Web Appendix for Buhai and Teulings (2013): "Tenure Profiles and Efficient Separation in a Stochastic Productivity Model"

This version: April 2013

Abstract

This web appendix contains technical notes concerning the estimation procedures in our paper Buhai and Teulings (2013). We present full details of our wage-equations system FGNLS estimation– including how to correct the standard errors of the FGNLS parameter estimates for the variability of the tenure distribution parameters estimated in the first SML step. In our context, this two-step variance adjustment turns out not to matter at all.

1 Summary of our 2-step estimation procedure

The parametric specification we used for the tenure distribution parameters Ω, π, and μ_z was the following:

$$\begin{aligned} \Omega &= \exp\left(\omega_0 + \omega_1 S + e_\Omega\right) \qquad (1) \\ \pi &= \pi_0 + \pi_1 S + e_\pi \\ \mu_z &= \mu_0 + \gamma\overline{\sigma}\left(\pi_1 S + e_\pi\right) \end{aligned}$$

where e_Ω and e_π are normally distributed, uncorrelated, random worker effects with mean 0 and standard deviations σ_Ω and σ_π.

The final log likelihood for the first estimation in the paper, namely the ML estimation of the tenure distribution parameters, is then given by (consult the paper for full derivations):

$$\log L = \ln \int\int \prod_{j=1}^{J} \frac{\overline{F}(\Psi_j)^{1-d_j} \cdot f(\Theta_j)^{d_j}}{\overline{F}(\tau_1)^{I(j=1)} \left(\int_0^1 \overline{F}(x)dx\right)^{I(j\neq 1)}} d\Phi\left(\frac{e_\Omega}{\sigma_\Omega}\right) d\Phi\left(\frac{e_\pi}{\sigma_\pi}\right) \qquad (2)$$

where $I(y)$ is the indicator function, taking value 1 if y is true and 0 otherwise. We estimate this likelihood above using Simulated Maximum Likelihood (SML)[1].

[1] We used simulated maximum likelihood (SML), cf. Stern (1997). Sampling from a joint normal distribution with mean 0 and variances σ_π^2 and σ_Ω^2, and using a sampling size of 500 sampling points we achieved strong convergence in a reasonable number of iterations (the results are robust to any starting values of the parameters, as well as to altering the sampling dimension to any size between 100 and 500 sampling points). We used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for convergence of derivatives, allowing for a tolerance of 1E-4 times the absolute value of the log likelihood.

Using the ML estimates of $\omega_{0,1}, \pi_{0,1}, \sigma_{\Omega}$ and σ_{π} from (2) above[2], we can calculate the conditional expectations and variances of $\Delta\Omega_{\tau}$. Expressions for all these are given in the standard Appendix attached to the paper.

These expressions are then used for the the 2nd estimation, namely that of the parameters $\overline{\sigma}, \gamma, \mu_0, \sigma_z$, and σ_u by Feasible Generalized Nonlinear Least Squares (FGNLS), using data on wage changes. As in the paper, the general system of equations to estimate is the following (as explained in the text, when estimating we allow for a a second-order polynomial experience profile: t, t^2; for brevity and space we do not enter those components below).

$$\begin{aligned}
\Delta w_t &= \mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_{\tau} + \varepsilon_t \qquad (3)\\
w_t - w^*_{t-1} &= \mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega^*_{\Theta} + \overline{\sigma}\Omega_0 + \zeta_t \\
\Delta w_t^2 &= \sigma_z^2 + 2\sigma_u^2 + \overline{\gamma}^2\overline{\sigma}^2\mathrm{Var}\Delta\Omega_{\tau} + \left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_{\tau}\right)^2 + \eta_t \\
\left(w_t - w^*_{t-1}\right)^2 &= \sigma_z^2 + 2\sigma_u^2 + \overline{\gamma}^2\overline{\sigma}^2\mathrm{Var}\Delta\Omega^*_{\Theta} + \left(\mu_0 + \gamma\overline{\sigma}\pi S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega^*_{\Theta} + \overline{\sigma}\Omega_0\right)^2 + \nu_t \\
\Delta w_t \Delta w_{t-1} &= -\sigma_u^2 + \left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_{\tau}\right)\left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_{\tau-1}\right) + \upsilon_t
\end{aligned}$$

