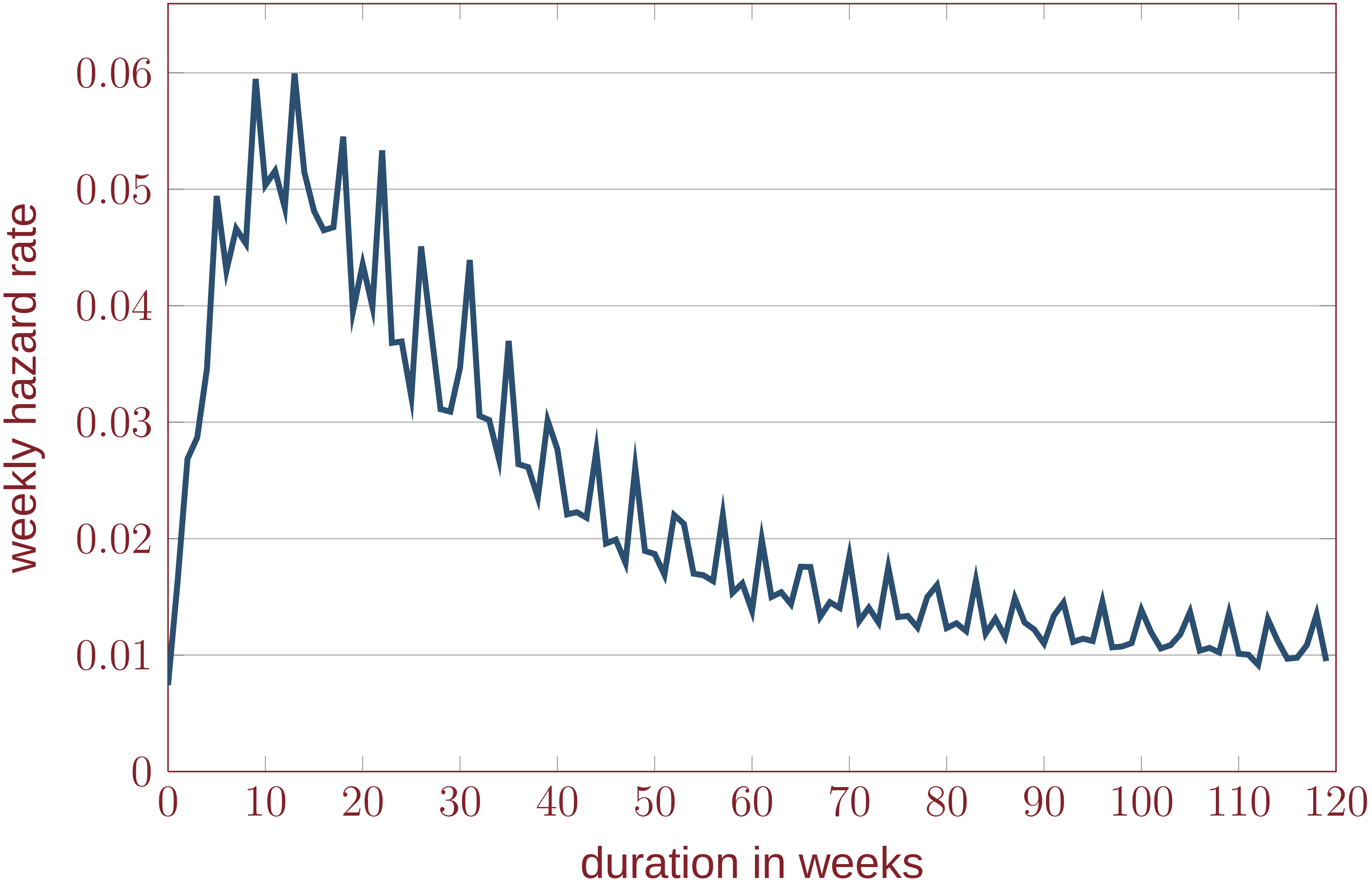
Decomposing Duration Dependence in a Stopping Time Model

Fernando Alvarez Katarína Borovičková Robert Shimer

June 13, 2015

Non-employment Exit Hazard



Introduction

- why do hazard rates decrease with duration?
  - heterogeneity across workers
  - structural duration dependence for each worker

- answer using a structural stopping time model
  - log net benefit of employment is a Brownian motion with drift
  - fixed cost of switching employment status
  - distribution of completed duration is inverse Gaussian (α, β)
  - parameters of the distribution differ across individuals, $G(\alpha, \beta)$

- identify G using distribution of two completed spells, density $\phi(t_1, t_2)$
  - estimate with Austrian administrative panel, ~ 1 million workers

Summary

- economically-motivated assumptions about individual durations
  - individual duration follows inverse Gaussian w/parameters (α, β)
  - arbitrary heterogeneity across individuals, w/distribution $G(\alpha, \beta)$
- identify the joint distribution of individual parameters G
  - identification relies on functional form of inverse Gaussian
  - test whether ϕ can be generated by any such a model
- estimate using observations on non-employment duration
  - hazard rate decomposition: “structure” vs “heterogeneity”
  - estimate small upper bound on fixed switching cost
  - model w/zero switching cost: soundly rejected

Stopping Time Literature

- Lancaster (1972), Newby and Winterton (1983): strikes
- Aalen-Gjessing (2001), Lee-Whitmore (2006, 2010): death
- Alvarez-Shimer (2009), (2011): theory
- Buhai-Teulings (2014): job tenure, no heterogneity
- Shimer (2008): no heterogeneity, no switching costs
- Abbring (2012): Lévy process, heterogeneity only in boundaries

Structural Model

Assumptions

- risk-neutral worker, discount rate r in continuous time
- worker can be either employed, $s(t) = e$, or nonemployed, $s(t) = n$
  - employed: wage $e^{w(t)}$, $dw(t) = \mu_{w,s(t)} dt + \sigma_{w,s(t)} dB_w(t)$
  - nonemployed: benefits $b_0\, e^{b(t)}$, $db(t) = \mu_{b,s(t)} dt + \sigma_{b,s(t)} dB_b(t)$
  - Brownian motions have correlation $\rho_{s(t)}$
- switching from nonemployment to employment: cost $\psi_e\, b_0\, e^{b(t)}$
- switching from employment to nonemployment: cost $\psi_n\, b_0\, e^{b(t)}$
- decision rules depend only on "net" log benefit of employment:

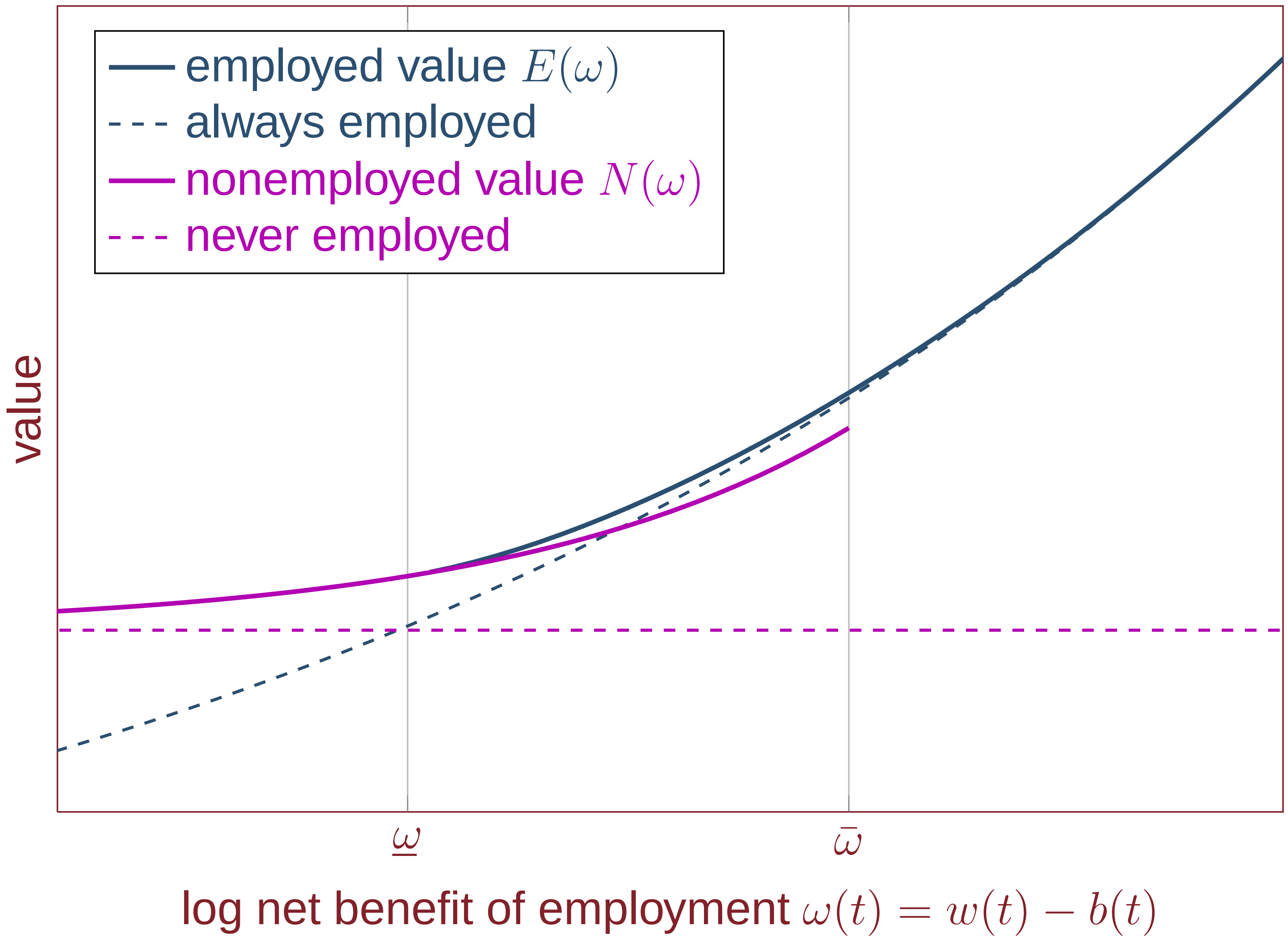
$$\omega(t) \equiv w(t) - b(t) \text{ with } d\omega(t) = \mu_{s(t)} dt + \sigma_{s(t)} dB(t)$$

