Structural Estimation: Duration Dependence & Non-parametric Estimation

Christine Braun

Previously

- Up until now we have assumed jobs arrive at a poisson rate
  - the hazard rate is constant over the duration

$$h = \lambda[1 - G(w_R)]$$

- Does this seem like a reasonable assumption?

Previously

- Up until now we have assumed jobs arrive at a poisson rate
  - the hazard rate is constant over the duration

$$h = \lambda[1 - G(w_R)]$$

- Does this seem like a reasonable assumption? No
  - λ might change over the spell, there might be stigma, people might change their search effort
  - w_R might change over the spell, may lose unemployment benefits

Hazard Rate Definition

- **Definition:** Let f and F be the pdf and cdf of t, then the hazard (failure) rate is

$$h(t) = \lim_{dt \to 0} \frac{P(T \in [t, t + dt) | T \geq t)}{dt}$$
$$h(t) = \frac{f(t)}{1 - F(t)}$$

- Integrate both sides and solve for $F(t)$

$$\int_0^t h(u)\ du = \int_0^t \frac{f(u)}{1 - F(u)}\ du$$

$$F(t) = 1 - \exp\left(- \int_0^t h(u)\ du \right)$$

More Flexibility

- **Poisson Process:** $h(t) = h$, plugging into $F(t)$, gives exponential arrival times

$$F(t) = 1 - e^{-ht}$$

$$f(t) = he^{-ht}$$

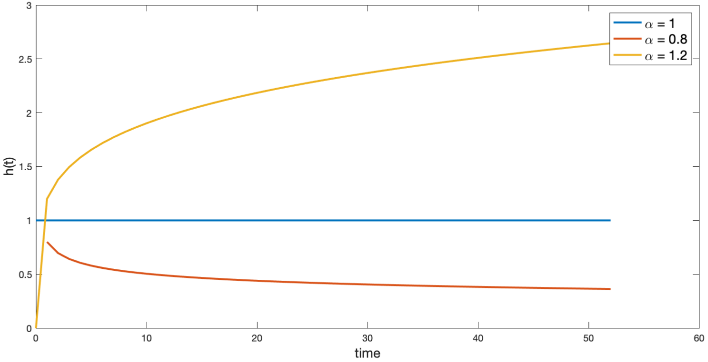
- **Weibull hazard:** $h(t) = \alpha t^{\alpha-1}$, plugging into $F(t)$, gives arrival times following a Weibull distribution

$$F(t) = 1 - e^{-t^{\alpha}}$$

$$f(t) = \alpha t^{\alpha-1} e^{-t^{\alpha}}$$

Duration Dependence

- With a hazard rate $\alpha t^{\alpha-1}$
  - $\alpha = 1$: $h(t)$ is flat (poisson process)
  - $\alpha < 1$: $h(t)$ is decreasing, negative duration dependence
  - $\alpha > 1$: $h(t)$ is increasing, positive duration dependence



MLE with Weibull hazard rate

- **Individual's Contribution:** Probability of observing a duration t

$$f(t_i;\alpha) = \alpha t_i^{\alpha-1} e^{-t_i^{\alpha}}$$

- **Log-Likelihood function:**

$$\begin{aligned}\mathcal{L}(\alpha;\{t_i\}) &= \sum_{i=1}^{N} \ln f(t_i;\alpha) \\ &= \sum_{i=1}^{N} \ln \alpha + (\alpha - 1)\ln t_i - t_i^{\alpha}\end{aligned}$$

Estimation in Matlab

- Using data3.csv
- File 1: SE2_main.m
    - read in data
    - extract just duration from data matrix
    - create lower bound and initial guess
    - estimate
- File 2: loglike3.m
    - inputs: parameters, duration
    - output: negative log-likelihood value

Weibull Hazard Answer

- Estimates and Standard Errors

Parameter	Estimate	Standard Error
α	0.5221	0.0005

- Log-Likelihood Value

$$logL = -2.6073e + 4$$

- Why do we get negative duration dependence?

