean ($[B_n + M + B_w]/3$), mode (M), variance ($[B_n (B_n - B_w) + M (M - B_n) + B_w (B_w - M)]/18$) and any other relevant statistics of the distribution.

Table 1 shows how to calculate B_n, M, and B_w given α_n and given the limits of the intervals L_n, L_w, and L_m. The equations in the table hold independently of whether the left interval is wide or narrow. The parameter s_n denotes the relative size of the narrow interval, $W_n / (W_n + W_w)$.[4]

[4]See the Appendix for a detailed derivation of the equations in Table 1.

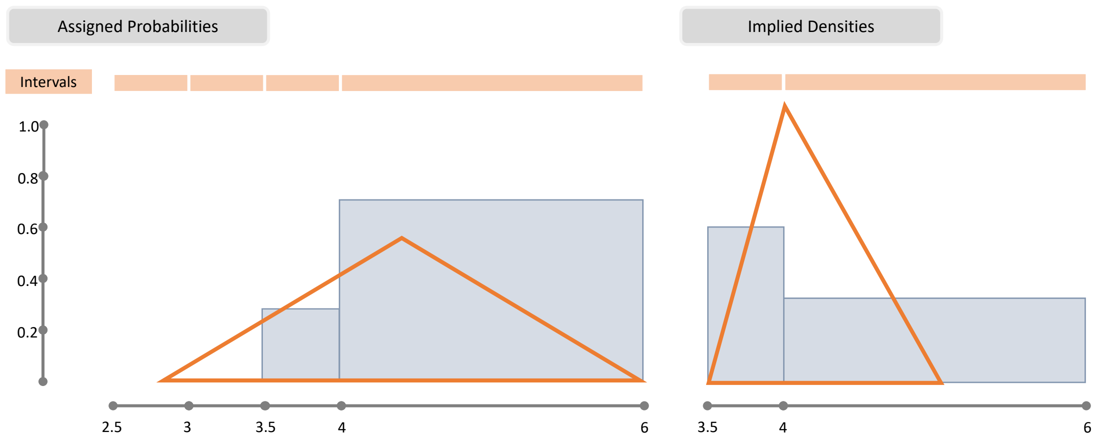

Figure 2: Fitting an isosceles triangle to the response of a respondent who assigns a probability of 30 percent to the interval $[3.5\%, 4\%)$ and a probability of 70 percent to the interval $[4\%, 6\%)$. Left Panel: Original procedure with mean of 4.37 and variance of 1.44. Right Panel: Modified procedure with mean of 4.22 and variance of 0.77.

Columns 1 and 4 in Table 1 correspond to the setting in EMW where the distribution takes the form of an isosceles triangle. The interior cases of Columns 2 and 3 cannot occur when both intervals are equally wide since then $2s_n^2 = s_n = 1/2$. In Column 2, the triangle covers the entire support of the two intervals and, as α_n increases, the mode moves towards L_m to assure that the two parts of the triangle to the left and to the right of M correspond to the probabilities assigned by the respondent (assumption 4). M is bounded by L_m. In Column 3, $\alpha_n < \alpha_w$ but the ordering of the intervals is reversed for the densities, i.e., $\delta_n > \delta_w$. This is the situation illustrated in the following example.

In September 2020, a professional forecaster of the ECB-SPF survey reported expectations of future GDP growth to fall into the interval $[3.5\%, 4\%)$ with a probability of 30 percent and into the interval $[4\%, 6\%)$ with a probability of 70 percent.[5] The original procedure calls for an isosceles triangle whose support "contains the entirety of the more probable" (here: right) interval. This fixes one endpoint of the support. The other endpoint of the support is found by constructing an isosceles triangle in such a way that assumption 4 is met (see the left hand panel of Figure 2). The result is an im-

[5]The ten interior intervals of the original ECB-SPF survey were equally-spaced with a width of half a percentage point. Anticipating sharp swings in forecasters' expectations because of the Covid-19 pandemic, the ECB-SPF appended several intervals for the GDP growth expectations in 2020Q2, each four times as wide as the original intervals.

plausibly wide triangle whose support exceeds the support of the respondent's original response, violating assumption 3. That is, the procedure assigns positive probabilities to events that the respondent explicitly ruled out, placing positive probability in the intervals $[2.5\%, 3\%)$ and $[3\%, 3.5\%)$ despite the respondent stating a probability of 0 in both.