Alternative Interpretation

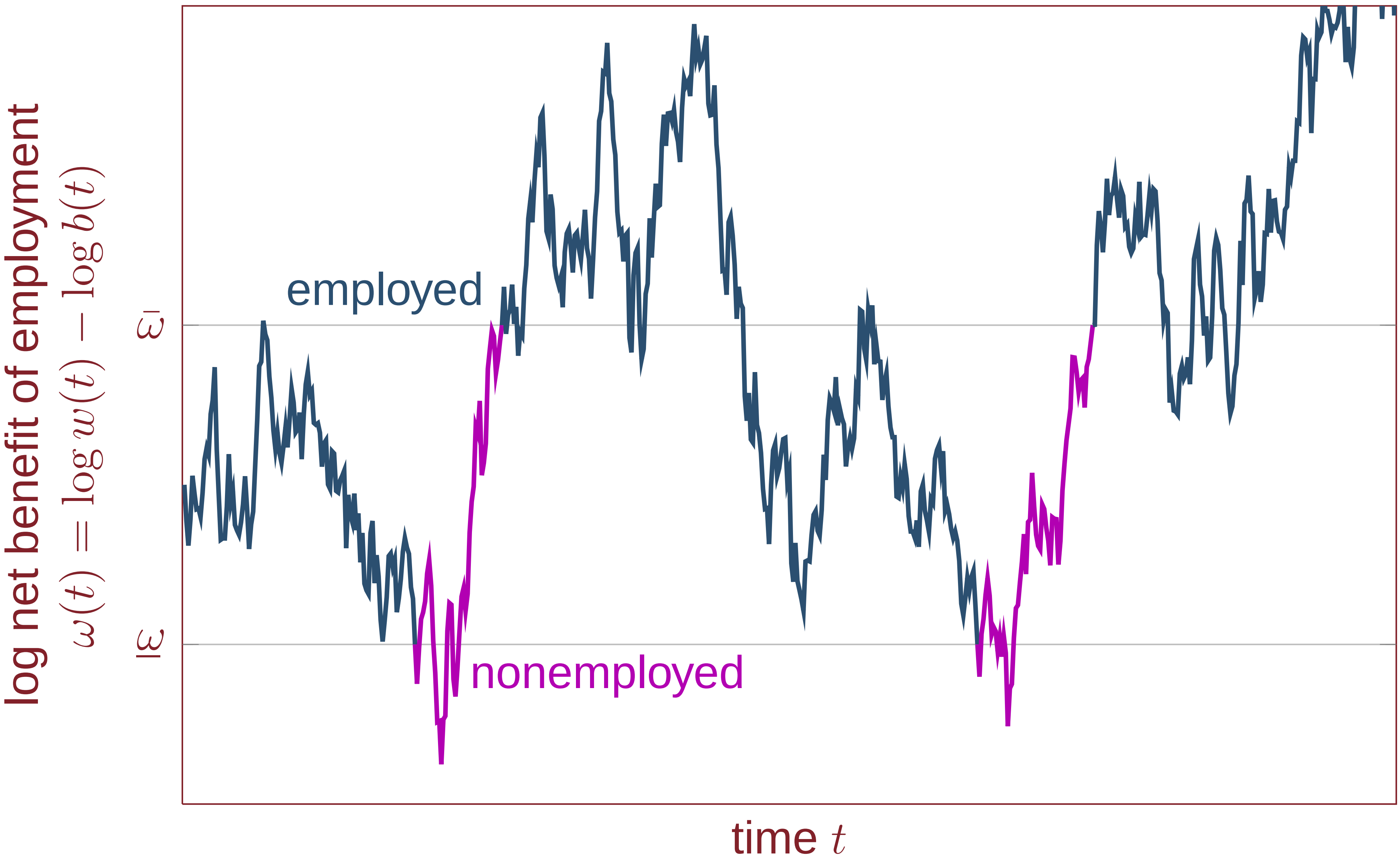
- worker's log productivity is $w(t)$ and log wage is $b(t)$
- monopsonist can hire worker at fixed cost $\psi_e \, b_0 \, e^{b(t)}$
- monopsonist can fire worker at fixed cost $\psi_n \, b_0 \, e^{b(t)}$
- monopsonist discounts the future at rate r

- same Bellman equations (differ by a constant)
  - our approach cannot tell if nonemployment is voluntary

Value Functions



Example of Sample Path



Determinants of Barriers

- assume non-employment benefits constant: $\mu_{b,s} = \sigma_{b,s} = 0$
- accurate approximation for width of inaction:

$$(\bar{\omega} - \underline{\omega})^3 \approx \frac{12\, r\, \sigma_e^2 \sigma_n^2}{(\mu_e + \sqrt{\mu_e^2 + 2r\sigma_e^2})(-\mu_n + \sqrt{\mu_n^2 + 2r\sigma_n^2})} \frac{\psi_e + \psi_n}{b_0}$$

  - uncertainty σ_n, σ_e widens inaction range (option value)
  - increase in μ_n or decrease μ_e widens inaction
  - large sensitivity of inaction range to cost, cubic root

Inverse Gaussian Distribution

- nonemployment duration is given by an inverse Gaussian distribution

$$f(t;\alpha,\beta) = \frac{\beta}{\sqrt{2\,\pi}\,t^{3/2}} \exp\left(-\frac{(\alpha t-\beta)^2}{2t}\right)$$

  where $\alpha = \mu_n/\sigma_n$ and $\beta = (\bar{\omega} - \underline{\omega})/\sigma_n$

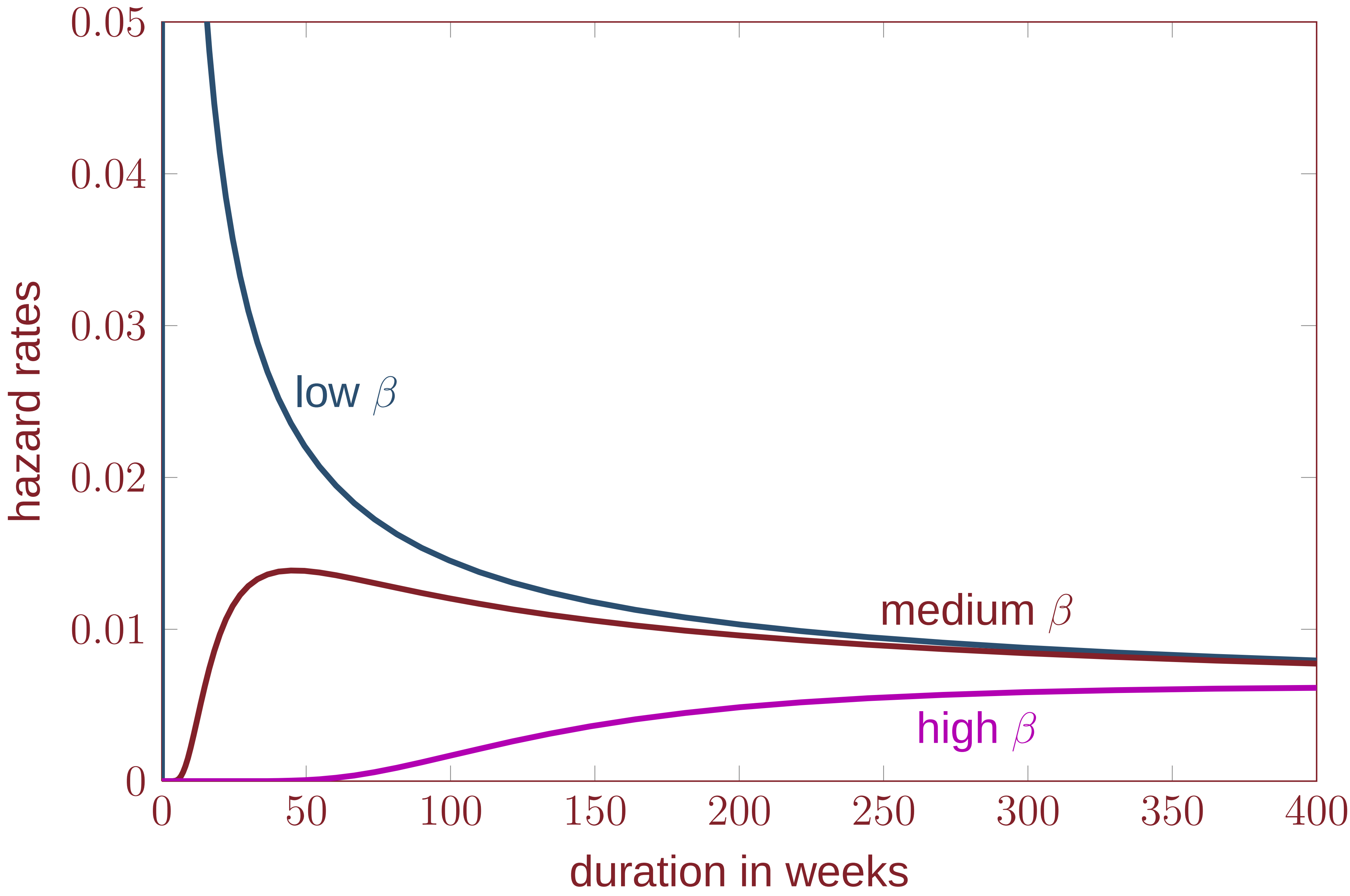
  - $\mu_n = \mu_{w,n} - \mu_{b,n}$ is the drift in ω while nonemployed
  - $\sigma_n^2 = \sigma_{w,n}^2 - 2\rho_n\sigma_{w,n}\sigma_{b,n} + \sigma_{b,n}^2$ is its variance