Selection Effect

- Observable characteristics could affect the hazard rate
- **Example:** h_{he} is the hazard rate of high educated and h_{le} is the hazard rate of low educated, both constant over time
  - $h_{he} > h_{le}$
  - $u_{he}(t)$: fraction of high educated in pool of unemp.
  - $u_{le}(t)$: fraction of low educated in pool of unemp.

$$\Rightarrow h(t) = u_{he}(t) \times h_{he} + u_{le}(t) \times h_{le}$$

- If we estimate $h(t)$ without covariates we will get negative duration dependence because of a **selection effect**
  - high educated people leave unemp. first ($h_{he} > h_{le}$) so the average hazard rate decreases over time

Proportional Hazard Model

- Define the hazard as

$$h(t|x) = \psi(t) \times h_0(x)$$

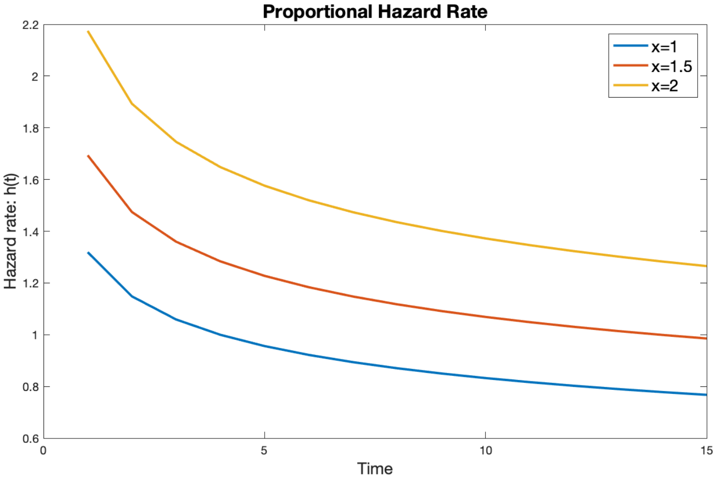
  $h_0(x)$ is called the *systematic part* and $\psi(t)$ is called the *baseline hazard*.

- The systematic part is commonly given an functional form assumption

$$h_0(x) = exp(x'\beta)$$

  covariates affect the hazard rate log-linearly. We then estimate β.

Proportional Hazard Model



Plotted: $h(t) = 0.8t^{0.8-1}exp(0.5x)$

Proportional Hazard Model

- Assume Weibull baseline

$$\psi(t) = \alpha t^{\alpha-1}$$

- Assume log-linear covariates

$$h_0(x) = \exp(x'\beta)$$

- The cdf of duration

$$F(t|x) = 1 - \exp\left(- \int_0^t \exp(x'\beta)\alpha u^{\alpha-1}\, du \right)$$

$$F(t|x) = 1 - \exp(-\exp(x'\beta)t^{\alpha})$$

- The pdf of duration

$$f(t|x) = \exp(x'\beta)\alpha t^{\alpha-1} e^{-\exp(x'\beta)t^{\alpha}}$$

MLE with Weibull baseline & Log-linear Covariates

- **Individual's Contribution:** Probability of observing a duration t

$$f(t_i|x_i;\alpha,\beta) = \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\exp(x_i'\beta)t_i^{\alpha}}$$

- **Log-Likelihood function:**

$$\begin{aligned}\mathcal{L}(\alpha,\beta;\{t_i\},\{x_i\}) &= \sum_{i=1}^{N} \ln f(t_i|x_i;\alpha,\beta) \\ &= \sum_{i=1}^{N} x_i'\beta + \ln\alpha + (\alpha-1)\ln t_i - \exp(x_i'\beta)t_i^{\alpha}\end{aligned}$$

Estimation in Matlab

- Using data3.csv
- File 1: SE2_main.m
  - create a vector x that contains a dummy for women
  - create lower bound and initial guess
  - estimate
- File 2: loglike4.m
  - inputs: parameters, duration, covariates
  - output: negative log-likelihood value

Weibull Hazard & Log-linear Covariates Answer

- Estimates and Standard Errors

Parameter	Estimate	Standard Error
α	0.5809	0.0025
β_{FE}	-0.5956	0.0345

- Log-Likelihood Value

$$logL = -2.5202e + 4$$

- What happened to the estimate of α?
- Let's add the education covariates

$$educDummy = dummyvar(\quad)$$

Weibull Hazard & Log-linear Covariates Answer

- Estimates and Standard Errors

Parameter	Estimate	Standard Error
α	0.6503	0.0038
β_{FE}	-0.3628	0.0067
β_{educ2}	-0.5817	0.0194
β_{educ3}	-0.5583	0.0044

- Log-Likelihood Value

$$logL = -2.4363e + 4$$

- What happened to the estimate of α and β_{FE}?
- Could we still have a selection effect?