2 Consistent estimation of FGNLS parameters

This system (3) can be written in matrix notation as follows:

$$y = X(\beta, \theta) + u \qquad (4)$$

where y is a vector of the five dependent variables, stacked on top of each other; u is a corresponding vector of error terms, with a different variance for each of the five components, and cross-equation correlations; β is the vector of parameters in the current estimation step (the dimension of this vector differs in function of which variant of nested FGNLS model we use, see paper), and θ is the vector of parameters of the first step SML estimation (see earlier footnote) ; $X(\beta, \theta)$ is the matrix of functions of exogenous regressors and parameters. Equation (4) is estimable by feasible generalized nonlinear least squares (FGNLS). β is estimated, as standard in this case, as an iterative process: start with a value of $\widehat{\beta}$, then calculate $\widehat{\beta}^*$ as

$$\widehat{\beta}^* = \left[X_\beta\left(\widehat{\beta}, \widehat{\theta}\right)' X_\beta\left(\widehat{\beta}, \widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta}, \widehat{\theta}\right)' \left[y - X\left(\widehat{\beta}, \widehat{\theta}\right) + X_\beta\left(\widehat{\beta}, \widehat{\theta}\right)\widehat{\beta}\right]$$

where $X_\beta(\beta, \theta)$ is the matrix of partial derivatives of $X(\beta, \theta)$ with respect to β; repeat the procedure using $\widehat{\beta}^*$ as the new starting point, until $\widehat{\beta}^*$ converges to $\widehat{\beta}$. This procedure allows the calculation of $\widehat{u} = y - X\left(\widehat{\beta}, \widehat{\theta}\right)$, which enables us to compute an estimate of the variances and covariances of each of the five components of u. Let V be the variance-covariance matrix for these error terms. Then, a consistent estimator for β requires a similar iterative process on $\widehat{\beta}$ as

[2] We keep only four parameters, namely $\omega_{0,1}$ and $\pi_{0,1}$, as parameters σ_{Ω} and σ_{π} were estimated to be 0 in the first SML stage, see paper.

above, but using V instead of the identity matrix– as stated in the first line below.

$$\begin{aligned}
\widehat{\beta} &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[y - X\left(\widehat{\beta},\widehat{\theta}\right) + X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\widehat{\beta}\right] \qquad (5)\\
&= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[X(\beta,\theta) + u - X\left(\widehat{\beta},\widehat{\theta}\right) + X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\widehat{\beta}\right]\\
(p\lim) &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\left(\beta - \widehat{\beta}\right) + u + X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\widehat{\beta}\right]\\
&= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\beta + u\right] \Rightarrow\\
p\lim\left(\beta - \widehat{\beta}\right) &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[y - X\left(\widehat{\beta},\widehat{\theta}\right)\right]\\
\mathrm{Var}\left(\widehat{\beta}|\widehat{\theta}\right) &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1}
\end{aligned}$$

We further expanded the expression of the estimator for β after the first line in (5), in order to obtain $p\lim\left(\beta - \widehat{\beta}\right)$ in the fifth line, as that will be useful for our derivation of the correct, two-step adjusted variance, below. In the third line from above, we applied a Taylor expansion that holds for small estimation errors; given small errors, we can take the plim (formal notation to be corrected– but that is easy, e.g. standard in all textbooks).

3 Computing correct variance for the 2nd step FGNLS parameters

3.1 Murphy-Topel two-step variance adjustment

The variance $\mathrm{Var}\left(\widehat{\beta}|\widehat{\theta}\right)$, estimated in the last line of (5), does not account for the estimation error introduced by the parameters estimated in the SML analysis from (2), hence assuming $\theta = \widehat{\theta}$ as given. We show below how to do the required variance adjustment, and thus compute $\mathrm{Var}\left(\widehat{\beta}\right)$, by using derivations first presented in Newey (1984) and Murphy and Topel (1985). More recent coverage appears in the books by Wooldridge (2002, Ch.12) and Cameron and Trivedi (2005, Ch. 6).

