EU Start-up Calculator
Technical Manual

December 8, 2020

Cristiana Benedetti Fasil
European Commission
Joint Research Centre

Petr Sedláček
University of Oxford
CFM-LSE, CEPR

Vincent Sterk
University College London
CFM-UCL, CEPR

1 The Start-up Calculator

1.1 Data

The start-up calculator uses publicly available information from Eurostat Business Demography Statistics spanning the period of 2008 to 2017 and 2013 to 2017 for some EU Member States. This dataset includes (among other things) information on the number of firms with at least one employee, average number of persons employed and survival rates by firm age, where the latter is considered in the following age categories: 0 (start-ups), 1, 2, 3, 4, 5 and all.

The *number of firms* of age a in year t, $n_{a,t}$, is directly observable in Eurostat, as is *firm size* (persons employed) by age, $s_{a,t}$. We denote the *survival rate* (also observable in Eurostat) by age as $1 - x_{a,t}$.

1.2 Accounting for start-ups: methodology

Because firms older than 5 are grouped together in Eurostat, it is necessary to interpolate information for each of the individual age categories. In addition, because the sample period ends in 2017, it is necessary to extrapolate the information up until 2019, just before we perform our scenario analysis. In what follows, we describe the interpolation and extrapolation methods employed in the start-up Calculator.

1.2.1 Age profile of exit and average size

Two inputs to the calculator are the profiles of average size and the survival rates by age in the baseline scenario (i.e. without shock), for firms up to age 15. For firms up to age 5, we measure directly in the data as averages over the sample period. For older firms, we assume a functional form for both profiles and fit these to the available data. Specifically, for the exit rate we assume the following functional form:

$$x_a = \beta_0 + \beta_1 \left(\frac{\exp \beta_2}{1 + \exp \beta_2} \right)^{a-1}.$$

and note this implies a smooth profile, gradually decaying from an initial point $x_{a=1} = \beta_0 + \beta_1$ to a limit point $x_{a \to \infty} = \beta_0$. The parameter β_2 controls the speed of decay.

Regarding the average size profile we assume a simple linear form:

$$n_a = \gamma_0 + \gamma_1 a.$$

The functional forms for these two profiles capture well patterns documented using data sets for which exit rates can be computed for all age groups (such as the US Longitudinal Business Database, see e.g. Pugsley, Sedlacek and Sterk (2017)[1]).

To estimate the parameters of these profiles we use a minimum distance estimator, targeting the following outcomes which we can observe in the data:

1. The average exit rate by age, for firms up to age 5
2. Average size of firms by age, for firms up to age 5
3. The average exit rate among all firms
4. Average size among all firms

Note that given a profile for the exit rate by age, one can compute the firm age distribution, and then the average exit rate by weighting the exit rates by age with the firm shares in each age bin.[2] Then, given the age distribution and the average size profile by age, one can compute average size across all firms.[3] The estimation is implemented in Matlab.

Finally we export the two profiles to Excel, for firms between age 5 and 15. For firms up to age 5, we use the averages measured in the data as opposed to the fitted profiles.

1.2.2 Extrapolation of information until 2019

Information on start-ups and young firms. In order to extrapolate the necessary data between 2017 and 2019, we assume that firm size by age and exit rates by age (up to age 15), and the number of start-ups, all linearly converge to their

[1] "The Nature of Firm Growth", working paper, see https://ideas.repec.org/p/cfm/wpaper/1737.html

[2] In particular, the share of firms in age bin a is given by $\frac{\prod_{k=1}^{a}(1-x_k)}{1+\sum_{a=1}^{T}\prod_{k=1}^{a}(1-x_k)}$, where T is a truncation age which we set at 100 years. When computing the age distribution, we use the data averages for firms up to age 5, rather than the fitted profile.

[3] In doing so, we again use actual data average for firms up to age 5, as opposed to the fitted profiles.

2008-2017 averages:

$$x_{a,2017+\tau} = x_{a,2017} + \frac{\tau}{2}(\overline{x}_a - x_{a,2017}),$$

$$s_{a,2017+\tau} = s_{a,2017} + \frac{\tau}{2}(\overline{s}_a - s_{a,2017}),$$

$$n_{0,2017+\tau} = n_{0,2017} + \frac{\tau}{2}(\overline{n}_0 - n_{0,2017}),$$

for $\tau = 1, 2$ and $a = 1, 2, ..., 15$, and where $\overline{x}_a$, $\overline{s}_a$ and $\overline{n}_0$ denote the 2008 to 2017 averages of age-specific exit rates, firm sizes and the number of start-ups, respectively. Using the above, we can then recover the number of firms for the ages of 1 to 15 as $n_{a,t} = n_{a-1,t-1}(1 - x_{a,t})$, for $a = 1, 2, ..., 15$ and $t = 2018, 2019$.