- structural parameters all determine $\bar{\omega}$ and $\underline{\omega}$

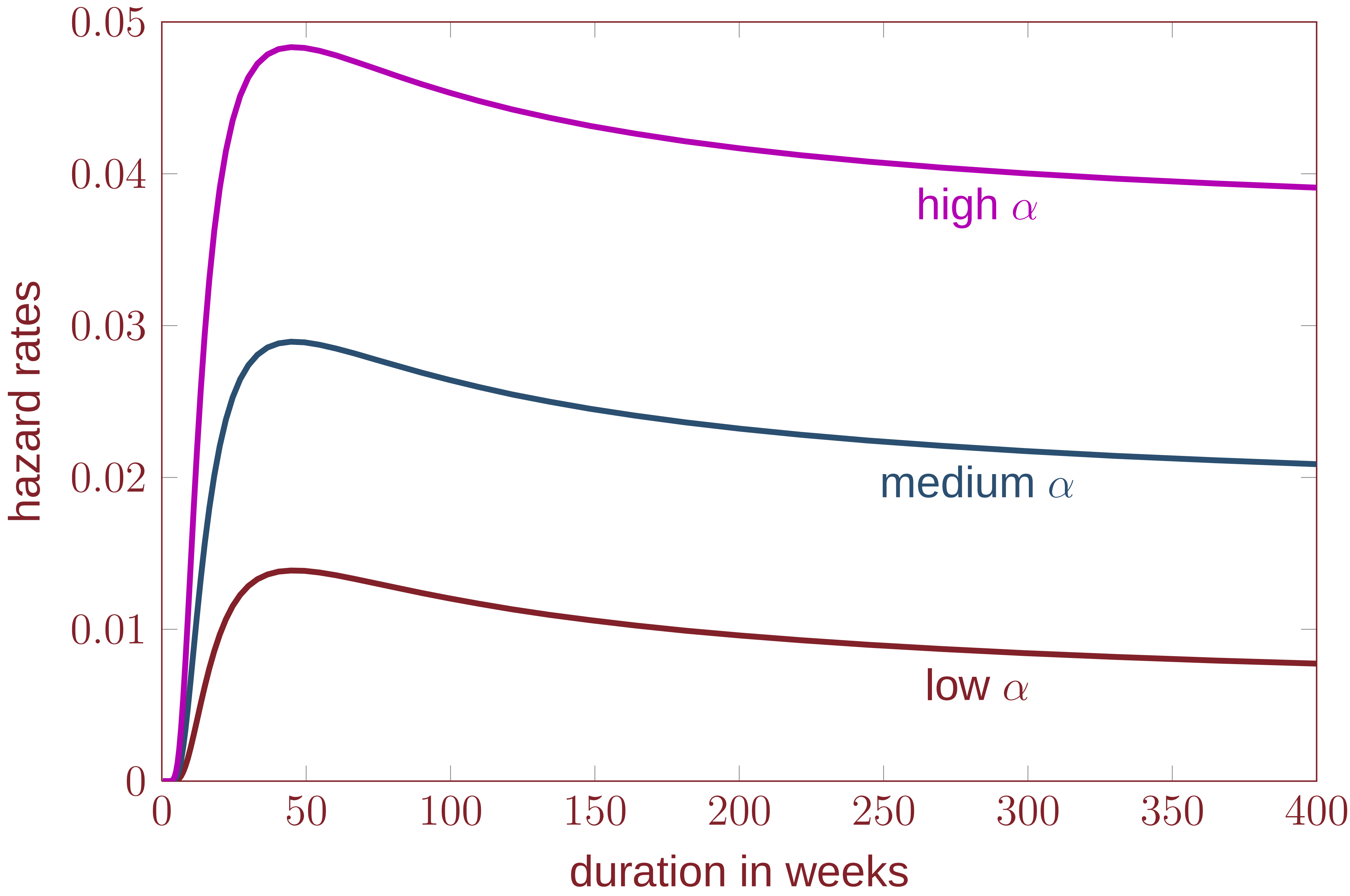
- worker eventually finds a job if and only if $\alpha \geq 0$

- parameter heterogeneity maps into cumulative distribution $G(\alpha,\beta)$

Hazard Rates



Hazard Rates



Identification and Testing

Idea

- input: data on non-employment duration – distribution ϕ
- goal: recover distribution $G(\alpha, \beta)$
- model implies a mapping $\phi = M(G)$
- identification
  - for each ϕ, there is at most one G such that $\phi = M(G)$
- testing
  - there exist ϕ's for which $\phi \neq M(G)$ for all G's

Idea

- input: data on non-employment duration – distribution ϕ
- goal: recover distribution $G(\alpha, \beta)$
- model implies a mapping $\phi = M(G)$
- identification
  - for each ϕ, there is at most one G such that $\phi = M(G)$
- testing
  - there exist ϕ's for which $\phi \neq M(G)$ for all G's
- model is not identified with data on one spell per individual
- model is identified with two spells per individual

Single Nonemployment Spell

- model is not identified with a single nonemployment spell
- generate data from a model with one type (α, β)
- alternative way to fit the model:
  - individual i with duration d: $\sigma_n^i = 0$, $\mu_n^i = (\bar{\omega}^i - \underline{\omega}^i)/d$
- special cases of the model are identified with one spell:
  - everyone has the same expected duration
  - no switching costs, so $\bar{\omega} = \underline{\omega}$

Two Nonemployment Spells

- fix a nonempty, open set of completed durations $T \subseteq \mathbb{R}$

- the joint density of duration of two spells $(t_1, t_2) \in T^2$ is given by

$$\phi(t_1, t_2) = \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)$$

Two Nonemployment Spells

- fix a nonempty, open set of completed durations $T \subseteq \mathbb{R}$

- the joint density of duration of two spells $(t_1, t_2) \in T^2$ is given by

$$\phi(t_1, t_2) = \frac{\iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\iint_{T^2} \iint f(t_1'; \alpha, \beta) f(t_2'; \alpha, \beta) dG(\alpha, \beta)\, dt_1'\, dt_2'}$$

- defines a linear map: $\phi = M(G)$

Sketch of the Identification Proof

1. ϕ is ∞-differentiable, except at $t_1 = t_2$

2. k^{th} derivatives of ϕ identify k^{th} conditional moments of (α^2, β^2)

    - unique moments of (α^2, β^2) conditional on surviving to (t_1, t_2)
    - unique distribution (α^2, β^2) conditional on surviving to (t_1, t_2)

3. extend to unconditional distribution of (α^2, β^2) (Bayes rule)

Sketch of the Identification Proof

1. ϕ is ∞-differentiable, except at $t_1 = t_2$
2. k^{th} derivatives of ϕ identify k^{th} conditional moments of (α^2, β^2)
   - unique moments of (α^2, β^2) conditional on surviving to (t_1, t_2)
   - unique distribution (α^2, β^2) conditional on surviving to (t_1, t_2)
3. extend to unconditional distribution of (α^2, β^2) (Bayes rule)

- requires only *local* information, i.e. ϕ and all derivatives at a point
- does not require existence of any moments of ϕ
- does not imply existence of any moments of (α, β)

A Limitation to Identification

- recall $f(t;\alpha,\beta) = \dfrac{\beta}{\sqrt{2\pi}\,t^{3/2}} \exp\left(-\dfrac{(\alpha t-\beta)^2}{2t}\right)$
- easy to prove $f(t;\alpha,\beta) = e^{2\alpha\beta} f(t;-\alpha,\beta)$ for all (α,β)
- complete spells only identify distribution up to sign of α

A Limitation to Identification

- recall $f(t;\alpha,\beta) = \dfrac{\beta}{\sqrt{2\pi}\,t^{3/2}} \exp\left(-\dfrac{(\alpha t-\beta)^2}{2t}\right)$
- easy to prove $f(t;\alpha,\beta) = e^{2\alpha\beta} f(t;-\alpha,\beta)$ for all (α,β)
- complete spells only identify distribution up to sign of α
- let G^+ w/density g^+ have all mass on $\alpha > 0$
- define G^-: $g^-(-\alpha,\beta) = e^{4\alpha\beta} g^+(\alpha,\beta)$
- then $M(G^+) = M(G^-)$
- so is any convex combination of $g^+(\alpha,\beta)$ and $g^-(\alpha,\beta)$

A Limitation to Identification

- recall $f(t;\alpha,\beta) = \dfrac{\beta}{\sqrt{2\pi}\, t^{3/2}} \exp\left(-\dfrac{(\alpha t - \beta)^2}{2t}\right)$
- easy to prove $f(t;\alpha,\beta) = e^{2\alpha\beta} f(t;-\alpha,\beta)$ for all (α,β)
- complete spells only identify distribution up to sign of α
- let G^+ w/density g^+ have all mass on $\alpha > 0$
- define G^-: $g^-(-\alpha,\beta) = e^{4\alpha\beta} g^+(\alpha,\beta)$
- then $M(G^+) = M(G^-)$
- so is any convex combination of $g^+(\alpha,\beta)$ and $g^-(\alpha,\beta)$
- proceed as if $\alpha \geq 0$ for all individuals
- incomplete spells provide partial information on sign α

Incomplete Spells and Sign of α

- G^+ distribution consistent w/ ϕ with $\alpha \geq 0$
- we identify G^+ from distribution of two complete spells
- $c \equiv$ fraction of two consecutive complete spells in $T \subset \mathbb{R}^2$:

$$c = \iint_{(t_1,t_2)\in T^2} \iint f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta) dt_1 dt_2$$

- we can measure $\hat{c}$ in actual data
- G^+ or G^- are likely not to be consistent with $\hat{c}$
- define $\bar{G}, \underline{G}$ consistent with ϕ and $\hat{c}$:
  - distribution $\bar{G}$ w/largest number of negative α
  - distribution $\underline{G}$ w/smallest number of negative α

Test: Conditional Moments of α^2, β^2

- differentiate $\phi(t_1, t_2)$ w.r.t. t_1 and t_2

$$\frac{\phi_i(t_1,t_2)}{\phi(t_1,t_2)} = \frac{1}{2}\left[\frac{1}{t_i^2}\mathbb{E}(\beta^2|t_1,t_2) - \frac{3}{t_i} - \mathbb{E}(\alpha^2|t_1,t_2)\right] \text{ for } i = 1, 2$$

$$\text{where } \mathbb{E}(\alpha^2|t_1,t_2) = \frac{\iint \alpha^2 f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}{\iint f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}$$

$$\text{and } \mathbb{E}(\beta^2|t_1,t_2) = \frac{\iint \beta^2 f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}{\iint f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}$$

