**NBER Industry Uncertainty Data**

**Overview**

This dataset contains two industry by year uncertainty measures from the US ASM Census data and three from CRSP. These are described in detail in Really Uncertain Business Cycles (by Bloom, Floetotto, Jaimovich, Saporta-Eksten and Terry), NBER WP 182245 and <http://www.stanford.edu/~nbloom/RUBC.pdf>

Any questions please e-mail Itay Saporta (isaporta@stanford.edu) or Nick Bloom (nbloom@stanford.edu). These measures are also briefly described below:

**Census data measures**

**Uncert\_tfp**: This is the standard-deviation of the innovations to establishment TFP in each industry-year in the Annual Survey of Manufacturing. So, for example, if broad-woven cotton textiles (SIC 2211) had 32 establishments in 1991 our measure of uncert\_tfp for that industry year would be the standard deviation of the TFP innovation across those 32 establishments.

TFP is measured in the standard-way, which is log(output) – factor-share weighted log(inputs). The innovation to TFP is obtained as the residual by regressing current establishment TFP against its lagged TFP, an establishment plant fixed-effect and a year fixed-effect. The ASM sample is all plants with 25+ years of data to control for sample selection issues, but results are very similar for plants with 2+ years or plants with 39 years (a balanced panel). The results are also similar if the TFP innovation is the residual for a larger regression including multiple lags (and powers) of TFP, employment, capital, materials and energy, which tries to include a fuller information set (including factor input choices) so that the TFP innovation is more a “shock”.

The underlying measure – and the one used in the paper “Really Uncertain Business Cycles” - is taken over all establishments within an industry-year. To avoid potential disclosure problems[[1]](#footnote-1) with the public use microdata this is then smoothed over time to increase the number of establishments in each industry-year cell. This is done by weighting over the current year plus (1/T2) other years where T is the distance in years+1. So, for example, the 1995 value is the underlying 1995 value, plus ¼ of the 1994 and 1996 values, plus 1/9 of the 1993 and 1997 values, plus 1/16 of the 1992 and 1998 values etc. These weights are normalized so they add to 1 for each year. This induces some light temporal smoothing, although the results using this and the un-smoothed data are very similar (the fully unsmoothed data is available in the Census RDC and the results and d-stats for the unsmoothed data are available in the paper).

After smoothing, the mean industry-year cell has a sample size of 1390 establishments (noted in the variable “indobs”), which reflects that fact that the ASM covers over 500,000 establishment-year observations, so the underlying dataset is very large.

**Uncert\_sales**: This is the cross-sectional standard-deviation of real sales growth in each industry-year in the ASM. It is constructed using the same sample and smoothing procedures as for uncert\_tfp. The only difference is this uses the raw sales – to be precise, the $ value of shipments reported by establishments on the Census AM form deflated using the NBER-CES sales deflator at the industry-year level. So for anyone that worries what TFP really measures (it is “a measure of our ignorance”) then this alternative measure has the benefit of being the dispersion in the raw survey response data taken directly from the survey. As it happens uncert\_sales and uncert\_tfp behave very similarly (correlation log of these two measures, weighted by indobs, is 0.5987).

**Indobs**: This is the cell-size of establishments for which uncert\_tfp and uncert\_sales were calculated. Since this smooths across all years in each industry, it does not vary over time within an industry, even though the weighting clearly varies heavily. Note that the 1/T2 was chosen to give very high weight to current years and low weight to years more than a couple of years away. Ideally we would have had no smoothing over time, but Census clearance cell-size requirements required us to have some smoothing to cover more establishments in each cell, so this was the best compromise possible between getting publicly available industry\*year uncertainty data and not breaching Census cell-size clearance rules.

Finally, note that because of the switch from SIC-1987 to NAICS in 1997 the 1996 data is about 5% too high. The reason is that when switching from NAICS to SIC some establishments appeared to re-categorize some of their inputs and outputs, leading to a larger jump than usual in the difference between years in these variables (i.e. measurement error in the TFP innovation is higher in that particular year). The cross-industry rankings of the various uncertainty measures look similar in 1996 compared to other years (so industry rankings look stable), but the overall level appears to jump by about 5%, due to more measurement error in that year. For macro plots we typically drop that year, but for applied micro regressions which include a time fixed-effect we usually include that year in the data.

**CRSP data measures**

**Uncert\_ret:** This is the standard-deviation of the monthly stock-returns of all CRSP firms within the industry-year in the dataset. So, for example, if 10 firms were in an industry then this would be the standard-deviation across the 120 months (12\*10) of stock returns for that industry-year.

**Uncert\_tret**: This is the mean of the within year standard deviation of firms stock-returns within an industry year. For example, if an industry had 10 firms then each one would have a within-year standard-deviation of returns (each firm would have 12 monthly observations within that year over which a standard-deviation could be taken). This measure then takes the average of this across all firms within that industry year. It tries to get a measure of the average time-series nature of stock-returns volatility within an industry year.

**Uncert\_iret:** This is the standard-deviation of the mean annual return for each firm within an industry year. For example, if an industry had 10 firms then each would have a return for that year (the mean return across the 12 monthly returns) and this measure then takes the cross-sectional standard-deviation of this. This measure, thus, tries to get at the cross-sectional spread of returns within an industry year.

It should be clear from the above that uncert\_ret can be roughly broken down into uncert\_tret (which is time series component of the industry-year stock-returns uncertainty measure) and uncert\_iret (which is the cross-firm cross-sectional component). It turns out these are all very closely cross correlated, but we provided all three measures for completeness.

1. The issue is that without smoothing in some years we would have too few establishments in the industry\*year cell, but by smoothing over all years we have a larger group. For example, is an industry in which is a few years only 10 establishments are alive but across all 38 years there have been at least 25 establishments alive at some point for 25+ years (our underlying sampling criteria) would have a cell size of 25 using our smoothed measure and 10 using the unsmoothed measure. [↑](#footnote-ref-1)