For the derivation of the correct variance, $\mathrm{Var}\left(\widehat{\beta}\right)$, we need to use the estimation errors of parameters θ in the first SML step, $\theta - \widehat{\theta}$. Denote by V_θ be the variance-covariance matrix obtained in the first step. An upperbound of the correct variance of β can then be derived as follows:

$$\begin{aligned} p\lim\left(\beta-\widehat{\beta}\right) &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1}\left[y - X\left(\widehat{\beta},\widehat{\theta}\right) + X_\theta\left(\widehat{\beta},\widehat{\theta}\right)\left(\theta-\widehat{\theta}\right)\right] \quad (6) \\ \mathrm{Var}\left(\widehat{\beta}\right) &= \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} + \left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} \\ & X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\theta\left(\widehat{\beta},\widehat{\theta}\right) V_\theta^{-1} X_\theta'\left(\widehat{\beta},\widehat{\theta}\right) V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\left[X_\beta\left(\widehat{\beta},\widehat{\theta}\right)' V^{-1} X_\beta\left(\widehat{\beta},\widehat{\theta}\right)\right]^{-1} \end{aligned}$$

In the first line of (6) above, we used the fifth line from (5), accounting for the contribution of the estimation error in that first step, $X_\theta\left(\widehat{\beta},\widehat{\theta}\right)\left(\theta-\widehat{\theta}\right)$. Compare Amemiya (1985), section 10.4.3, Murphy and Topel (1985, p. 374-375), or Cameron and Trivedi (2005, section 6.6).

In practice, the only component we still have to derive for obtaining $\mathrm{Var}\left(\widehat{\beta}\right)$ in (6) above is the matrix of derivatives $X_\theta\left(\widehat{\beta},\widehat{\theta}\right)$. All other components have been stored while estimating $\mathrm{Var}\left(\widehat{\beta}|\widehat{\theta}\right)$ in (5), while V_θ comes stored from the first-step SML estimation in the paper.

The derivation of the matrix $X_\theta\left(\widehat{\beta},\widehat{\theta}\right)$ involves first simplifying the components of $X_\theta\left(\widehat{\beta},\widehat{\theta}\right)$, and then applying standard numerical differentiation on the remaining terms.

3.2 Deriving explicit matrix form and simplifying $X_\theta\left(\widehat{\beta},\widehat{\theta}\right)$

We want to find the derivative of X with respect to the parameter vector $\theta = \begin{bmatrix}\pi_0\\ \pi_1\\ \omega_0\\ \omega_1\end{bmatrix}$. Write again $X(\beta,\theta)$:

$$\begin{aligned} \Delta w_t &= \mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_\tau + \varepsilon_t \\ w_t - w_{t-1}^* &= \mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_\Theta^* + \overline{\sigma}\Omega_0 + \zeta_t \\ \Delta w_t^2 &= \sigma_z^2 + 2\sigma_u^2 + \overline{\gamma}^2\overline{\sigma}^2\mathrm{Var}\Delta\Omega_\tau + \left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_\tau\right)^2 + \eta_t \\ \left(w_t - w_{t-1}^*\right)^2 &= \sigma_z^2 + 2\sigma_u^2 + \overline{\gamma}^2\overline{\sigma}^2\mathrm{Var}\Delta\Omega_\Theta^* + \left(\mu_0 + \gamma\overline{\sigma}\pi S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_\Theta^* + \overline{\sigma}\Omega_0\right)^2 + \nu_t \\ \Delta w_t\Delta w_{t-1} &= -\sigma_u^2 + \left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_\tau\right)\left(\mu_0 + \gamma\overline{\sigma}\pi_1 S + \overline{\gamma\sigma}\mathrm{E}\Delta\Omega_{\tau-1}\right) + \upsilon_t \end{aligned}$$

The form of the derivative is a 5x4 matrix (in this case where we use 5 equations, adjusted accordingly in the specifications where we use more equations, as, e.g., where we allow heterogeneity between movers and non-movers, see paper), where each of the 5 horizontal rows represent blocks of equation-rows in the data. For clarity and brevity, it is easiest to write this matrix as a block matrix formed of 4 adjacent column vectors that will contain the derivatives with respect to each of the 4 components of θ.

$$X_\theta(\beta,\theta) = \left[\begin{array}{cccc} X\pi_0(\beta,\theta) & X\pi_1(\beta,\theta) & X\omega_0(\beta,\theta) & X\omega_1(\beta,\theta)\end{array}\right]$$

Below we explicitly write down each of these 4 vector components, preparing them for the final format in which they will be computed.