- if $t_1 \neq t_2$, gives 2 linear equations in $\mathbb{E}(\alpha^2|t_1,t_2), \mathbb{E}(\beta^2|t_1,t_2)$
- test: these moments are non-negative
- higher order partial derivatives give more tests

Power of Test: Constant Hazard

- data generating process: constant hazard rate of finding a job h

- observe $\phi(t_1, t_2) = h^2 e^{-h(t_1+t_2)}$ and perform our test

$$\mathbb{E}(\alpha^2|t_1, t_2) = 2 - \frac{3}{t_1 + t_2} > 0 \Leftrightarrow t_1 + t_2 > \frac{3}{2h}$$

and $\mathbb{E}(\beta^2|t_1, t_2) = \frac{3t_1 t_2}{t_1 + t_2} > 0$

- conclusion: data was not generated by our model

- paper extends to case of distribution of constant hazard rates h

Power of Test: Log Normal

- data generating process: log-normal distribution (μ, σ)

- observe $\phi(t_1, t_2)$ and perform our test

$$\mathbb{E}(\alpha^2|t_1, t_2) = \frac{2}{\sigma^2(t_1+t_2)}\left(\frac{t_1 \log t_1 - t_2 \log t_2}{t_1 - t_2} - (\mu + \tfrac{1}{2}\sigma^2)\right)$$

and $\mathbb{E}(\beta^2|t_1, t_2) = \frac{2t_1t_2}{\sigma^2(t_1+t_2)}\left(\frac{t_2 \log t_1 - t_1 \log t_2}{t_1 - t_2} + (\mu + \tfrac{1}{2}\sigma^2)\right).$

  - $\mathbb{E}(\alpha^2|t_1, t_2)$ is negative at small (t_1, t_2), e.g. $t_1 = t_2 < e^{\mu + \frac{1}{2}\sigma^2 - 1}$
  - $\mathbb{E}(\beta^2|t_1, t_2)$ is negative at large (t_1, t_2), e.g. $t_1 = t_2 > e^{\mu + \frac{1}{2}\sigma^2 + 1}$

- conclusion: data was not generated by our model

Decompositions

Notation

- type-specific hazard rate $h(t;\alpha,\beta) = \dfrac{f(t;\alpha,\beta)}{1-F(t;\alpha,\beta)}$
- time-t survivor distribution $dG(\alpha,\beta;t) = \dfrac{(1-F(t;\alpha,\beta))dG(\alpha,\beta)}{\int\int(1-F(t;\alpha',\beta'))dG(\alpha',\beta')}$
  - note: $dG(\alpha,\beta;0) = dG(\alpha,\beta)$
- aggregate hazard is $H(t) = \int\int h(t;\alpha,\beta)dG(\alpha,\beta;t)$

Hazard Rate Decomposition

- hazard rate $\dot{H}(t) = \dot{H}^s(t) + \dot{H}^h(t)$ where

$$\dot{H}^s(t) = \iint \dot{h}(t;\alpha,\beta)dG(\alpha,\beta;t)$$

$$\dot{H}^h(t) = \iint h(t;\alpha,\beta)d\dot{G}(\alpha,\beta;t)$$

$$= -\iint (h(t;\alpha,\beta) - H(t))^2 dG(\alpha,\beta;t) < 0$$

- this is R.A. Fisher’s fundamental theorem of natural selection
  - “the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time”

Austrian Data

Description

- universe of private sector employees, 1986–2007
  - social security records
  - observe employment, unemployment, retirement, maternity leave
  - full-time, part-time, “marginal” jobs
  - start and end date for each spell
- definition of non-employment spell:
  - end of one full-time job to start of next full-time job (in weeks)
  - registered as unemployed at some point (avoid job-to-job)
- only spells if individual > 25 at the time (avoid school)
- age criterium individual is (potentially) for at least 15 yrs in sample

Data Construction

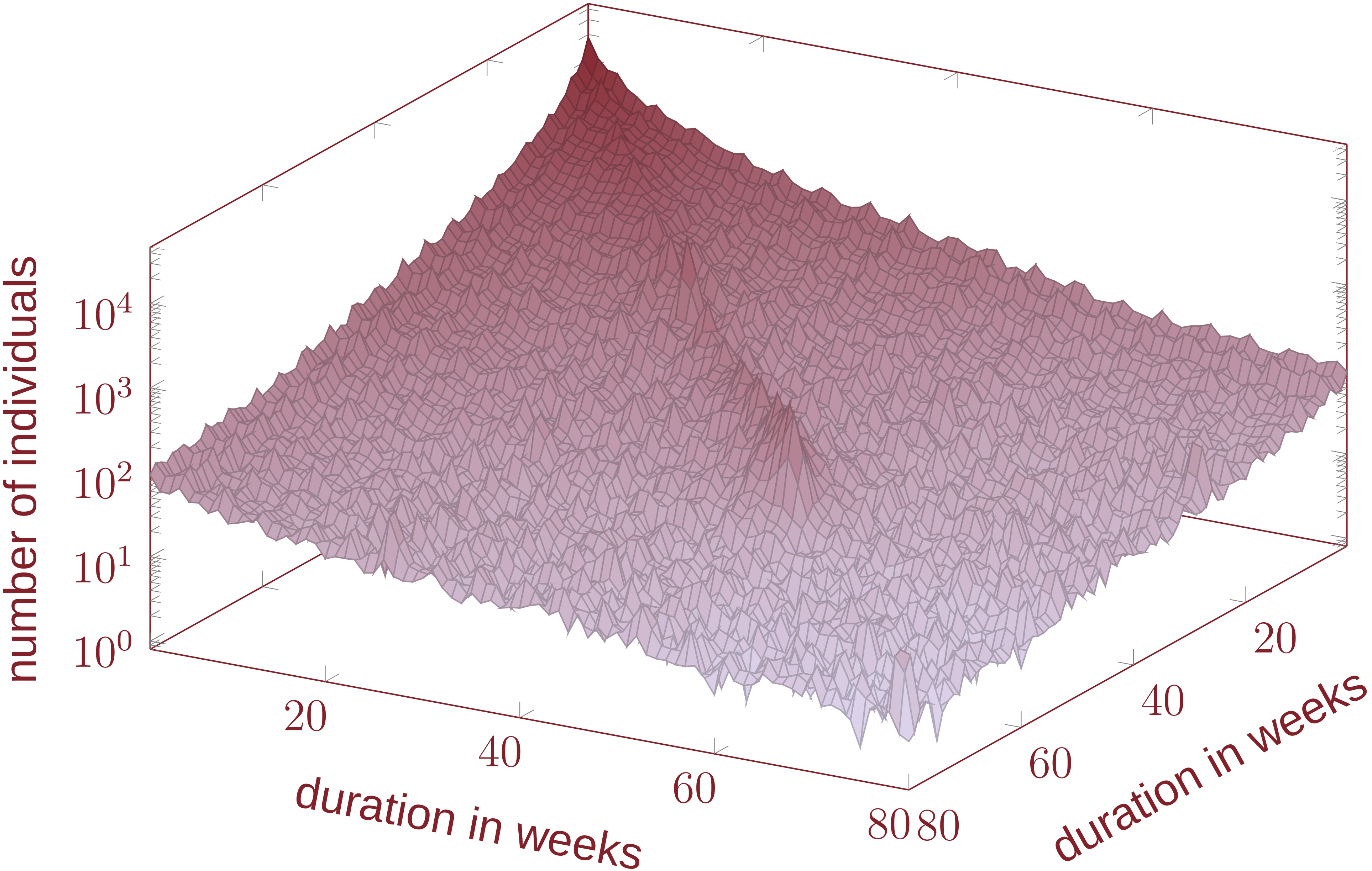
- set $T = [0, 260]$ weeks

individuals age < 45 in 1986 and > 40 in 2007	
w/at least one completed non-employment	1,266,716
w/at least two completed non-employment	852,570
w/ one of the first 2 spells longer than 260 weeks	56,760
final sample	795,810

