

APPENDIX

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:

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

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 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:

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

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 $\phi_t = [\alpha_t; \beta_t; \gamma_t; \lambda_t]$ 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:

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

Where \bar{M} denotes a vector of the mean value of all covariates and $\hat{\vartheta}_t$ is the vector of coefficient estimates from the logistic regression. ¹ 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:

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

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

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

Using these probabilities we then calculate the following:

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

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

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

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

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

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

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 $\Pr(\text{new job by } t \mid \text{never recalled \& firm growth rate } g)$.

¹ 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 $\delta_{4t} = 0$.