Mixed Proportional Hazard Model

- Define the hazard rate as

$$h(t|x,\nu) = \nu \times \psi(t) \times h_0(x)$$

  - $\psi(t)$: baseline hazard
  - $h_0(x)$: systematic part
  - ν: unobserved heterogeneity, "error term"

- $\nu \sim G(\nu)$ where G is called the mixing distribution

  - can make a parametric assumption (usually Gamma)

  - can estimate non-parametrically

Mixed Proportional Hazard Model

- Assume Weibull baseline

$$\psi(t) = \alpha t^{\alpha-1}$$

- Assume log-linear covariates

$$h_0(x) = \exp(x'\beta)$$

- Assume a there exists a mixing distribution $G(\nu)$
- The cdf of duration

$$F(t|x,\nu) = 1 - \exp(-\nu \exp(x'\beta)t^{\alpha})$$

- The pdf of duration

$$f(t|x,\nu) = \nu \exp(x'\beta)\alpha t^{\alpha-1} e^{-\nu \exp(x'\beta)t^{\alpha}}$$

Parametric Estimation

- Parametric estimation of mixing distribution
  - Choose $G(\nu;\theta)$ with support $[0,\infty)$ and parameters θ
  - Integrate out of duration pdf

$$f(t|x) = \int_0^\infty f(t|x,\nu) \times g(\nu)\ d\nu$$

  - This is often a difficult integral ($\nu \sim$ Gamma has a closed-form solution)
  - We would get an MLE of θ
  - Heckman & Stinger (1984) show instability of parameter estimates depending on the assumptions on the mixing distribution

Non-Parametric Estimation

- Non-Parametric estimation of mixing distribution
  - We discretize G
  - $\{\nu_j\}_{j=1}^K$: set of points in G
  - $\{\pi_j\}_{j=1}^K$: the probability of point j
- Sum over the points to get the full distribution of durations

$$f(t|x) = \sum_{j=1}^{K} \pi_j \times f(t|x, \nu_j)$$

- The likelihood function we be a function of $\{\nu_j\}_{j=1}^K$ and $\{\pi_j\}_{j=1}^K$ and we get ML estimates of each point and it's probability.

Non-Parametric Estimation: Example

- Let's estimate with $K = 2$
- **Individual's Contribution:** Probability of observing a duration t

$$f(t_i|x_i;\alpha,\beta,\nu_1) = \nu_1 \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\nu_1 \exp(x_i'\beta)t_i^{\alpha}}$$
$$f(t_i|x_i;\alpha,\beta,\nu_2) = \nu_2 \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\nu_2 \exp(x_i'\beta)t_i^{\alpha}}$$

- **Log-Likelihood function:**

$$\mathcal{L}(\alpha,\beta,\{\nu_j\},\{\pi_j\};\{t_i\},\{x_i\}) = \sum_{i=1}^{N} \ln[\pi_1 \times f(t_i|x_i;\alpha,\beta,\nu_1) + \pi_2 \times f(t_i|x_i;\alpha,\beta,\nu_2)]$$

Non-Parametric Estimation: Example

- Maximize $\mathcal{L}(\alpha, \beta, \{\nu_j\}, \{\pi_j\}; \{t_i\}, \{x_i\})$ with respect to
  - $\alpha > 0$
  - β: no restrictions
  - ν_1, ν_2, all > 0
  - π_1, $\pi_2 \in [0, 1]$
- Subject to $\pi_1 + \pi_2 = 1$

Syntax

```
x = fmincon(fun,x0,A,b)
x = fmincon(fun,x0,A,b,Aeq,beq)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon,options)
x = fmincon(problem)
[x,fval] = fmincon( __ )
[x,fval,exitflag,output] = fmincon( __ )
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon( __ )
```

Description

Nonlinear programming solver.

