Sensitivity Analysis on Samples and Specifications with JLS

We explored a number of differences between our sample design and specification and those of JLS and found them to be unable to explain the difference in our main results. The possible explanations we explored (estimates not reported but available upon request) include:

a. JLS included in their comparison group workers who were observed to separate and later returned to the same employer (recalls), while we omit these individuals.

b. JLS included in their sample separators from firms that closed, while we omit these individuals.

c. JLS restricted their sample to workers with at least six years of tenure, while our tenure restriction is three years.

d. In pooling the sample across dates of separation, JLS hold coefficients constant over time, and therefore across macroeconomic conditions, whereas our separate samples allow those coefficients to vary.

e. JLS's data do not allow them to follow workers who become re-employed in another state, while our data infrastructure allows us to track individuals who cross state lines.[1]

f. JLS include individual time trends in their estimation of specification analogous to (2) in the main text. [2]

This sensitivity analysis suggests that the differences between our results for non-distressed separators and those in JLS are due to the differences in time and place of our data.[3] The main text has further discussion of this argument including discussion of the related literature.

Winsorization

Let z_i be the greater of the median of earnings observed for individual i 24 quarters before or after the reference quarter and 10,000 ($z_i = \max\{\text{median}\{y_{it}\},10000\}$). Then define the earnings growth rate for each individual and quarter as:

$$\Delta_{it} = (y_{it} - z_i)/[\frac{1}{2} * (y_{it} + z_i)]$$

The growth rate, Δ_{it}, captures the extent to which the current earnings exceed the typical earnings of that individual in a given quarter. This growth rate, made popular by Davis et al. (1996) and commonly referred to as the DHS growth rate, is bounded between -2 and 2. We use this growth

[1] In addition, JLS restricted their sample to workers with positive earnings in every calendar year, whereas we require positive earnings within eight quarters of separation. Von Wachter, Song, and Manchester (2009) show that the earnings losses for non-distressed separators are larger and more persistent when separators with zero annual earnings are included in the sample. JLS also appear to limit their sample of stayers to stayers at firms that experienced some separations. We have not replicated these sample restrictions, but we expect that the differences between them and our restrictions are too small to account for the large difference in estimated outcomes.

[2] The main text Figure 8 shows the impact of including individual-specific time trends for our generalized specification (4). In unreported results, we have also included individual-specific time trends as a robustness check in estimating (2) and find that this does not make distressed separators have larger and more persistent earnings losses.

[3] JLS also did not have the benefit of the extensive data quality controls currently used by the Census Bureau in the LEHD program. The resulting measurement error could also contribute to the difference in findings.

rate to identify large increases in quarterly earnings that are likely driven by data errors. The choice of the minimum value of z as 10,000 is made such that we do not accidentally winsorize earnings for low earners. We chose to edit the earnings values if they exceed the 95th percentile of earnings growth rates such that if we were to recalculate Δ_{it} using the edited earnings then Δ_{it} would be equal to the 95th percentile of earnings growth rates. Specifically, let $\Delta(p95)$ be the 95th percentile of Δ_{it}. The earnings data used in the analysis are equal to:

$$y_{it} = \begin{cases} y_{it} & if\ \Delta_{it} < \Delta(p95) \\ z_i * \dfrac{1 + .5 * \Delta(p95)}{1 - .5 * \Delta(p95)} & else \end{cases}$$

Relative to standard winsorization methods that identify outliers in levels, this method has the advantage of correctly retaining the earnings records of high wage individuals.

Inverse Survival Functions

We start by estimating the logistic regressions presented in equations 7 and 8. For notational simplicity, let M=[1;X;Z;g] be a matrix of the concatenation of all the right-hand-side variables and let $\emptyset_t = [\alpha_t; \beta_t; \gamma_t; \lambda_t]$ be the corresponding vector of coefficients. For each t, we use the output from the logistic regression from equation 7 to calculate the baseline probability of finding a new job in that period conditional on not being re-employed prior to t and having a firm of growth rate g=[1,2,3,4]. We denote this conditional probability as h_t^n, and it is calculated as follows:

$$h_t^{n,g} = \frac{\exp\left(\bar{M}_t\widehat{\emptyset}_t + \frac{\delta_{gt}}{2}\right)}{\left[1 + \exp\left(\bar{M}_t\widehat{\emptyset}_t + \frac{\delta_{gt}}{2}\right)\right]} - \frac{\exp\left(\bar{M}_t\widehat{\emptyset}_t - \frac{\delta_{gt}}{2}\right)}{\left[1 + \exp\left(\bar{M}_t\widehat{\emptyset}_t - \frac{\delta_{gt}}{2}\right)\right]}$$