Number of older firms. In order to compute aggregate employment, it is also necessary to assume a particular time-path for employment of 16+ year old firms. However, because 16+ year old firms are unaffected by our scenarios, the particular time-path is quantitatively unimportant for the results which are reported in deviations from the assumed trend. For this reason, we simply assume that employment in 16+ year old firms stays fixed at its 2017 level.

1.2.3 Constructing alternative scenarios

Having the above information, we are ready to conduct scenarios starting in 2020 and running through to 2030. We consider three types of margins: (i) changes in the number of start-ups, (ii) changes in growth potential and (iii) changes in survival rates.

Scenarios involving (i) and (iii) are straightforward. Upon impact, we lower the number of start-ups and/or the survival rates of young firms by a certain value and keep this value for a certain period. Growth potential works on the same principle, but applies to the *cohort* of start-ups which enters in 2020. Therefore, lowering the growth potential by a certain percentage value results in the entire *growth profile* of firms born in 2020 shifting downwards. Importantly, the size of firms which in 2020 are older than 0 years is unaffected.

To be concrete, for a given scenario, let us denote the initial percentage decreases in the number of start-ups, the growth potential of start-ups and the survival rate of young firms by $\zeta_j \in (0, 1)$, where $j = \{n, s, x\}$, respectively. Let us further denote the duration of these effects by $\tau_j > 0$, where $j = \{n, s, x\}$, respectively. The given scenarios are then given by

$$
\begin{aligned}
n_{0,2019+t} =& n_{0,2019}(1-\zeta_n), \quad \text{for } t=1,...,\tau_n, \\
s_{a,2019+t+a} =& s_{a,2019}(1-\zeta_s), \quad \text{for } t=1,...,\tau_s, \text{ and } a=0,1,2,...,15, \\
x_{a,2019+t} =& x_{a,2019}(1-\zeta_x), \quad \text{for } t=1,...,\tau_n, \text{ and } a=1,2,...,15.
\end{aligned}
$$

Notice that in the above, the changes in growth potential apply to *cohorts* of start-ups. For instance, if the effect of the pandemic lasts only for one year ($\tau_s = 1$), then only start-ups in 2020 are affected. In 2021, it is one year old firms which have lower growth potential, i.e. the cohort born in 2020, while firms of all other ages (including new start-ups), are unaffected. In contrast, the pandemic affects the survival rates of all young firms simultaneously and therefore businesses aged 0 to 10 years experience a drop in survival rates in 2020. Also note that the number of businesses older than (i.e. $a > 0$) years is given by $n_{a,t} = (1 - x_{a,t})n_{a-1,t-1}$.

Our calculator can also accommodate bounce-back scenarios. These are always defined as certain values above the 2008-2017 averages of the number of start-ups, average sizes and survival rates of young firms. Recall that all these margins converge precisely to the respective 2008-2017 averages by 2019.

Specifically, let us denote the percentage increase (above the respective long-run average) in the bounce-back scenario related to the number of start-ups, the growth potential of young firms and their survival rates by χ_j, where $j = \{n, s, x\}$, respectively. Furthermore, let us denote the length of the bounce-back period by σ_j, where $j = \{n, s, x\}$, respectively. The given bounce-back scenarios are then given by

$$
\begin{aligned}
n_{0,2019+\tau_n+t} =& n_{0,2019}(1+\chi_n), \quad \text{for } t=1,...,\tau_n, \\
s_{a,2019+\tau_s+t+a} =& s_{a,2019}(1+\chi_s), \quad \text{for } t=1,...,\tau_s, \text{ and } a=0,1,2,...,15, \\
x_{a,2019+\tau_x+t} =& x_{a,2019}(1+\chi_x), \quad \text{for } t=1,...,\tau_n, \text{ and } a=1,2,...,15.
\end{aligned}
$$

Finally, in all scenarios aggregate employment in a given year is computed simply as the sum of employment in firms aged 0 to 15 and the (extrapolated) employment of firms older than 16 years. Therefore, we are being conservative in the sense that we are not allowing businesses aged 16 and more years to be affected by the crisis. Our results should, therefore, be considered as a lower bound on the given scenarios. While the margins of start-ups and growth potential would only “kick in” after 2030 for these older firms, their survival rates may very well be affected in 2020 already.