Armantier, Topa, Van der Klaauw, and Zafar (2017, footnote 28), the operators of the SCE, suggest to alleviate this problem by replacing the threshold of 50:50 of the original procedure with a conditional threshold of 40:60, where the support of the triangle is assumed to include the entire narrow interval if its probability exceeds 40 percent and includes the entire wide interval otherwise. While this somewhat arbitrary threshold may ease the problem (though not in the example shown in Figure 2), it does not assure that assumption 3 is met. Their suggestion points, however, in the right direction: The decision whether an interval is entirely covered by the triangle should be based on a combination of probabilities and interval widths. The appropriate combinations are the intervals' densities (equation 1). In order to satisfy assumptions 2, 3, and 4, we have to allow for general triangular distributions. In the example, this means that the mode of the triangle is slightly tilted towards the narrow interval (see right hand panel of Figure 2). This has a strong intuitive appeal: First, the triangular distribution in the right panel of Figure 2 assigns a higher probability to the interval $[4\%; 6\%)$, congruent with the respondent's estimate. Second, the tilted mode of the distribution takes into account that the interval $[3.5\%, 4\%)$ has a higher density indicating that the respondent considers events in this interval to occur with a higher likelihood.

2.2 Non-Standard Scenarios

There are two non-standard scenarios that require some discussion. First, respondents may place positive probabilities in one (or both) of the two open intervals. The parametric approach in EMW requires a closure of the open intervals. When a respondent uses three or more intervals, EMW's suggestion to determine the limit(s) of the generalized Beta distribution endogenously is sensible independently of whether one is faced with intervals of equal or unequal widths. When a respondent uses two intervals, EMW's original procedure forces the researcher to assume that the open interval has the same width as the adjacent interval. The modified procedure proposed here provides more flexibility in this case and allows the researcher to specify any width that the researcher deems appropriate.

Second, EMW require responses to be unimodal. Especially in household surveys, unimodality is frequently violated and care is necessary in order not to discard valuable information. In addition, with two types of responses (densities and probabilities) it is a priori not clear whether one should require both the probabilities and the densities to be single peaked or whether unimodal probabilities but bimodal densities merit an exclusion of the response. One way to proceed is to assume unimodality only in case of one or two (adjacent) intervals and to take a more flexible approach for the Beta distribution, dropping the constraint that the shape parameters of the Beta distribution have to be greater than one as is done in Armantier et al. (2017). There are situations where bimodal responses may be rational and, especially for household surveys, the additional flexibility seems warranted.

3 Conclusion

The procedure of EMW to convert probabilistic survey responses into continuous distributions has become a standard tool in the analysis of survey data. The original procedure implictly assumes intervals of equal width. When applied to surveys with intervals of varying widths, the procedure may assign positive probabilities to events that the respondent has explicitly ruled out, leading to systematically higher measures of respondents' uncertainty. In addition, intervals of varying width require us to differentiate between the probabilities a respondent assigns to the intervals and the implied densities. Since all major surveys are constructed such that the outer intervals are wider than the interior intervals, working with probabilities rather than densities may systematically overweight tail risks.

This note extends the procedure to allow for survey questions with varying interval widths. At the time when EMW proposed their procedure, varying interval widths were much less common than they are today. The proposed modification is natural and necessary given the presence of varying interval widths in almost all recently established surveys.

References

Armantier, O., G. Topa, W. Van der Klaauw, and B. Zafar (2017). An overview of the survey of consumer expectations. *Economic Policy Review* (23-2), 51–72.

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

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

Appendix

This Appendix derives the equations in Table 1.

Columns 1 and 4 In Columns 1 and 4, the triangular distribution is isosceles and the procedure corresponds to theoriginal in EMW. We derive the equations for Column 4. The derivation for Column 1 is similar.