- mean non-employment spell duration on T: 29.6 weeks

- mean employment duration between spells: 96.4 weeks

Joint Density of Two Spells



Results

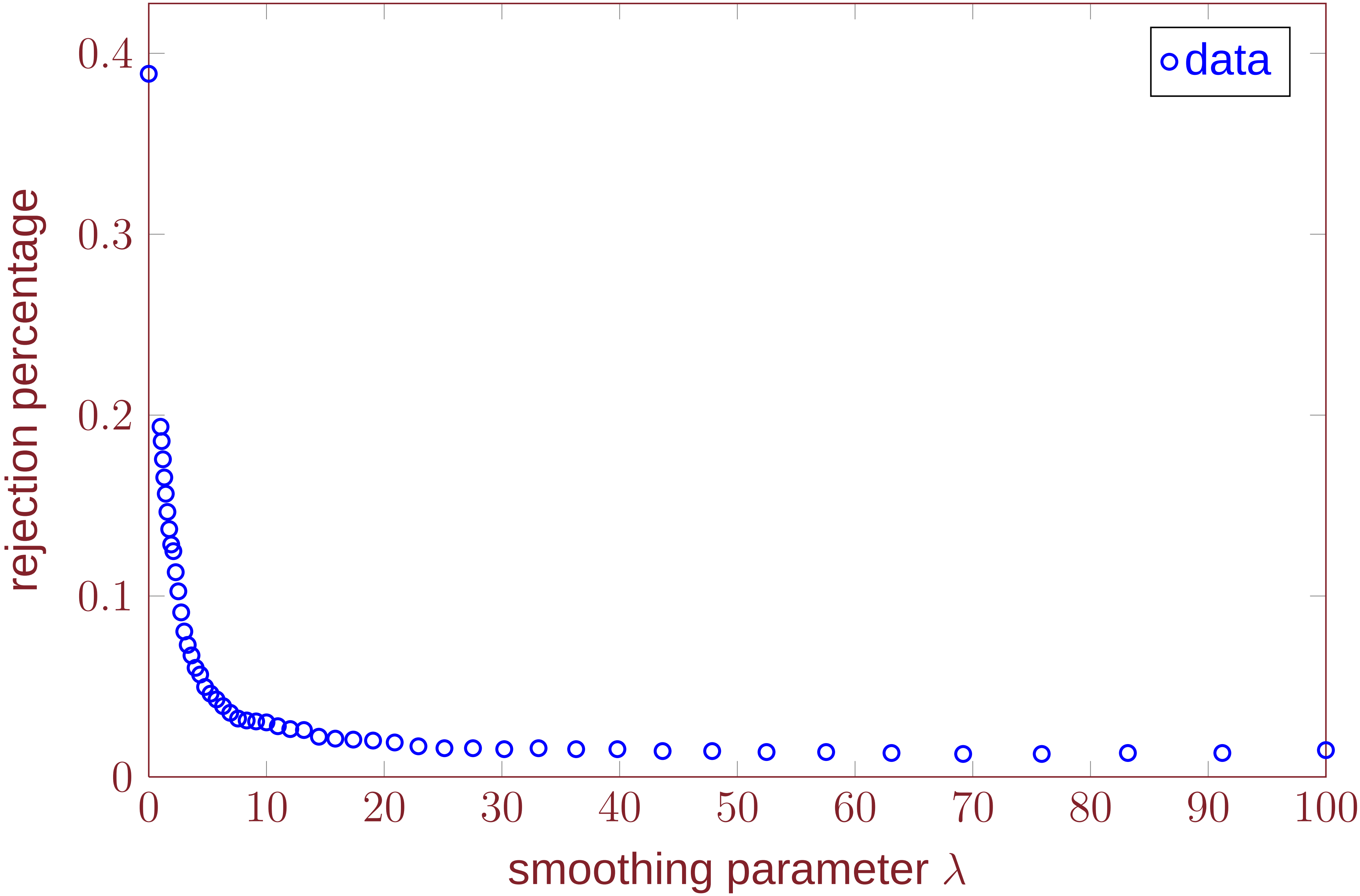
Test

- cannot measure the density at all $(t_1, t_2) \in \mathbb{R}_+$
- propose a discrete version of the test

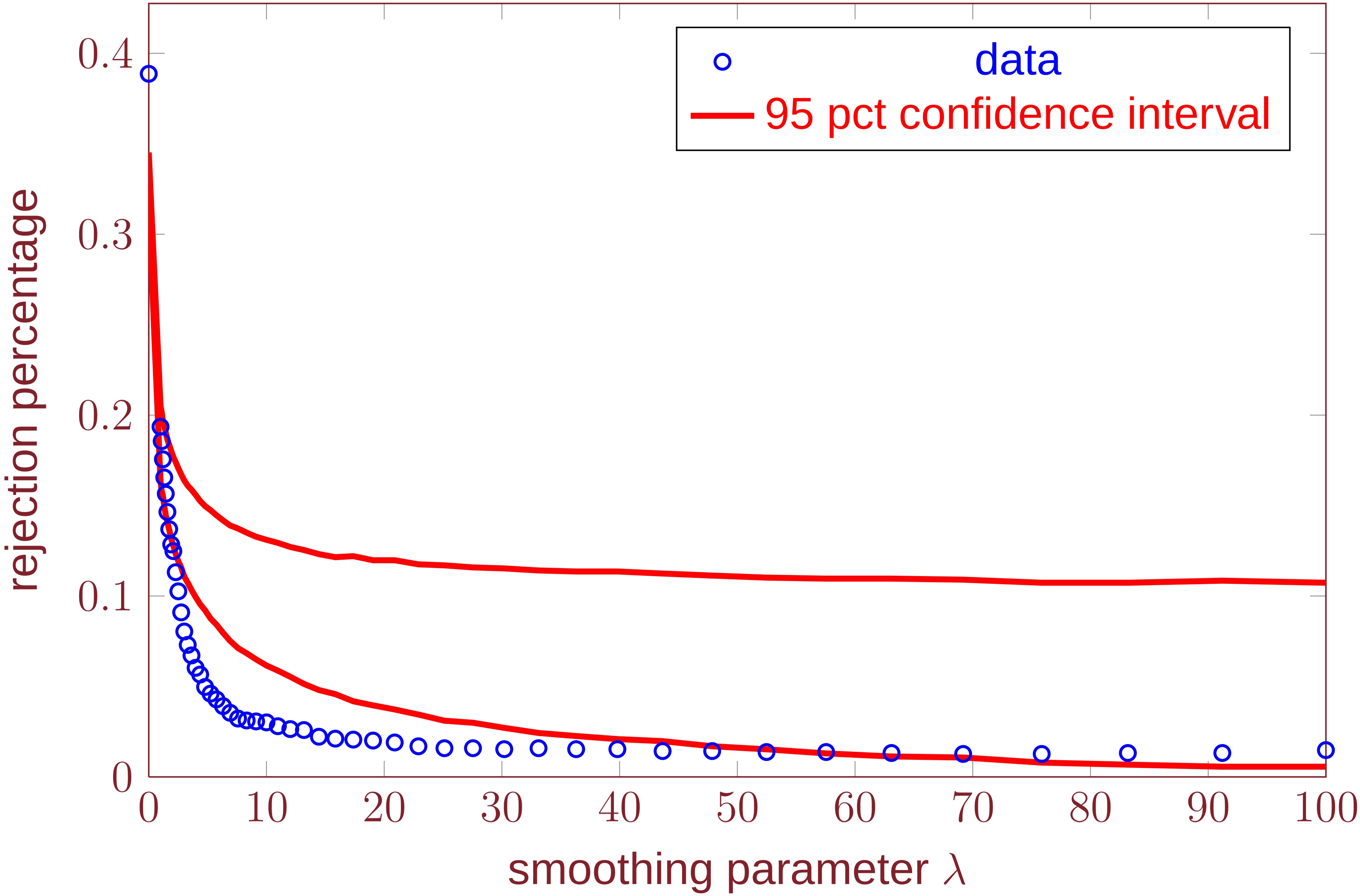
$$\log \phi(t_1 + 1, t_2) - \log \phi(t_1 - 1, t_2) = \frac{b(t_1, t_2)}{t_1^2} - \frac{3}{t_1} - a(t_1, t_2)$$
$$\log \phi(t_1, t_2 + 1) - \log \phi(t_1, t_2 + 1) = \frac{b(t_1, t_2)}{t_2^2} - \frac{3}{t_2} - a(t_1, t_2)$$

- model is symmetric
  - measure $(\phi(t_1, t_2) + \phi(t_2, t_1))/2$
- measured ϕ is noisy
  - “noise” mostly end of month effect
  - smooth with a two-dimensional HP filter, except on diagonal

Test Results



Test Results



Estimation: Part I

- model
  - start with a finite number of types, $k = 1, \ldots, K$
  - select pairs $(\alpha_k, \beta_k)_{k=1}^K$ on pre-specified grid, $\alpha_k \geq 0$
  - shares $g_k \geq 0$, $\sum_{k=1}^N g_k = 1$, or $g \in \Delta^{K-1}$
  - use densities $f(t; \alpha_k, \beta_k)$
  - condition on all types having $t \in T = [0, 260]$
- data
  - smoothed estimate of density $\hat{\phi}(t_1, t_2)$ (2-dimensional HP)
  - evaluate at pairs of times $(t_1, t_2) \in T \equiv \{0, 260\}^2$

Estimation: Part II

- find $\hat{g}$ in Δ^{K-1} which minimizes:

$$\sum_{(t_1,t_2)\in T^2}\left(\sum_{k=1}^{K}\left[\hat{\phi}(t_1,t_2)\bar{F}(\alpha_k,\beta_k)^2 - f(t_1;\alpha_k,\beta_k)f(t_2;\alpha_k,\beta_k)\right] g_k\right)^2$$

where $\bar{F}(\alpha_k,\beta_k) \equiv \sum_{(t'_1,t'_2)\in T^2} f(t'_1;\alpha_k,\beta_k)f(t'_2;\alpha_k,\beta_k)$

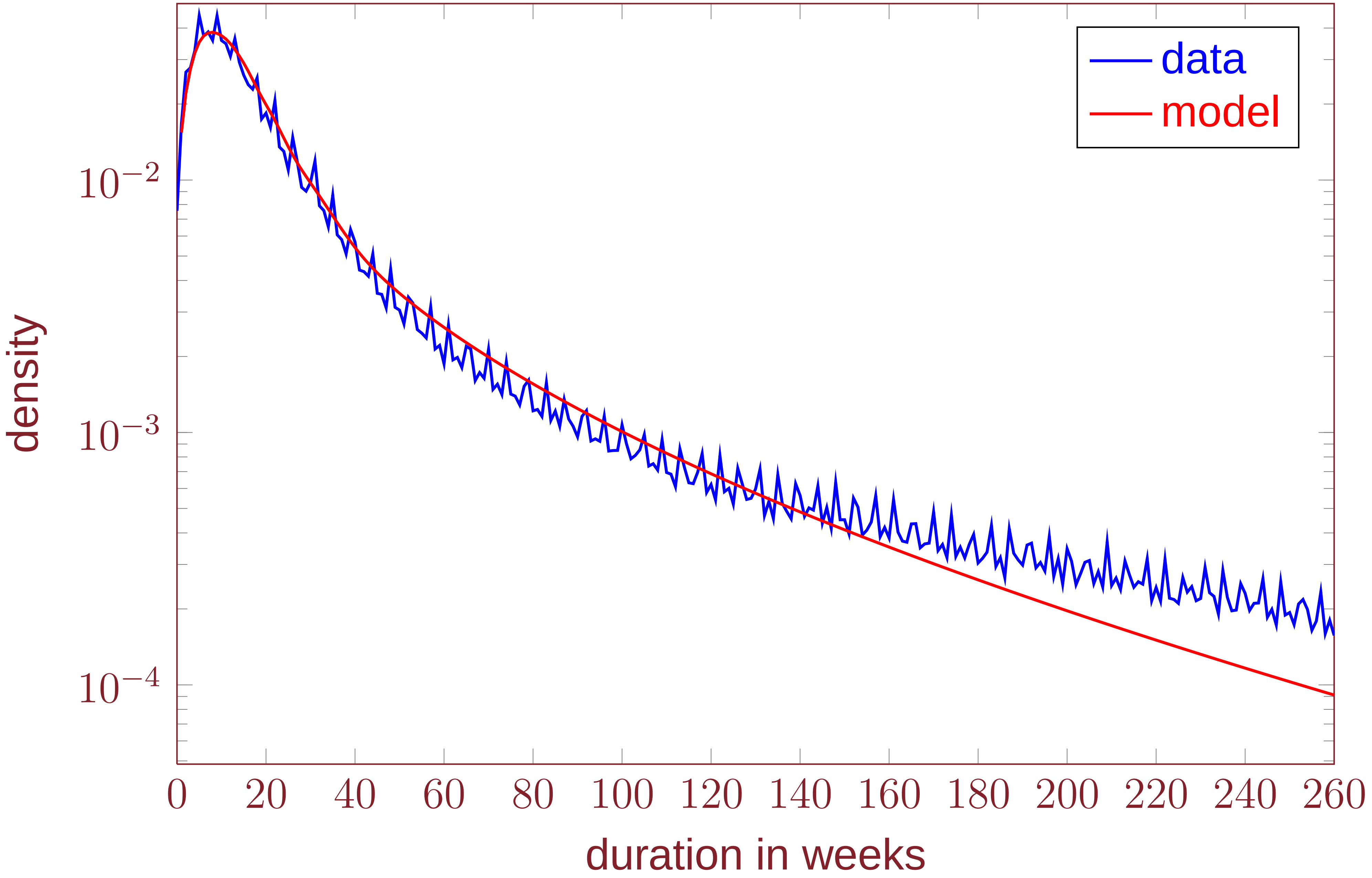
regularize it by adding penalty on $||g||^2$ – ill-posed inverse problem

