

# Wage Estimation

Kai-Jyun Wang\*

Spring 2025

## 1. Model Specification

Let  $y_t^*$  be the latent log wage at time  $t$  and  $a_t$  be the age at time  $t$ .  $d_t$  is the dummy for working. The specification of the model is as follows:

$$\begin{aligned}y_t^* &= f(a_t; \beta) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \\ \epsilon_t &= \rho \epsilon_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma_\nu^2) \\ d_t &= \mathbb{1} \{y_t^* \geq g(a_t, z_t; \gamma) + \eta_t\}, \quad \eta_t \sim N(0, \sigma_\eta^2) \\ y_t &= y_t^* \times d_t.\end{aligned}$$

We take the assumption that  $\eta$  and  $\epsilon$  are independent. The  $\sigma_t^2$  must obey that  $\sigma_t^2 = \rho^2 \sigma_{t-1}^2 + \sigma_\nu^2$  for  $t > 1$ . Suppose  $a_1$  is the first age that enters the labor market. We need to estimate the variance of the first period in the data,  $\sigma_1^2$ .  $z_t$  indicates whether or not the individual has a child at time  $t$ .

## 2. Estimation Procedure

### 2.1. A Less Efficient Method

1. For  $t = 1$ , consider the contemporary log likelihood

$$\begin{aligned}L_c(\beta_t, \gamma_t, \sigma_t, \sigma_{\eta,t}) &= \sum_{i=1}^N d_i \left[ \log \Phi \left( \frac{y_t - g(a_t, z_t; \gamma_t)}{\sigma_{\eta,t}} \right) + \log \phi \left( \frac{y_t - f(a_t; \beta_t)}{\sigma_t} \right) - \log \sigma_t \right] \\ &\quad + (1 - d_i) \log \Phi \left( \frac{f(a_t; \beta_t) - g(a_t, z_t; \gamma_t)}{\sqrt{\sigma_t^2 + \sigma_{\eta,t}^2}} \right),\end{aligned}$$

where  $\Phi$  and  $\phi$  are the cumulative distribution function and the probability density function of the standard normal distribution, respectively. Estimate the  $\beta$  for period  $t = 1$  by maximizing the contemporary log likelihood with the first period data.

2. Using the estimated  $\beta$  in first step, compute the residuals  $y_t^* - f(a_t; \beta)$ . Regress the residuals on the lagged residuals without intercept to estimate  $\rho$  and  $\sigma_\nu$ .

Estimate the  $\beta, \gamma, \sigma_1, \sigma_\eta$  for male and female separately.

---

\*National Taiwan University, Department of Economics. Student ID: B11303072.

### 3. Some Notes

- The specification of  $g$  is non-parametric.  $g(a_t, z_t; \gamma) = \gamma_{a_t, z_t}$ . Each possible combination of  $a_t$  and  $z_t$  corresponds to a  $\gamma_{a_t, z_t}$ .
- Try different specifications of  $f$ , for instance, quadratic, cubic, or quartic.
- Estimate the naive model first, i.e., regress  $y_t$  on  $f(a_t; \beta)$  and use the coefficients as initial values for the  $\beta$  and  $\gamma$ .