In Column 4, $B_n = L_n$ and M is the midpoint between B_n and B_w, which can be derived by noting that the triangle with support $|x, B_w|$ in Panel A of Figure A.1 is congruent to the triangle of interest (with support $|B_n, B_w|$). Given the triangles' areas ($2\alpha_w$ and 1, respectively), we can express their heights as $\frac{2\alpha_w}{|L_m - B_w|}$ and $\frac{2}{|L_n - B_w|}$. Congruence implies that the ratio of the triangles' heights equals the ratio of the triangles' supports. That is,

$$\frac{\frac{2\alpha_w}{|L_m - B_w|}}{2\,|L_m - B_w|} = \frac{\frac{2}{|L_n - B_w|}}{|L_n - B_w|}$$

so that, using $1 = \alpha_n + \alpha_w$, we have

$$\sqrt{\frac{1-\alpha_n}{2}}\,(L_n - B_w) = (L_m - B_w)\,.$$

Solving for B_w gives

$$B_w = \frac{1}{1-\sqrt{\frac{1-\alpha_n}{2}}}L_m - \frac{\sqrt{\frac{1-\alpha_n}{2}}}{1-\sqrt{\frac{1-\alpha_n}{2}}}L_n = L_m + (L_m - L_n)\frac{\sqrt{\frac{1-\alpha_n}{2}}}{1-\sqrt{\frac{1-\alpha_n}{2}}}$$

as stated in Table 1.

Column 2 The triangle in Column 2 covers the entire support of the two intervals so that $B_n = L_n$ and $B_w = L_w$. See panel C of Figure A.1. We determine M by taking advantage of the congruence of two rectangular triangles: First, the triangle with support $|L_m - L_n|$ whose area equals α_n, and whose height is, therefore, given by $\frac{2\alpha_n}{|L_m - L_n|}$. Second, the triangle with support $|M - L_n|$ whose height is equal the height of the large triangle with area 1. This height equals $\frac{2}{|L_w - L_n|}$. Congruence implies that

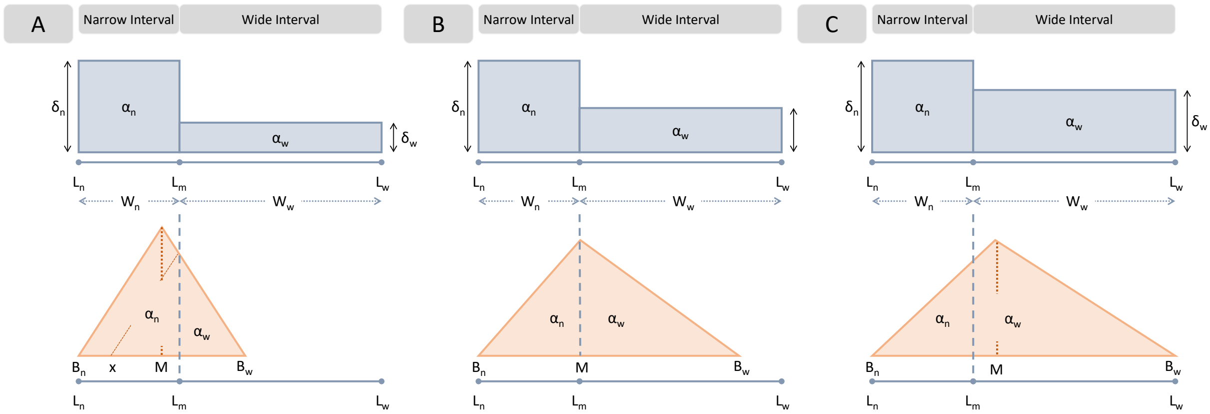

Figure A.1: Constructing a triangular subjective probability distribution from survey responses when the intervals have different widths. Panel A corresponds to Column 4, panel B to Column 3, and panel C to Column 2 in the table. The derivation of the equations in Column 1 is similar to the derivation of the equations in Column 4.

the ratio of the triangles' heights equals the ratio of the triangles' supports. That is,

$$\frac{\frac{2\alpha_n}{|L_m - L_n|}}{|L_m - L_n|} = \frac{\frac{2}{|L_w - L_n|}}{|M - L_n|}$$

Solving for M, we have

$$M = L_n + \frac{1}{\alpha_n} \frac{(L_n - L_m)^2}{(L_w - L_n)}$$

as stated in Table 1.

Column 3 In Column 3, $B_n = L_n$ and $M = L_m$. In order to determine B_w, consider the two rectangular triangles to the right and the left of L_m in panel B of Figure A.1. The first has support $|L_m - L_n|$ and area α_n, the second has support $|B_w - L_m|$ and area α_w. Both have the same height $\frac{2\alpha_n}{|L_m - L_n|} = \frac{2\alpha_w}{|B_w - L_m|}$. Solving this equality for B_w, we have

$$B_w = L_m + \frac{1 - \alpha_n}{\alpha_n} (L_m - L_n)$$

as stated in Table 1.