Finds the minimum of a problem specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases}$$

Estimation in Matlab

- Using data3.csv
- File 1: SE2_main.m
  - create lower bound and initial guess
  - create Aeq (1×8) and beq (1×1)
  - estimate
- File 2: loglike5.m
  - inputs: parameters, duration, covariates
  - output: negative log-likelihood value

Estimation Answer

- Estimates and Standard Errors

Parameter	Estimate	Standard Error
α	0.8854	0.1226
ν_1	0.0936	0.0373
ν_2	0.3795	0.0182
π_1	0.0807	0.1211
π_2	0.9193	1.2941
β_{FE}	0.0597	0.2088
β_{educ2}	0.0069	0.3952
β_{educ3}	0.0276	0.1594

- Log-Likelihood Value

$$logL = -2.2976e+4$$

- What happened to α and β?

Estimation in Matlab

- Let's estimate with $K = 3$

- Use the same likelihood function but add another point in the mixing distribution

$$f(t_i|x_i;\alpha,\beta,\nu_1) = \nu_1 \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\nu_1 \exp(x_i'\beta) t_i^{\alpha}}$$
$$f(t_i|x_i;\alpha,\beta,\nu_2) = \nu_2 \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\nu_2 \exp(x_i'\beta) t_i^{\alpha}}$$
$$f(t_i|x_i;\alpha,\beta,\nu_3) = \nu_2 \exp(x_i'\beta)\alpha t_i^{\alpha-1} e^{-\nu_3 \exp(x_i'\beta) t_i^{\alpha}}$$

Estimation Answer

- Estimates and Standard Errors

Parameter	Estimate	Standard Error
α	0.9810	0.0166
ν_1	0.0399	0.0394
ν_2	0.2005	0.0888
ν_2	0.6037	0.2449
π_1	0.0266	0.0493
π_2	0.5168	1.4205
π_3	0.4566	0.2708
β_{FE}	0.0713	0.0695
β_{educ2}	0.0008	0.2776
β_{educ3}	0.0267	0.0334

- Log-Likelihood Value

$$logL = -2.2945e + 4$$

- What happened to α and β?

How may points should we estimate?

- Adding points will improve fit
- Adding too many points is computationally costly
- Use likelihood ratio test to find best K
  - test goodness of fit of two competing models, one is a restricted version of the other
  - stop adding points when the information gained from $K + 1$ points is not statistically significant

Likelihood Ratio Test

- **Unrestricted model:** parameter space is Θ

$$\max_{\theta \in \Theta} L(\theta)$$

  where $rank(\theta) = r$

- **Restricted model:** constrained parameter space is Θ_0

$$\max_{\theta \in \Theta_0} L(\theta)$$

  where $rank(\theta) = r - q$

- **Likelihood-ratio test statistic:**

$$\lambda_{LR} = -2 \ln \left[\frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right]$$

  where $\lambda_{LR} \to \chi^2(q)$

Likelihood Ratio Test: Example

- **Unrestricted model:** the model where $K = 3$,

$$\theta^U = \{\alpha, \beta_{FE}, \beta_{educ1}, \beta_{educ2}, \nu_1, \nu_2, \nu_3, \pi_1, \pi_2, \pi_3\}$$

$$rank(\theta^U) = 10$$

$$\ln \max_{\theta \in \Theta} L(\theta) = -2.2945e + 4$$

- **Restricted model:** the model where $K = 2$, where we restricted $\nu_3 = 0$ and $\pi_3 = 0$

$$\theta^R = \{\alpha, \beta_{FE}, \beta_{educ1}, \beta_{educ2}, \nu_1, \nu_2, \pi_1, \pi_2\}$$

$$rank(\theta^R) = 8$$

$$\ln \max_{\theta \in \Theta} L(\theta) = -2.2976e + 4$$

Likelihood Ratio Test: Example

- **Likelihood-ratio test statistic:**

$$\lambda_{LR} = -2[-2.2976e + 4 - (-2.2945e + 4)] = 61.9539$$

- **P-value:** Probability that a chi-squared RV with 2 degrees of freedom is larger than 61.9539

$$1 - chi2cdf(61.9539, 2) = 3.5194e - 14$$

so we reject the null hypothesis, i.e. the restricted model. $K = 3$ points is statistically significantly better than $K = 2$.

- Keep estimating by adding one more point until we fail to reject restricted model.

So do we have duration dependence?

- We need a lot of data to estimate a good mixing distribution
- Can not tell if negative duration dependence is selection driven or structural
- Kroft, Lange, Notowidigdo (2013): investigate employer behavior in duration dependence
  - send out many fake resumes
  - vary the length of unemployment duration
  - show call-back rate decrease with unemployment duration