- drop types with $\hat{g}_k < 10^{-6}$, refine estimates $(\hat{\alpha}_k,\hat{\beta}_k,\hat{g}_k)_{k=1}^{K}$ using EM
    - maximum likelihood with a given number of types ($= K$)
    - EM algorithm uses step 1 as initial guess
    - allows us to find values of (α,β) outside the grid

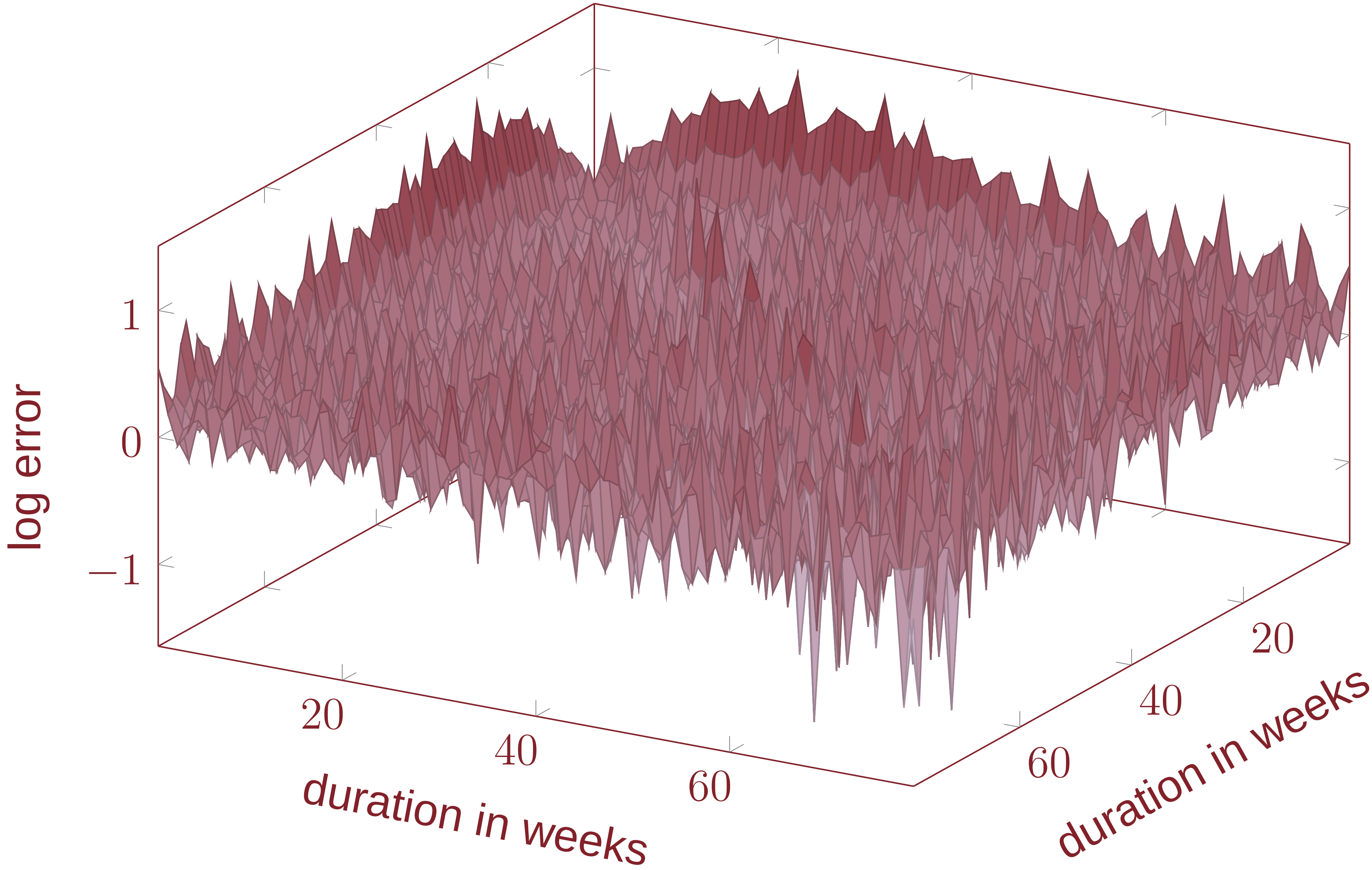
Summary Statistics from Estimation

	minimum distance estimate				EM estimate			
	mean	median	st.dev.	min	mean	median	st.dev	min
α	0.36	0.20	0.51	0.007	391	0.12	2776	0.08
β	7.48	5.03	5.94	1.466	2510	6.01	15623	1.43
$\frac{\mu_n}{\bar{\omega}-\underline{\omega}}$	0.04	0.04	0.03	0.005	0.04	0.04	0.04	0.02
$\frac{\sigma_n}{\bar{\omega}-\underline{\omega}}$	0.21	0.20	0.12	0.005	0.22	0.17	0.14	10^{-5}

Empirical and Theoretical Marginal Distribution



Error in Joint Density Estimates



Sign of α

- consider three G consistent w/ distribution ϕ of two complete spells
- G^+ assumes all $\alpha \geq 0$
  - implies fraction of completed spells $c = 0.98$

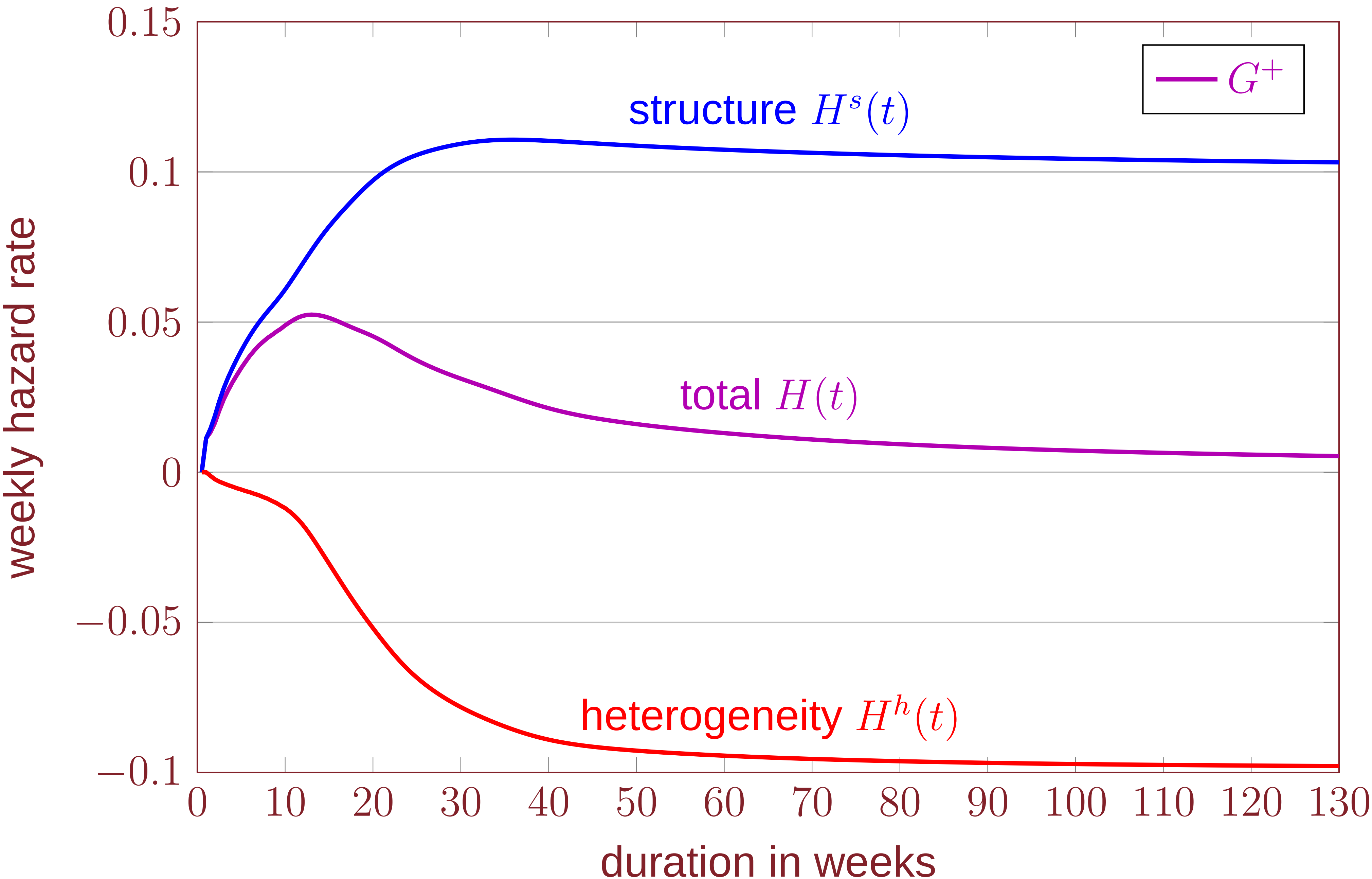
Sign of α

- consider three G consistent w/ distribution ϕ of two complete spells
- G^+ assumes all $\alpha \geq 0$
  - implies fraction of completed spells $c = 0.98$
- construct $\underline{G}$, $\bar{G}$ consistent with $\hat{c} = 0.80$
  - $\underline{G}$ with minimum number of $\alpha < 0$, about 18% of workers
  - $\bar{G}$ with maximum number of $\alpha < 0$, about 51% of workers

