Appendix: Why Use a Diffusion Index?

Economic Brief No. 22-22, June 22, 2022

This appendix provides a deeper dive into the methodology of diffusion indexes. A diffusion index is a statistic used to report on multiple-choice survey questions that measure whether the item being studied improved, deteriorated or stayed the same. Broadly, a typical sample of responses might be summarized by the vector *{u,s,u,d,d,d,u,s,…}* where *u* indicates an increase (up), *s* indicates no change (same), and *d* indicates a decrease (down).

Diffusion Index Calculation

In this case, we define a diffusion index *D* as $D = \frac{n^u}{n} - \frac{n^d}{n}$ where *n* is the total number of respondents, n^u is the number of respondents who reported an increase and n^d is the number of respondents who reported a decrease.

If there are *n* survey respondents, then the survey process can be interpreted as *n* independent trials. If we generalize that the possible responses are *a* (where *a* is u, d or s), then each answer type has a "success" probability of p^a. We then define $\hat{p}^a = n^a/n$ as the proportion of survey answers that are type *a*. If type-*a* answers are assigned a value of ω^a in the diffusion index — where ω^a is a number between $\underline{\omega}^a$ and $\overline{\omega}^a$ (that is, values of -1, 0 or 1) — then, a general diffusion index statistic can be defined as

$$\widehat{D} = \sum_{a=1}^{r} \omega^a \hat{p}^a$$

In the Richmond Fed survey, $(\omega^u, \omega^s, \omega^d) = (1,0,-1)$, so that $\widehat{D} = \hat{p}^u - \hat{p}^d$. In this case, an expansion would therefore be indicated by a value of $\widehat{D}$ greater than 0. One problem with this interpretation is that $\widehat{D}$ can be equal to zero under two completely different conditions:

$$\hat{p}^u = \hat{p}^d \text{ or } \hat{p}^s = 1$$

In other words, half of the respondents can say "up," and half can say "down," or all respondents can say "same." To distinguish between these two cases, we would need information about the variance of $\widehat{D}$. The equation is as follows:

$$\widehat{D} \pm z\sqrt{\frac{\left[(1-\hat{p}^s) - \widehat{D}^2\right]}{n}}$$

Thus, when $p^u = p^d = 0.5$ and $p^s = 0$, $CI = \pm\, z\left(1/\sqrt{n}\right)$, and when $p^u = p^d = 0, p^s = 1, CI = 0$.

Confidence Intervals

Regarding calculating confidence intervals, the table below can help provide insight. It is constructed based on the Richmond Fed manufacturing employment index from 2002 through 2018 and provides a rule of thumb for how to calculate the confidence interval and assess the statistical significance of the diffusion index.

		n								
		40	**60**	**80**	**100**	**120**	**140**	**160**	**180**	**200**
p^s	**0.0**	16	13	11	10	9	8	8	7	7
	0.1	15	12	11	9	9	8	7	7	7
	0.2	14	11	10	9	8	8	7	7	6
	0.3	13	11	9	8	8	7	7	6	6
	0.4	12	10	9	8	7	7	6	6	5
	0.5	11	9	8	7	6	6	6	5	5
	0.6	10	8	7	6	6	5	5	5	4
	0.7	9	7	6	5	5	5	4	4	4
	0.8	7	6	5	4	4	4	4	3	3
	0.9	5	4	4	3	3	3	2	2	2
	1.0	0	0	0	0	0	0	0	0	0

Each cell in the table reports (half of the length of) the confidence interval for the null hypothesis that the diffusion index is equal to zero. Specifically, it shows the values $z\sqrt{(1-p^s)/n}$ for different combinations of p^s (the proportion of individuals that responded "stay the same") and n (the sample size). The table uses $z = 1$, or a 68 percent confidence level.

In our example, the mean/median share that responded "stay the same" was 0.7 (ranging from 0.4 to 0.8), and n ranged from 60 to 140, with a mean of 90 and a median of 85. The cell for p^s = 0.7 and n = 80 shows a value of 6. This means that the diffusion index D is significantly different from zero when its (absolute) value is larger than 6.

Variance

In the Richmond Fed monthly business surveys, the variance equals $(1 - p^s - D^2)/n$, where $D = (p^u - p^d)$. Thus, we observe the following with respect to the variance of the observed diffusion index ($\widehat{D}$):

1) Because $\widehat{D}$ is derived from a weighted sum of means, its variance decreases at rate $n.$ In other words, the larger the sample, the smaller the uncertainty.
2) The variance decreases with the magnitude of the diffusion index $D^2 = (p^u - p^d)^2$. Note that D^2 is equal to 1 when all survey participants respond either "up," "down" or "same." In all three cases, $(1 - p^s - D^2) = 0$.
3) On the other hand, the variance will be maximized if $p^u = p^d$ and $p^s = 0.$

Observations (2) and (3) are key to this point: An equal split of "up" and "down" responses might lead to a reported index of 0 — just like all respondents reporting "same" — but the variance in the former will be higher than in the latter. This is because uncertainty in the first case is much higher. Formally:

- If $\hat{p}^u = \hat{p}^d$, then $Var(\widehat{D}) = (1 - p^s)$.
- If $\hat{p}^s = 1$, then $Var(\widehat{D}) = 0$.

In this sense, the variance of the diffusion index can be thought of as reflecting polarization or disagreement across responses.