$X\pi_0(\beta,\theta) =$

$$\begin{bmatrix}
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_0} \\
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi_0} \\
\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \pi_0} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1\right)\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_0} \\
\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \pi_0} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi_0} \\
(\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S)\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi_0} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_0}\right) + \overline{\gamma}^2\overline{\sigma}^2\left(\mathrm{E}\Delta\Omega_\tau\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi_0} + \mathrm{E}\Delta\Omega_{\tau-1}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_0}\right)
\end{bmatrix}$$

$X\pi_1(\beta,\theta) =$

$$\begin{bmatrix}
\gamma\overline{\sigma}S + \overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_1} \\
\gamma\overline{\sigma}\pi S + \overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi_1} \\
((\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \pi_1} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1\right)\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_1} + \\
2\gamma^2\overline{\sigma}^2\pi_1 S^2 + 2\mu_0\gamma\overline{\sigma}S + \gamma\overline{\sigma}^2\overline{\gamma}S\mathrm{E}\Delta\Omega_\tau)) \\
(\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \pi_1} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi_1} + \\
2\gamma^2\overline{\sigma}^2\pi_1 S^2 + 2\mu_0\gamma\overline{\sigma}S + \gamma\overline{\sigma}^2\overline{\gamma}S\mathrm{E}\Delta\Omega^*_\Theta + \overline{\sigma}^2\Omega_0\gamma S) \\
((\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S)\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi_1} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_1}\right) + \overline{\gamma}^2\overline{\sigma}^2\left(\mathrm{E}\Delta\Omega_\tau\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi_1} + \mathrm{E}\Delta\Omega_{\tau-1}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_1}\right) + \\
2\mu_0\gamma\overline{\sigma}S + 2\gamma^2\overline{\sigma}^2\pi_1 S^2 + \gamma\overline{\sigma}^2 S\overline{\gamma}\left(\mathrm{E}\Delta\Omega_{\tau-1} + \mathrm{E}\Delta\Omega_\tau\right))
\end{bmatrix}$$

$X\omega_0(\beta,\theta) =$

$$\begin{bmatrix}
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_0} \\
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \omega_0} + +\overline{\sigma}\frac{\partial \Omega_0}{\partial \omega_0} \\
\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \omega_0} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1\right)\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_0} \\
(\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \omega_0} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \omega_0} + \\
\left(2\overline{\sigma}\Omega_0 + \overline{\sigma}^2\overline{\gamma}\mathrm{E}\Delta\Omega^*_\Theta + 2\overline{\sigma}\mu_0 + 2\gamma\overline{\sigma}^2\pi_1 S\right)\frac{\partial \Omega_0}{\partial \omega_0}) \\
(\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S)\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \omega_0} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_0}\right) + \overline{\gamma}^2\overline{\sigma}^2\left(\mathrm{E}\Delta\Omega_\tau\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \omega_0} + \mathrm{E}\Delta\Omega_{\tau-1}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_0}\right)
\end{bmatrix}$$

$X\omega_1(\beta,\theta) =$

$$\begin{bmatrix}
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_1} \\
\overline{\gamma\sigma}\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \omega_1} + +\overline{\sigma}\frac{\partial \Omega_0}{\partial \omega_1} \\
\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \omega_1} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1\right)\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_1} \\
(\overline{\gamma}^2\overline{\sigma}^2\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \omega_1} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \omega_1} + \\
\left(2\overline{\sigma}\Omega_0 + \overline{\sigma}^2\overline{\gamma}\mathrm{E}\Delta\Omega^*_\Theta + 2\overline{\sigma}\mu_0 + 2\gamma\overline{\sigma}^2\pi_1 S\right)\frac{\partial \Omega_0}{\partial \omega_1}) \\
(\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S)\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \omega_1} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_1}\right) + \overline{\gamma}^2\overline{\sigma}^2\left(\mathrm{E}\Delta\Omega_\tau\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \omega_1} + \mathrm{E}\Delta\Omega_{\tau-1}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_1}\right)
\end{bmatrix}$$

Observations for further computations:

First, we can immediately compute $\frac{\partial \Omega_0}{\partial \omega_0} = \frac{\partial \Omega}{\partial \omega_0} = \exp(\omega_0 + \omega_1 S) = \Omega$, and $\frac{\partial \Omega_0}{\partial \omega_1} = \frac{\partial \Omega}{\partial \omega_1} = S \exp(\omega_0 + \omega_1 S) = S\Omega$. This simplifies the expressions $X\omega_0(\beta, \theta)$ and $X\omega_1(\beta, \theta)$above.