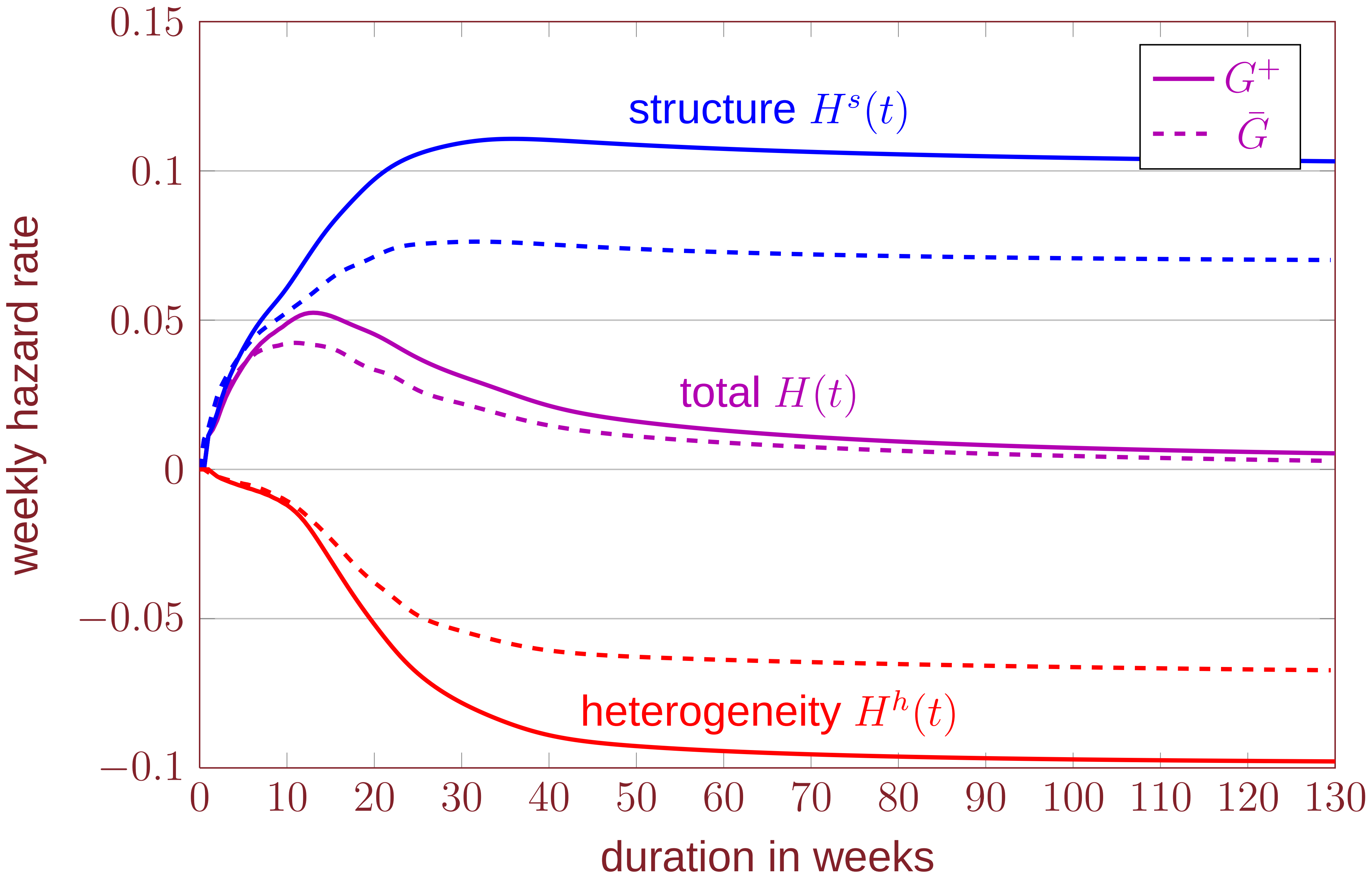
Sign of α

- consider three G consistent w/ distribution ϕ of two complete spells
- G^+ assumes all $\alpha \geq 0$
  - implies fraction of completed spells $c = 0.98$
- construct $\underline{G}$, $\bar{G}$ consistent with $\hat{c} = 0.80$
  - $\underline{G}$ with minimum number of $\alpha < 0$, about 18% of workers
  - $\bar{G}$ with maximum number of $\alpha < 0$, about 51% of workers
- hazard rate decomposition for all three cases

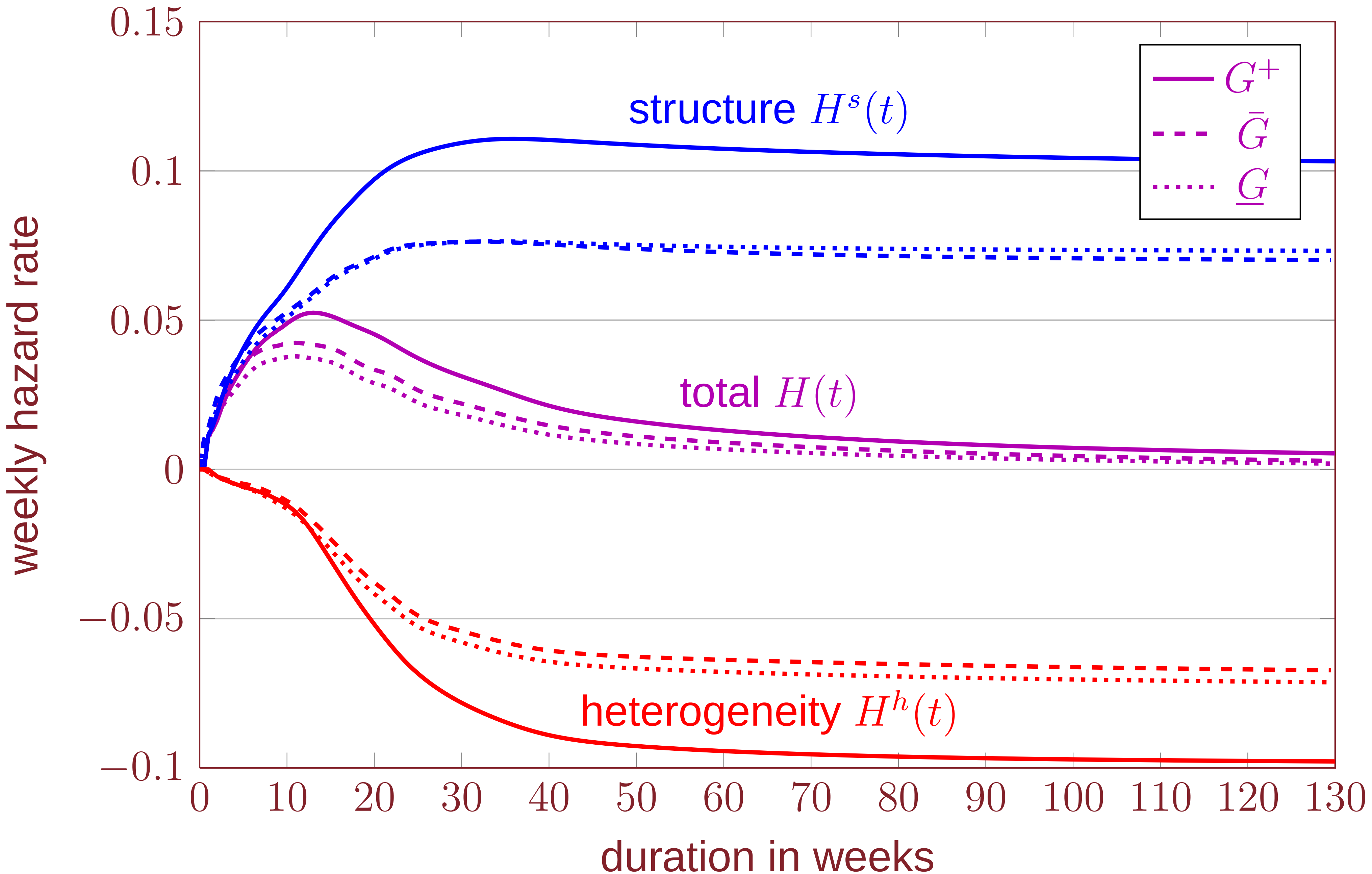
Hazard Rate Decomposition



Hazard Rate Decomposition



Hazard Rate Decomposition



Fixed Cost

Size of Fixed Costs

- use model and estimated G to infer the distribution of $(\psi_n + \psi_e)/b_0$
  - fixed costs (for each type) unidentified without more information
  - common values of other parameters
    - discount rate $r = 0.03$, constant benefit of non-employment
    - cost of switching from employ. to non-employ.: $\psi_n = 0$
    - drift of wages $\mu_e = 0.015$ and $\mu_n = -0.03$
    - volatility of wages when employed $\sigma_e = 0.05$
  - volatility σ_n and width $\bar{\omega} - \underline{\omega}$ from estimates

- mean (median) fixed costs $\frac{\psi_e}{b_0} = 1.8\,(0.5)$ percent per year

- mean (median) width of inaction $\bar{\omega} - \underline{\omega} = 1.8\,(1.7)$ percent

Should We Set Fixed Cost to Zero? No.

