HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Extending the Procedure of Engelberg et al. (2009) to surveys with varying interval-widths

Christoph Becker

Peter Duersch

Thomas Eife

Alexander Glas

AWI DISCUSSION PAPER SERIES NO. 707
November 2021

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]FAU Erlangen-Nürnberg

November 29, 2021

Abstract

The approach by Engelberg, Manski, and Williams (2009) to convert probabilistic survey responses into continuous probability distribution requires 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.

1 Introduction

One of the less expected consequences of the current spell of low interest rates 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 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

*We would like to thank Christian Conrad, Timo Dimitriadis, 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.

probabilistic questions.[1] In probabilistic surveys, respondents are shown a number of intervals and are then asked to attach probabilities to all intervals that represent their beliefs about some future outcome (e.g., the inflation rate). It is standard to leave the two outermost intervals open ended. Probabilistic questions have the advantage that they allow survey participants to express the uncertainty they feel when answering the question (Manski, 2004).

The important contribution of Engelberg, Manski, and Williams (2009, EMW) was to show how to turn the survey data into rigorous measurements of a respondent's subjective probability distribution. The procedure provides a full analytical distribution from which important statistics (e.g., subjective measures of location, spread, or tail risk) may be computed. The original methodology requires that all intervals are equally wide. EMW analyze inflation and real GDP growth expectations from the Survey of Professional Forecasters (SPF) operated by the Federal Reserve Bank of Philadelphia in which this assumption is satisfied. However, most of the recently established household and firm surveys include probabilistic questions with varying interval width. In this note, we show (i) that applying the original procedure to surveys with varying widths gives implausible and potentially misleading results and (ii) that the original procedure can easily be modified to allow for intervals of unequal widths. We begin with a description of the original procedure before turning to examples of intervals whose widths vary.

2 The Original Procedure (assuming equal widths)

The survey responses (i.e., the probabilities assigned by the respondents to the distinct outcome intervals) are usually assumed to be discrete representations of the respondent's subjective probability distribution. Following earlier work by Dominitz and Manski (1997), EMW propose to fit a continuous distribution to the probabilities.

[1]Examples are the forthcoming Consumer Expectations Survey conducted by the European Central Bank (which will survey representative samples 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.

Their choice of the continuous distribution depends on the number of intervals the respondent uses. When a respondent assigns a positive probability to one or two intervals, the underlying distribution is assumed to have the shape of an isosceles triangle. When positive probability is assigned to three or more intervals, the procedure assumes an underlying Beta distribution. Unimodality of the responses is assumed throughout. Section 2.3 addresses the unimodality assumption in more detail.

2.1 One or Two Intervals

Panels I and II of Figure 1 illustrate the procedure for a respondent who uses two adjacent intervals $[L, M]$ and $[M, R]$ with $L < M < R$.[2] In line with EMW, the intervals are equally wide, i.e., $R - M$ equals $M - L$. In this exampe, the respondent considers the right interval more probable. When a forecaster places more probability mass in one interval than the other, the procedure assumes that the support of the triangle contains the entirety of the more probable interval. This restricts one endpoint of the support of the triangle. The other endpoint is determined by the assumption that the triangle is isosceles and that the areas of the triangle to the left and right of M match the probability masses originally assigned by the forecaster.

Formally, let $[A, B]$ be the support of such a triangle. Since their area equals one by definition, isosceles triangular distributions are completely characterized by their support. The mode of the distribution is located at $C = (A + B)/2$. Let α denote the probability mass the respondent assigns to the left interval. Depending on whether $\alpha \lesseqgtr 1/2$, there are two cases. When $\alpha < 1/2$ (as in Figure 1), B coincides with R, and A is adjusted according to

$$A = M - \frac{\sqrt{\frac{\alpha}{2}}}{1-\sqrt{\frac{\alpha}{2}}} (R - M) . \qquad (1)$$

[2]Many surveys do not explicitly specify the inclusion (or exclusion) of the interval limits so that, say, an expected inflation rate of M percent may either be assigned to the left or right interval.