Next, we need to evaluate all $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi_0}$, $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \pi_0}$, $\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi_1}$, $\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \pi_1}$ and respectively $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \omega_0}$, $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \omega_0}$, $\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \omega_1}$, $\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \omega_1}$. Note that all those conditional expectations and variances are functions of Ω and π, hence we use the chain rule for the derivations in functions of the components π_0, π_1, ω_0 and ω_1. We can then write the vectors above as such. For this purpose note that for some general functions n and m of respectively Ω and π:

$\frac{\partial n(\pi)}{\partial \pi_0} = \frac{\partial n(\pi)}{\partial \pi} \frac{\partial \pi}{\partial \pi_0} = \frac{\partial n(\pi)}{\partial \pi}$

$\frac{\partial n(\pi)}{\partial \pi_1} = \frac{\partial n(\pi)}{\partial \pi} \frac{\partial \pi}{\partial \pi_1} = S \frac{\partial n(\pi)}{\partial \pi}$

$\frac{\partial m(\Omega)}{\partial \omega_0} = \frac{\partial m(\Omega)}{\partial \Omega} \frac{\partial \Omega}{\partial \omega_0} = \Omega \frac{\partial m(\Omega)}{\partial \Omega}$

$\frac{\partial m(\Omega)}{\partial \omega_1} = \frac{\partial m(\Omega)}{\partial \Omega} \frac{\partial \Omega}{\partial \omega_1} = S\Omega \frac{\partial m(\Omega)}{\partial \Omega}$

Then, the 4 column vector components from above can be written as follows:

$X\pi_0(\beta, \theta) =$

$$\begin{bmatrix} \overline{\gamma\sigma} \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi} \\ \overline{\gamma\sigma} \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi} \\ \overline{\gamma}^2\overline{\sigma}^2 \frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \pi} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1\right) \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi} \\ \overline{\gamma}^2\overline{\sigma}^2 \frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \pi} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right) \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi} \\ \left(\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S\right)\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi}\right) + \overline{\gamma}^2\overline{\sigma}^2 \left(\mathrm{E}\Delta\Omega_\tau \frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi} + \mathrm{E}\Delta\Omega_{\tau-1} \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi}\right) \end{bmatrix}$$

$X\pi_1(\beta, \theta) =$

$$\begin{bmatrix} \gamma\overline{\sigma} S + \overline{\gamma\sigma} S \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi} \\ \gamma\overline{\sigma}\pi S + \overline{\gamma\sigma} S \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi} \\ (\overline{\gamma}^2\overline{\sigma}^2 S \frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \pi} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1\right) S \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi} + \\ 2\gamma^2\overline{\sigma}^2\pi_1 S^2 + 2\mu_0\gamma\overline{\sigma} S + \gamma\overline{\sigma}^2\overline{\gamma} S \mathrm{E}\Delta\Omega_\tau) \\ (\overline{\gamma}^2\overline{\sigma}^2 S \frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \pi} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right) S \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \pi} + \\ 2\gamma^2\overline{\sigma}^2\pi_1 S^2 + 2\mu_0\gamma\overline{\sigma} S + \gamma\overline{\sigma}^2\overline{\gamma} S \mathrm{E}\Delta\Omega^*_\Theta + \overline{\sigma}^2\Omega_0\gamma S) \\ ((\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S) S \left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi}\right) + \overline{\gamma}^2\overline{\sigma}^2 S \left(\mathrm{E}\Delta\Omega_\tau \frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \pi} + \mathrm{E}\Delta\Omega_{\tau-1} \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \pi}\right) + \\ 2\mu_0\gamma\overline{\sigma} S + 2\gamma^2\overline{\sigma}^2\pi_1 S^2 + \gamma\overline{\sigma}^2 S\overline{\gamma}\left(\mathrm{E}\Delta\Omega_{\tau-1} + \mathrm{E}\Delta\Omega_\tau\right)) \end{bmatrix}$$