Where $\bar{M}$ denotes a vector of the mean value of all covariates and $\widehat{\emptyset}_t$ is the vector of coefficient estimates from the logistic regression.[4] Using this same methodology we use the estimates from equation 8 to calculate the condition probabilities for recalls, denoted h_t^r. To summarize, the estimates from the logistic regression allow us to calculate two types of conditional probabilities:

$$h_t^{n,g} = Pr(new\ job\ in\ t\ |\ not\ reemployed\ before\ t\ \&\ firm\ growth\ rate\ g)$$

$$h_t^{r,g} = Pr(recall\ in\ t\ |\ not\ reemployed\ before\ t\ \&\ firm\ growth\ rate\ g)$$

Note that $h_t^r = 0$ for t<2 by construction.

Using these probabilities we then calculate the following:

$$h_t^g = \Pr(reemployment\ in\ t\ \& firm\ growth\ rate\ g) = h_t^{n,g} + h_t^{r,g}$$

$$P_0^{n,g} = \Pr(new\ job\ by\ t = 0\ \& firm\ growth\ rate\ g) = h_0^{n,g}$$

[4] Note that when we estimate the logistic regression g=4 is the reference firm growth rate category. The above notation is consistent with this if you simply assume that $\delta_{4t} = 0$.

$$P_t^{n,g} = \Pr(new\ job\ by\ t > 0\ \&\ firm\ growth\ rate\ g) = \sum_{\tau=0}^{t}\left(\prod_{s=0}^{\tau-1}(1-h_s^g)\right)h_\tau^{n,g}$$

$$P_t^{r,g} = \Pr(recall\ by\ time\ t\ \&\ firm\ growth\ rate\ g) = \sum_{\tau=2}^{t}\left(\prod_{s=0}^{\tau-1}(1-h_s^g)\right)h_\tau^{r,g}$$

Lastly, we calculate the probability of finding a new job by time t, conditional on never being recalled as:

$$\Pr(\text{new job by t} \mid \text{never recalled}\ \&\ firm\ growth\ rate\ g) = P_t^{n,g} \Big/ (1-P_8^{r,g})$$

We present the results as an inverse survival plot in which we create a plot in which the x-axis is t and the y-axis is the probability of re-employment by t and we plot four separate lines for the four estimates of Pr(new job by t | never recalled & firm growth rate g).

AKM Firm Effects

To estimate the AKM firm fixed effect, we use data on the earnings of all workers who appear in the LEHD between 2002 and 2009. For each worker and each year, we identify the main employer (i.e., the employer that provides over 50 percent of total earnings in that year) and we calculate the annual earnings associated with that employer in that year as the average quarterly earnings across all quarters in which the worker had strictly positive earnings at the employer. Using these worker-by-year data we then regress log of annual earnings on an individual fixed effect, a firm fixed effect, year fixed effect, and the interaction between education, sex, and a third-order polynomial in age.[5] To ease computational burden, we estimate this specification within nine distinct subsamples defined by the Census region in which the firm is located. Within each of these samples, we limit the sample to the largest connected set. To make the firm fixed effects comparable across Census regions, we normalize firm fixed effects by subtracting the mean value of the firm fixed effect for firms in the accommodation and food services industry. This normalization assumes that firms in this industry offer a pay premium of zero, on average.

Firm Productivity

We measure firm productivity using revenue and employment data from the Census Business Registrar and the Longitudinal Business Database (LBD). We measure productivity as log revenue per worker, which is a measure that has been commonly used to measure productivity at both the macro and micro level. While this is a relatively crude measure of productivity compared to total factor productivity (TFP), other research has found log revenue per worker is highly correlated with TFP within industries. We measure the productivity of each firm as the employment-weighted log revenue per worker between 2002 and 2009. We then calculate employment-weighted ranks within four-digit North American Industry Classification System (NAICS) industry codes. We are able to measure log revenue per worker for approximately 80 percent of

[5] To identify both year and age effects, we center age around 40 and include a quadratic and cubic transformation of age but omit the linear term.

firms in the LBD and the ranks are calculated within the universe of firms for which we can measure productivity between 2002 and 2009.