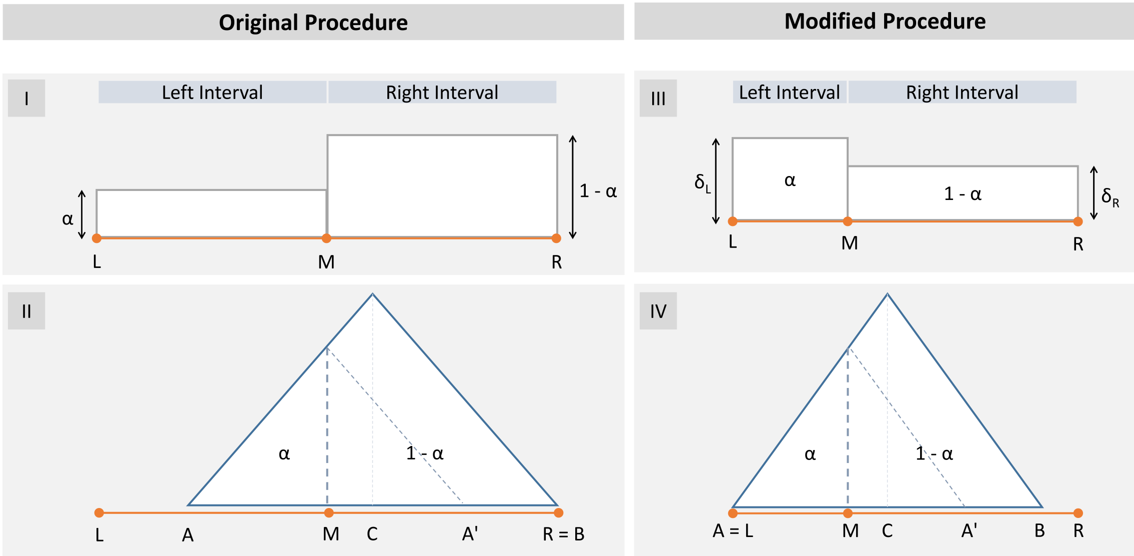

Figure 1: Isosceles Triangles. Panels I and II illustrate the original procedure where the intervals are assumed equally wide. Given $\alpha < 1/2$, the triangle in Panel II covers the right interval entirely and the left interval partially. Panels III and IV illustrate the modified procedure that allows for intervals of unequal width. The illustration assumes $\alpha < 1/2$, but since the wide right interval has a lower density, the procedure fixes $A = L$ and adjusts B. With $\alpha < 1/2$, the gravity center of the triangle remains in the right interval.

When $\alpha > 1/2$, A coincides with L, and B is adjusted according to

$$B = M + \frac{\sqrt{\frac{(1-\alpha)}{2}}}{1-\sqrt{\frac{(1-\alpha)}{2}}}\left(M - L\right). \tag{2}$$

When $\alpha = 1/2$, equations (1) and (2) imply that the support of the triangle spans the entire support of the two intervals $[L, R]$.[3] These triangular distributions match the respondent's answer in the sense that $\mathcal{T}\left([A, M]\right) = \alpha$ and $\mathcal{T}\left([M, B]\right) = 1 - \alpha$, where the notation $\mathcal{T}\left(I\right)$ indicates the probability mass assigned to interval I.[4]

[3]When a respondent uses only a single interval, such that $\alpha = 1$, equation (2) implies that $B = M$ so that the support of the triangle spans the entire support of this interval.

[4]Equation (1) can be derived by noting that the triangle with support $[A, A']$ in Figure 1 is congruent to the triangle of interest (with support $[A, R]$). Given the triangles' areas (2α and 1, respectively), we can express their heights as $\frac{2\alpha}{M-A}$ and $\frac{1}{R-A}$. Congruence implies that the ratio of the triangles' heights equals the ratio of the triangles' supports. Simplifying and rearranging this relationship gives equation (1). The derivation of equation (2) proceeds accordingly.

2.2 Three or More Intervals

When a respondent uses three or more intervals, EMW propose to fit a generalized Beta distribution. Instead of the limits 0 and 1 of the regular Beta distribution, the generalized Beta distribution has a flexible support. Let A and B denote the two location parameters and a and b denote the two shape parameters of the Beta distribution whose CDF is given by

$$F_{gBeta}\left(t;a,b,A,B\right)=\begin{cases} 0 & t\leq A \\ \frac{1}{\mathcal{B}(a,b)}\int_L^t \frac{(x-L)^{a-1}(R-x)^{b-1}}{(R-L)^{a+b-1}}dx & l<t\leq B \\ 1 & t>B, \end{cases} \tag{3}$$

where $\mathcal{B}\left(a,b\right)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma\left(a\right)=\int_0^\infty x^{a-1}\exp\left(-u\right)du$.