$X\omega_0(\beta, \theta) =$

$$\begin{bmatrix} \overline{\gamma\sigma}\Omega \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \Omega} \\ \overline{\gamma\sigma}\Omega \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \Omega} + +\overline{\sigma}\Omega \\ \overline{\gamma}^2\overline{\sigma}^2\Omega \frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial \Omega} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1\right)\Omega \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \Omega} \\ (\overline{\gamma}^2\overline{\sigma}^2\Omega \frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial \Omega} + \left(2\overline{\gamma}^2\overline{\sigma}^2 \mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma} S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)\Omega \frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial \Omega} + \\ \left(2\overline{\sigma}\Omega_0 + \overline{\sigma}^2\overline{\gamma} \mathrm{E}\Delta\Omega^*_\Theta + 2\overline{\sigma}\mu_0 + 2\gamma\overline{\sigma}^2\pi_1 S\right)\Omega) \\ \left(\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S\right)\Omega\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \Omega} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \Omega}\right) + \overline{\gamma}^2\overline{\sigma}^2\Omega\left(\mathrm{E}\Delta\Omega_\tau \frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial \Omega} + \mathrm{E}\Delta\Omega_{\tau-1} \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial \Omega}\right) \end{bmatrix}$$

$$X\omega_1(\beta,\theta) = \begin{bmatrix} \overline{\gamma\sigma}S\Omega\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega} \\ \overline{\gamma\sigma}S\Omega\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial\Omega} + \overline{\sigma}S\Omega \\ \overline{\gamma}^2\overline{\sigma}^2 S\Omega\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial\Omega} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega_\tau + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1\right)S\Omega\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega} \\ (\overline{\gamma}^2\overline{\sigma}^2 S\Omega\frac{\partial \mathrm{Var}\Delta\Omega^*_\Theta}{\partial\Omega} + \left(2\overline{\gamma}^2\overline{\sigma}^2\mathrm{E}\Delta\Omega^*_\Theta + 2\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}^2\overline{\gamma}S\pi_1 + \overline{\sigma}^2\Omega_0\overline{\gamma}\right)S\Omega\frac{\partial \mathrm{E}\Delta\Omega^*_\Theta}{\partial\Omega} + \\ \left(2\overline{\sigma}\Omega_0 + \overline{\sigma}^2\overline{\gamma}\mathrm{E}\Delta\Omega^*_\Theta + 2\overline{\sigma}\mu_0 + 2\gamma\overline{\sigma}^2\pi_1 S\right)S\Omega) \\ (\mu_0\overline{\gamma\sigma} + \gamma\overline{\sigma}\pi_1 S)S\Omega\left(\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial\Omega} + \frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega}\right) + \overline{\gamma}^2\overline{\sigma}^2 S\Omega\left(\mathrm{E}\Delta\Omega_\tau\frac{\partial \mathrm{E}\Delta\Omega_{\tau-1}}{\partial\Omega} + \mathrm{E}\Delta\Omega_{\tau-1}\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega}\right) \end{bmatrix}$$

3.3 Numerical differentiation

The remaining exercise, given the simplifications in the subsection above, is to evaluate $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega}$, $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\pi}$, $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial\Omega}$ and $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial\pi}$ at the values of the estimated parameters. With those quantities computed, we can fill in and store the matrix of derivatives $X_\theta(\widehat{\beta},\widehat{\theta})$.

In order to obtain the quantities from above, $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\Omega}$, $\frac{\partial \mathrm{E}\Delta\Omega_\tau}{\partial\pi}$, $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial\Omega}$ and $\frac{\partial \mathrm{Var}\Delta\Omega_\tau}{\partial\pi}$,we use numerical differentiation of $\mathrm{E}\Delta\Omega_\tau$ and $\mathrm{Var}\Delta\Omega_\tau$ with respect to Ω and respectively π; we need to do that separately for the completed spells case, and incomplete spells case, as in the case where we computed $\mathrm{E}\Delta\Omega_\tau$ and $\mathrm{Var}\Delta\Omega_\tau$ for the completed and incomplete spells cases in the paper. For the numerical differentiation, we use a standard forward difference rule for chosen very small h:

$$f'(x) \simeq \lim_{h->0}\frac{f(x+h)-f(x)}{h}$$

3.4 Impact of the two-step variance correction

After we perform the numerical differentiations above, we construct $X_\theta(\widehat{\beta},\widehat{\theta})$, and then we multiply all the relevant matrices as explained in (6) above, we obtain that *only from the 11th decimal digit onwards* are the elements in the FGNLS variance-covariance matrix affected by the adjustment. Hence, although the exercise of correcting the FGNLS estimated variances for the possible error in the first SML step estimation is required, in the end the practical upshot is that it does not matter at all. The *ex post* intuition for this fact is that all but one of the tenure distribution parameters have been estimated with very small standard errors in our first step, see Table 2, the "Large Sample" panel in our paper.