Thesis Modeling

Tate Huffman

12/05/2021

Model Setup

This first regression is minimizing the objective function

$$Q = \sum_{i=1}^{20} [P_i - \pi_i(\Omega)]^2$$

with respect to the parameters of Ω. The values of i correspond to the exit or promotion of the individual in a given time period: for $1 \leq i \leq 10$, this denotes an exit in year i, and for $11 \leq i \leq 20$, this denotes a promotion of this person in year $i - 10$ (i.e., $i = 12$ means a promotion for the individual in year 2). P_i is the observed probability of a worker exiting or promoting in period i.

$\pi_i(\Omega)$ is the predicted probability of an exit or promotion in period i, where $\Omega = (\alpha, \tau, \theta_0, \mu, \sigma)$. These variables, respectively, correspond to: the probability of a Type A (able) worker producing a good signal; the probability of a Type B (unable) worker producing a good signal; the ex ante probability of a new worker being Type A; the mean of the worker promotion threshold θ_u; and its standard deviation.

When simplified, we see the probability of an exit in period i as

$$\pi_i = [\theta_0 \alpha^{i-1}(1-\alpha) + (1-\theta_0)\tau^{i-1}(1-\tau)] \int_{\theta(i-1,i-1)}^{1} g(\theta_u; \mu, \sigma) d\theta_u$$

and the probability of a promotion in period i as

$$\pi_{i+10} = [\theta_0 \alpha^{i} + (1-\theta_0)\tau^{i}] \int_{\theta(i-1,i-1)}^{\theta(i,i)} g(\theta_u; \mu, \sigma) d\theta_u$$

where $g(\theta_u; \mu, \sigma)$ is the beta distribution of θ_u.

Initial Regression

For this first attempt at the regression, we use its simplest form. This means that the statistics we use to qualify as a good signal are not normalized to league, and in fact are generalized across leagues: a good hitting signal is defined as an OPS over .825, and good hitting signal is defined as an ERA under 3.75. In future regressions, this performance will have appearance minimums, use more representative statistics, and normalize performance to the league's level of offense, but for now, this is a very rough first pass.