In order to fit the distribution in equation (3) to the array of probabilities assigned by the respondent, EMW fix the limits A and B to the endpoints of the intervals with positive probability. That is, A is the left limit of the respondent's leftmost interval and B is the right limit of the respondent's rightmost interval. The minimization problem in this case is

$$\min_{a>1,\ b>1}\sum_{k=1}^{K}\left[F_{gBeta}\left(t_k;a,b\right)-P_k\right]^2$$

where t_k is the right endpoint of the k^{th} interval and $P_k=\sum_{i=1}^{k}p_i$ is the cumulative probability mass assigned to the first k intervals. EMW impose the constraints that $a>1$ and $b>1$ to ensure unimodality. When a respondent uses one or both of the two open intervals, EMW suggest to treat the limits (A and/or B) as parameters to be estimated in the minimization problem with an additional restriction that assures that both location parameters are within reasonable bounds.[5]

[5] In the original paper, EMW assume that A and B lie within the most extreme values of inflation and GDP growth that have actually occured in the United States since 1930.

2.3 Non-Standard Scenarios

There are two non-standard scenarios that are not addressed by the procedure above. First, respondents may occasionally provide responses that violate the unimodality assumption. In expert surveys, such as the SPF which is analyzed in EMW, this scenario is rarely observed and EMW suggest to simply discard these observations. In contrast, responses with multiple peaks are observed more frequently in household surveys and additional care is necessary in order not to discard valuable information. One way to proceed is to drop the constraints that the shape parameters a and b have to be greater than one and thus to allow for a U-shaped Beta distribution (as it is done in the SCE as described in Armantier, Topa, Van der Klaauw, and Zafar (2017)).

Second, a respondent may choose two intervals of which one is open. EMW do not address this scenario as it is rarely observed in expert surveys, but it occurs more frequently in household surveys. One way to proceed is to close the open intervals and to assume that they are equally wide (or twice as wide) as the adjacent intervals. The case where a respondent assigns a positive probability to three or more intervals, of which one is open ended is addressed in EMW by optimizing over these limits as explained in the previous section.

2.4 Applying the Original Procedure to Intervals with Varying Widths

The example shown in the left panel of Figure 2 illustrates one of the problems that may arise when the original procedure is applied to intervals with varying width. In the example, which is taken from the ECB's SPF, the respondent expects 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.[6] Since $\alpha < 1/2$, the original procedure suggests to fit a triangle distribution whose left limit is specified by equation (1). The result is an implausibly wide triangle whose support exceeds the support of the

[6]In response to the sharp decline in economic activity caused by the Covid-19 pandemic, the ECB-SPF introduced outcome intervals for GDP growth with unequal width in 2020Q2. Up until that point, all intervals in the SPF data used to have equal width.

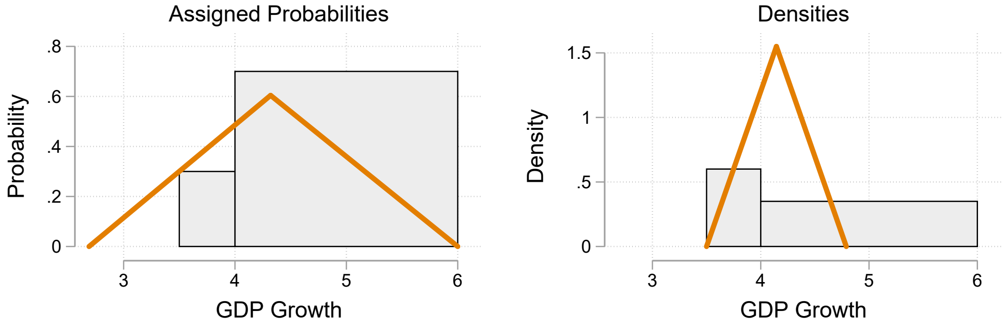

Figure 2: Fitting an isosceles triangle to the response of a forecaster who assigns a probability of 30 percent to the interval $[3.5, 4]$ and a probability of 70 percent to the interval $[4, 6]$. Left Panel: Original procedure. Right Panel: Modified procedure.

forecaster's original response. That is, the procedure assigns positive probabilities to events that the forecaster explicitly ruled out.

Armantier et al. (2017), the operators of the SCE, suggest to alleviate this problem by replacing the threshold of $\alpha = 0.5$ with $\alpha = 0.4$ in such cases. While this may ease the problem somewhat, the new threshold is essentially arbitrary. More importantly, with this threshold, one may still assign positive probabilities to events that the forecaster explicitly ruled out as in the case of Figure 2.

3 The Modified Procedure

In this section, we show how to adjust the original methodology of EMW in the presence of unequal interval widths. In this case, the relevant measures of a respondent's subjective probability distribution are the densities implied by the assigned probabilities, not the probabilities themselves.

