Notes to Accompany Discussion of “Can Unemployment Insurance Spur Entrepreneurial Activity? Evidence from France” by Hombert, Schoar, Sraer and Thesmar

Let $SP$ and $NE$ denote the number of new sole proprietors and new businesses with employees, respectively, let $i$ index industries, and let $t$ index time. Consider a very simple model in which the policy intervention works entirely through its effect on the aggregate number of new sole proprietors. Assumptions:

(A.1) The policy intervention does not alter the allocation equations for new entrepreneurs →

$$SP_{it} = p_iSP_t \quad \text{and} \quad NE_{it} = n_iNE_t$$

(A.2) The policy intervention has no effect on aggregate $NE$.

(A.3) The policy intervention raises aggregate $SP$ by $\Delta SP$, where $\Delta$ takes the same value for all levels of $SP$. That is, the policy effect is proportional to the no-intervention level of $SP$. (The algebra is less nice when $\Delta SP$ is independent of $SP$.)

(A.4) The environment is stationary over time in the sense that $SP_t/NE_t$ is the same in the Pre and Post periods, absent the policy intervention.

Given these assumptions and definitions, we can write the authors’ treatment variable as follows:

$$f_i = \frac{p_i SP_{Pre}}{p_i SP_{Pre} + n_i NE_{Pre}} = \frac{p_i Y}{p_i Y + n_i}, \quad \gamma = \frac{SP_{Pre}}{NE_{Pre}}$$

The authors investigate how policy intervention responses differ across industries with higher and lower values of $f$. For example, they consider how the log change in the number of new firms differs between industries with higher and lower values of $f$, as in Figure 3. Evaluating this difference-in-difference under Assumptions A.1 to A.4 for two industries with high and low values of $f$, we have

$$DD = \ln \left[ \frac{SP_{H,Post} + NE_{H,Post}}{SP_{H,Pre} + NE_{H,Pre}} \right] - \ln \left[ \frac{SP_{L,Post} + NE_{L,Post}}{SP_{L,Pre} + NE_{L,Pre}} \right]$$

$$= \ln \left[ \frac{p_H(1 + \Delta)SP + n_HNE}{p_H SP + n_HNE} \right] - \ln \left[ \frac{p_L(1 + \Delta)SP + n_LNE}{p_L SP + n_LNE} \right]$$

$$= \ln \left[ \frac{p_H(1 + \Delta)\gamma + n_H}{p_H \gamma + n_H} \right] - \ln \left[ \frac{p_L(1 + \Delta)\gamma + n_L}{p_L \gamma + n_L} \right]$$

$$= \ln \left[ \frac{p_H(1 + \Delta)\gamma + n_H}{p_H \gamma + n_H} \right] - \ln \left[ \frac{p_L(1 + \Delta)\gamma + n_L}{p_L \gamma + n_L} \right]$$
$DD(\Delta=0) = 0$. That is, the change over time in the outcome variable is the same in both industries absent the policy intervention.

**Point 1:** Diff-in-diff estimation yields $DD$, while $p_H$, $p_L$, $n_H$, $n_L$, and $\gamma$ are parameters that can be calculated directly from the data. Thus, equation (1) lets us solve for the size of the policy effect on the treatment variable, $\Delta$ -- an intrinsically interesting quantity. Moreover, using the calculated value of $\Delta$, we can recover the effect of the policy intervention on the outcome variable per unit of the induced effect on $SP$. For purposes of external comparison and the evaluation of other policy interventions, it is advantageous to express the response of the outcome variable per unit impact on $SP$.

Now calculate how the difference-in-difference for the outcome variable varies with the size of the policy intervention effect, as measured by $\Delta$:

$$
\frac{d\ DD}{d\ \Delta} = \left[ \frac{p_H Y}{p_H (1 + \Delta) \gamma + n_H} \right] - \left[ \frac{p_L Y}{p_L (1 + \Delta) \gamma + n_L} \right]
$$

(2)

**Point 2:** Thus, for a non-marginal policy intervention, the mapping from the policy effect on the treatment variable to the difference-in-difference for the outcome variable is complicated. It depends in a nonlinear manner on industry-level and aggregate parameters and on the size of the policy effect on $SP$. Figures 2.C and 2.D suggest that $\Delta$ is quite sizable, perhaps on the order of 0.4.

For a small policy intervention ($\Delta \approx 0$), this derivative reduces to

$$
\lim_{\Delta \to 0} \frac{d\ DD}{d\ \Delta} = f_H - f_L.
$$

(3)

**Point 3:** Thus, even for a marginal policy intervention, the mapping from the policy effect (on the treatment variable) to the difference-in-difference estimate depends critically on how the treatment and control group industries differ with respect to the relative prevalence of new sole proprietors and new employer businesses.

In practice, the authors consider dozens of industries in the treatment group $H$ and dozens of other industries in the control group $L$. So the difference-in-difference estimate involves differences in some averages of these parameters for $H$-type industries and $L$-type industries. One can plug average values of the parameters into equations (1)-(3) or explicitly derive the many-industry analogs to equations (1)-(3).

**Point 4:** Altering (A.1) or (A.2), or relaxing the proportionality condition in (A.3), would yield a different mapping from the policy effect on the treatment variable to the “treatment effects” on the outcome variables. (Relaxing (A.4) is probably best handled with regression controls, e.g., time effects.)