1.3 Adjusting for equilibrium effects

The calculations above abstract from potential equilibrium effects. In this subsection, we describe how to adjust for this, by placing the calculator within a "shell" formed by a basic but standard heterogeneous-firm model. This model also clarifies how the calculator connects to canonical equilibrium models of firm dynamics.

In the model, there is a measure M of heterogeneous firms.[4] Let the production function of firm i be given by

$$y_i = z_i n_i^{\alpha},$$

where y_i is the firm's output, n_i its employment level, z_i is the firm's productivity level, and $\alpha \in (0,1)$ is the elasticity of production with respect to labour input.[5] The wage per employee is taken as given by firms, and denoted by w. The firm chooses its level of employment in order to maximize profits, given by $y_i - wn_i$. This implies the following familiar solution for labour demand by firm i:

$$n_i = (z_i)^{\frac{1}{1-\alpha}} \left(\frac{w}{\alpha}\right)^{\frac{1}{\alpha-1}}$$

Aggregating over all firms, aggregate labour demand is given by:

$$N = M \left(\frac{w}{\alpha}\right)^{\frac{1}{\alpha-1}} \chi$$

where $\chi \equiv \int z^{\frac{1}{1-\alpha}} dF(z)$, where F is the CDF of the productivity distribution. Taking logs and differentiating (keeping idiosyncratic productivities constant), we can decompose changes in aggregate labour demand as:

$$d \ln N = \underbrace{d \ln M}_{\#\text{firms}} + \underbrace{d \ln \chi}_{\text{growth potential}} + \underbrace{\frac{1}{\alpha - 1} d \ln w}_{\text{wages}} \tag{1}$$

The first two terms reflect changes in, respectively, the number of firms and their

[4]Although the model is dynamic, it can be described entirely in static terms, hence we omit time subscripts.

[5]We abstract from capital for simplicity. Augmenting the model with capital would not change any of our results.

growth potential (productivity), whereas the third term captures equilibrium effects due to wage conditions.[6] Equation (1) can be understood as an aggregate labour demand curve, which is shifted by the number of firms and their growth potential.

To close the model, we need to specify how labour supply is determined. We assume there is a representative household with Greenwood-Hercowitz-Huffmann preferences. Specifically, the household's level of utility is given by: $U(C,N) = \frac{1}{1-\sigma}\left(C - \mu\frac{N^{1+\kappa}}{1+\kappa}\right)^{1-\sigma}$, where C denotes consumption and μ, κ, $\sigma > 0$ are preference parameters. The household chooses C and N to maximize utility, subject to a budget constraint given by $C = wN + \Pi$, where Π are aggregate firm profits. Utility maximization implies the following labour supply curve: $\mu N^{\kappa} = w$. Taking logs and differentiating gives the labour supply schedule:

$$d\ln N = \frac{1}{\kappa} d\ln w \tag{2}$$

Combining the labour demand and supply schedules, Equations (1) and (2), we can solve for the equilibrium level of aggregate employment:

$$d\ln N = \underbrace{\Psi}_{\text{equilibrium dampening}} (\underbrace{d\ln M + d\ln\chi}_{\text{calculator output}}) \tag{3}$$

where $\Psi \equiv \frac{1}{1-\kappa\epsilon_{nw}} \in (0,1)$, where $\epsilon_{nw} = \frac{1}{\alpha-1}$ is the wage elasticity of labour demand. Equation (3) expresses aggregate employment (in deviation from some baseline trend) as a function of the number of firms and their growth potential. The latter two we obtain as outputs from the calculator.[7] The parameter Ψ is an equilibrium dampening coefficient, which depends on the elasticity of labour demand (ϵ_{nw}) and the Frisch elasticity of labour supply ($\frac{1}{\kappa}$). Based on these two parameters and the output from the calculator, we can thus compute the equilibrium change in aggregate employment from Equation (3).

To gauge how large such equilibrium dampening effects could be we consider standard values for the model parameters. Specifically, we could assume a unit Frisch elasticity of labour supply ($\kappa = 1$) which is in the ballpark of the estimates in the micro and macro literature. The parameter α could be set in accordance with the

[6]Other sources of equilibrium dampening could derive from endogenous entry and exit, which we abstract from here.