3.1 One or Two Intervals

When a respondent assigns positive probabilities to two adjacent intervals of unequal width, the decision whether to adjust the left or the right limit of the triangle's support should be based on the corresponding densities (denoted as δ_L and δ_R). The probability α is still relevant as it determines the location of the triangle's gravity center. This

Case	**Density**	**Probability**	
(1a)	$\delta_L < \delta_R$	$\alpha < \frac{1}{2}$	$A = M - \frac{\sqrt{\frac{\alpha}{2}}}{1-\sqrt{\frac{\alpha}{2}}}(R - M)$
(1b)	$\delta_L < \delta_R$	$\alpha > \frac{1}{2}$	$A = M - \frac{1-\sqrt{\frac{1-\alpha}{2}}}{\sqrt{\frac{1-\alpha}{2}}}(R - M)$
(2a)	$\delta_L > \delta_R$	$\alpha > \frac{1}{2}$	$B = M + \frac{\sqrt{\frac{1-\alpha}{2}}}{1-\sqrt{\frac{1-\alpha}{2}}}(M - L)$
(2b)	$\delta_L > \delta_R$	$\alpha < \frac{1}{2}$	$B = M + \frac{1-\sqrt{\frac{\alpha}{2}}}{\sqrt{\frac{\alpha}{2}}}(M - L)$

Table 1: Adjustments of the support of the triangle. The densities determine whether the right or the left limit of the support is adjusted. The probabilities determine the location of the center of the triangle. Cases (1a) and (2a) correspond to the original procedure in Engelberg et al. (2009).

is illustrated in Figure 1. When the two intervals are equally wide (Panels I and II), the gravity center, which is above C, is always located in the interval that is entirely included in the support of the triangle (the right interval in the figure). Panels III and IV show that this may no longer be the case when the intervals have different widths. Panels III and IV illustrate the response of a forecaster who assigns a probability $\alpha < 1/2$ to the narrow (left) interval. The corresponding density of the narrow interval is, however, higher than the density of the wide interval (i.e., $\delta_L > \delta_R$), suggesting that the support of the subjective distribution should contain the entirety of the left interval. Nonetheless, despite assigning a lower density to the right interval, the respondent still considers the (entire) right interval more likely than the (entire) left interval, so that the gravity center of the triangle should be located in the right interval. Since both the densities and the probabilities determine the shape of the triangle, we have to distinguish between four cases.

Table 1 shows the adjustments of the modified procedure. There are four cases, depending on the relative densities and on whether the probability α is smaller or larger than 1/2. The densities and the probabilities are sufficient to jointly determine the appropriate shape of the triangle. When the intervals are equally wide, cases (1b) and (2b) cannot occur so that the original procedure (cases 1a and 2a) is contained as a special case of the modified procedure. The expressions for A and B in these cases correspond to the expressions in the original procedure of EMW (equations (1) and (2) above). When the right interval is larger than the left, case ($1b$) cannot occur. When the left interval is larger, case ($2b$) cannot occur. The derivation of the four expressions follows the steps outlined in footnote 4.

3.2 Three or More Intervals

When a respondent assigns positive probabilities to three or more intervals of equal width, fitting a Beta distribution directly to the assigned probabilities is incorrect as the probabilities do not describe a proper histogram. While this does not affect the support of the Beta distribution as it does with the triangular distributions, the estimated shape parameters (and thus the moments of the distribution) may not be appropriate. The extent of this misspecification is ultimately an empirical question. To avoid this problem, the Beta distribution should be fitted to the densities that are implied by the respondent's probabilities. The rest of the procedure follows the steps outlined in Section 2.2.

3.3 Non-Standard Scenarios

Apart from the non-standard scenarios discussed in Section 2.3, there is an additional point one has to consider. EMW suggest to restrict the analysis to unimodal (single peaked) responses. With two types of responses (densities and probabilities) we have to reconsider this restriction. A priori, it is 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. The answer to these questions will depend on the research question at hand and will require a careful consideration of the consequences. One way to proceed is to discard the unimodality requirement altogether, as it is suggested in Armantier et al. (2017), and allow for U-shaped responses.

4 Conclusion

The procedure of Engelberg, Manski, and Williams (2009) to convert probabilistic survey responses into continous distributions has become a standard tool in the analysis of survey data. This note extends the procedure to allow for survey questions with varying interval widths. At the time when Engelberg et al. (2009) proposed their procedure, varying interval widths were much less common than they are today. Given the presence of varying interval widths in most household and firm surveys, the proposed modification seems to be both necessary and natural. Finally, it should be noted that applying the original procedure of Engelberg et al. (2009) to intervals with varying widths may introduce systematic biases. Most survey questions are constructed such that the outer intervals are wider than the center intervals. By working with the probabilities rather than the implied densities, the procedure systematically assigns a higher weight on the outer intervals and therefore overweights tail risks.

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.

Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association 92*(439), 855–867.

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.