[7]Alternatively, one could model an explicit entry and exit block of the model.

labour share of aggregate income, which is around sixty percent in the US, implying $\alpha = 0.6$. Given these numbers, we obtain $\Psi = 0.29$, i.e. equilibrium effects dampen just over seventy percent of the decline in aggregate employment.

Note however, that the above model does not contain any labour market frictions. In the presence of such frictions, labour demand is likely to be less sensitive to wages. We therefore could use a direct empirical estimate of the labour demand elasticity. Lichter, Peichl and Siegloch (2015) conduct a meta study of empirical estimates and recommend an elasticity of -0.246. Setting $\epsilon_{nw} = -0.246$ (and again $\kappa = 1$) we obtain a coefficient of $\Psi = 0.80$, i.e. 20% dampening. Otherwise, we could use elasticities that are consistent with values adopted by the European Commission QUEST and RHOMOLO models. In this case, the labour supply elasticity is set at 0.25 and the labour demand elasticity at -0.1. These elasticities result in a dampening effect of 29%. These value have been used by Benedetti-Fasil, Sedláček and Sterk (2020a), Benedetti-Fasil, Sedláček and Sterk (2020b) and Benedetti-Fasil, Sedláček and Sterk (2020c) in their analysis of the impact of COVID-19 on employment in EU Member States. Moreover, they conforms with other evidence that equilibrium dampening effects may not be that strong. For instance, Sedláček (forthcoming) shows that a search and matching model with heterogeneous firms displays relatively weak equilibrium dampening effects. In a recession, the slack labour market (increasing the chances of hiring and reducing wages) is not a strong enough force to overturn the impact of a missing generation of start-ups. In any case, the tool allows to choose different elasticities according to the labour market structure of the EU Member State considered.

Finally, we note that if a scenario is based on empirical observations for average size of young firms (for the start-up growth potential margin), then it may be important to account for the fact that this number itself is subject to equilibrium dampening. Therefore, the true change in growth potential might be larger than what the data suggest. To do so, we use Equation (1), but this time aggregated over only start-ups, as opposed to all firms.[8] Using Equation (2) to substitute out the wage and rearranging, we obtain the following expression for start-up growth potential:

$$d \ln \chi^{start-up} = \underbrace{d \ln N^{start-up} - d \ln M^{start-up}}_{\text{avg start-up size}} - \underbrace{\kappa \epsilon_{nw} d \ln N}_{\text{equil. adjustment}} .$$

[8] This gives $d \ln N^{start-up} = d \ln M^{start-up} + d \ln \chi^{start-up} + \epsilon_{nw} \ln w$.

On the right hand side, the first two terms jointly are the change in average start-up size. From this one subtracts the $\kappa\epsilon_{nw}$ times the change in *aggregate* employment in order to obtain the change in the growth potential of start-ups.[9]

References

Benedetti-Fasil, Cristiana, Petr Sedláček, and Vincent Sterk, "EU start-up calculator: impact of COVID- 19 on aggregate employment. Scenario analysis for Austria, Belgium, Germany, Hungary, Italy and Spain," Technical Report, Publications Office of the European Union, Luxembourg 2020a.

_ , _ , and _ , "EU start-up calculator: impact of COVID- 19 on aggregate employment. Scenario analysis for Denmark, Estonia, Finland, France, Latvia, Lithuania, Portugal and Sweden," Technical Report, Publications Office of the European Union, Luxembourg 2020b.

_ , _ , and _ , "EU start-up calculator: impact of COVID- 19 on aggregate employment. Scenario analysis for Bulgaria, Croatia, Czechia, Luxemburg, Netherlands, Poland, Romania, Slovakia and Slovenia," Technical Report, Publications Office of the European Union, Luxembourg, forthcoming 2020c.

Lichter, Andreas, Andreas Peichl, and Sebastian Siegloch, "The Own-Wage Elasticity of Labor Demand: A Meta-Regression Analysis," *European Economic Review*, 2015, *80* (C), 94–119.

Sedláček, Petr, "Lost Generations of Firms and Aggregate Labor Market Dynamics," *Journal of Monetary Economics*, forthcoming.

[9]Note that the adjustment only matters when aggregate employment is away from its trend level. It turns out that in our application here, this adjustment has only negligible effects, and hence we omit it in our calculations.