- we estimate very small fixed cost

- what happens if we set them to zero?
  - $\bar{\omega} = \underline{\omega}$ and hence $\beta = 0$ for all workers
  - this restriction is emphatically rejected (much smaller likelihood)
  - implied durations are much too short
  - if $\alpha = \beta = 0$, mean duration conditional on $t \in [\underline{t}, \bar{t}]$ is $(\underline{t}\,\bar{t})^{1/2}$
    - other values of α give still shorter durations
  - if $T = [1, 260]$, upper bound ≈ 16, while mean in data ≈ 30

Conclusions

Conclusions

- we model nonemployment duration as a stopping time
- we allow for arbitrary parameter heterogeneity
- we prove the model is partially identified and testable
  - test and don't reject the model on Austrian data
  - estimate the distribution of unobserved types in Austrian data
- we decompose evolution of hazard rate and residual duration
  - hazard rate suggests heterogeneity is the dominant force
- we find that small fixed switching cost important for model fit

Decomposing Duration Dependence in a Stopping Time Model

Fernando Alvarez Katarína Borovičková Robert Shimer

June 13, 2015

Mixed Proportional Hazard Model

Nonparametric Test

- individual hazard rate: $\theta h(t)$
- survivor function: fraction of people with duration at least t_1 and t_2

$$\Phi(t_1, t_2) = \int e^{-\theta(Z(t_1)+Z(t_2))} dG(\theta)$$

- $Z(t)$ is the *integrated* baseline hazard

$$Z(t) \equiv \int_0^t h(\tau) d\tau$$

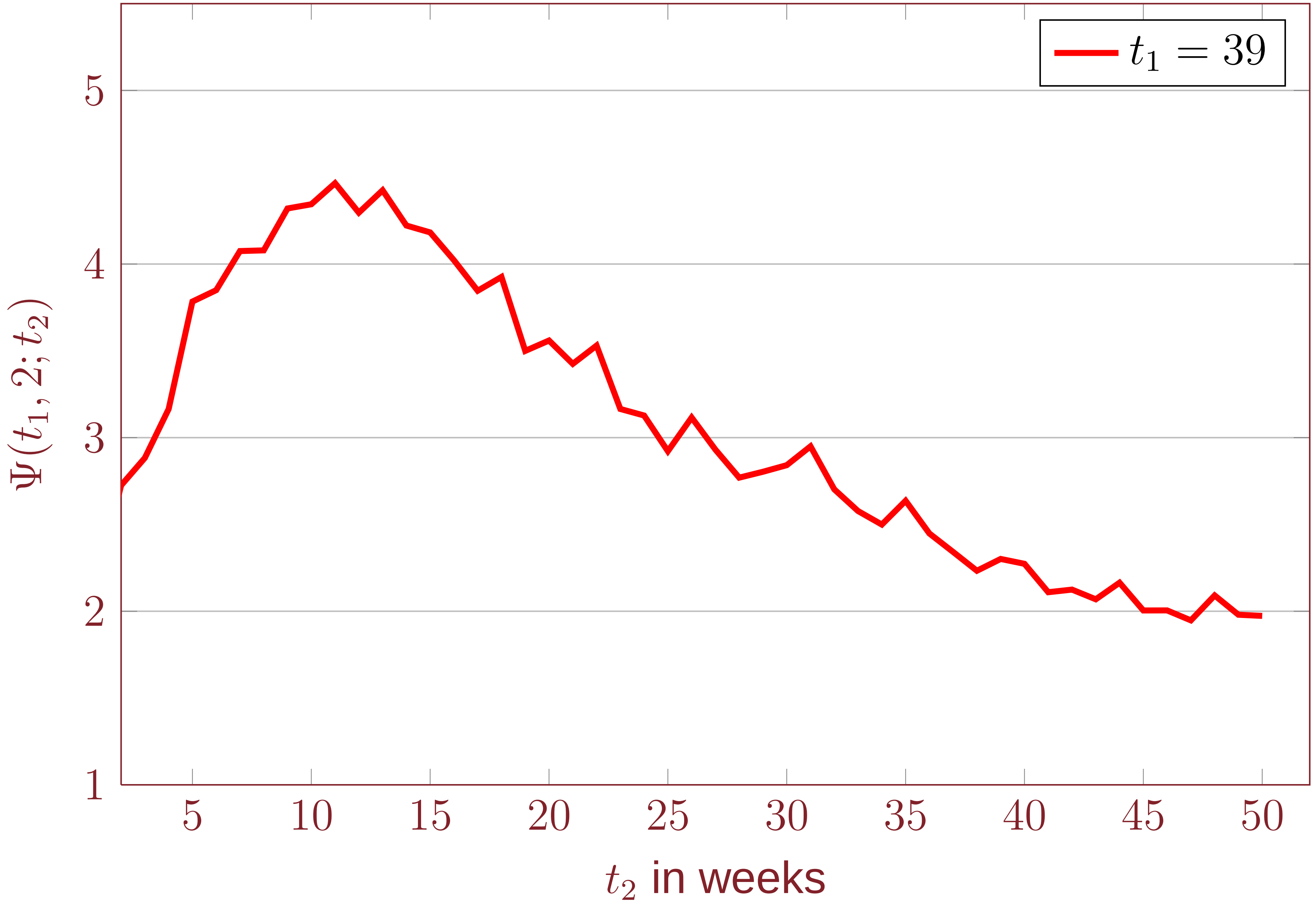
- test

$$\Psi(t_1, t_1'; t_2) \equiv \frac{\left(\frac{\Phi(t_1,t_2)-\Phi(t_1+1,t_2)}{\Phi(t_1,t_2)-\Phi(t_1,t_2+1)}\right)}{\left(\frac{\Phi(t_1',t_2)-\Phi(t_1'+1,t_2)}{\Phi(t_1',t_2)-\Phi(t_1',t_2+1)}\right)} = \frac{h(t_1)}{h(t_1')}$$

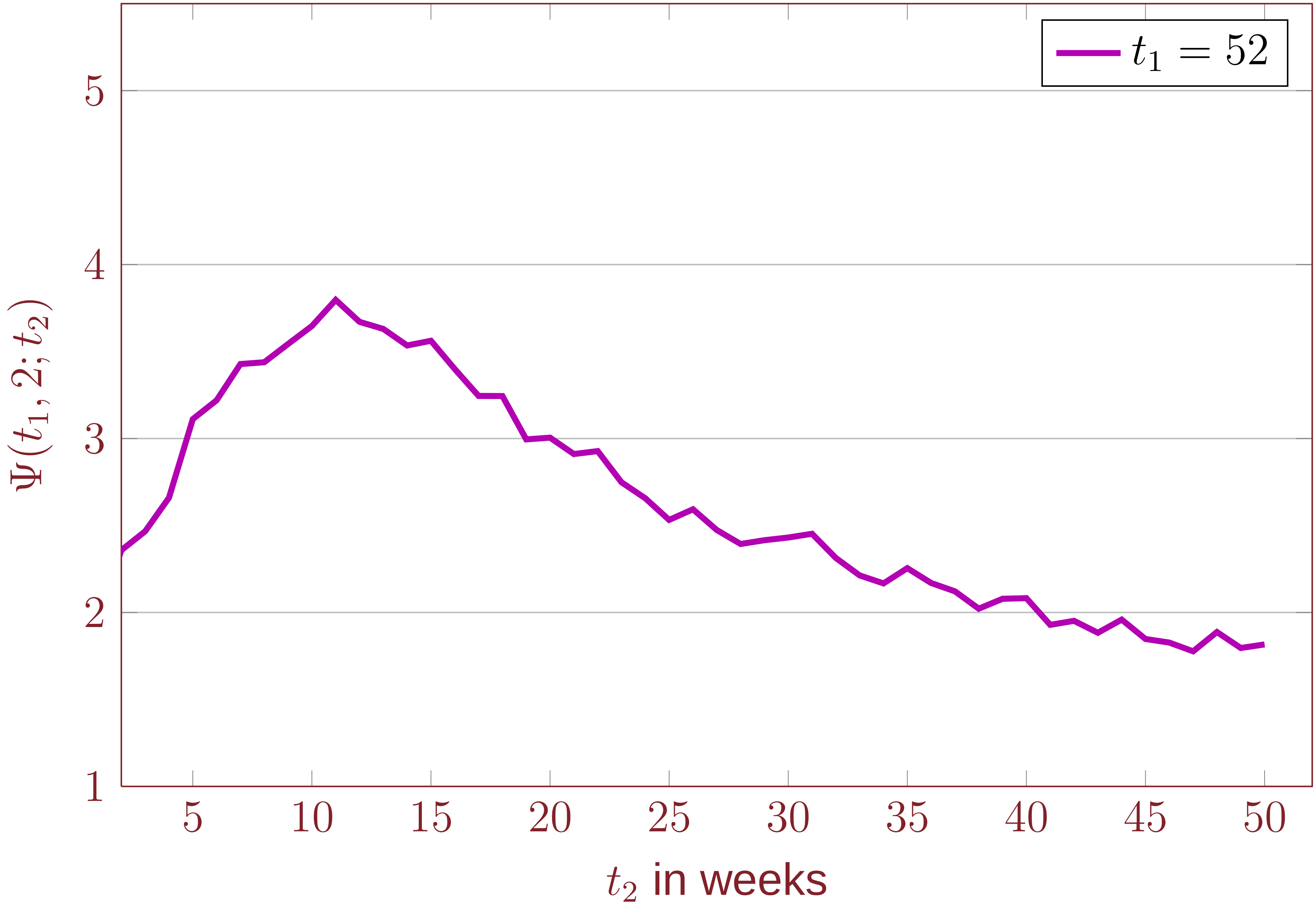
Non-employment Exit Rate

Non-employment Exit Rate

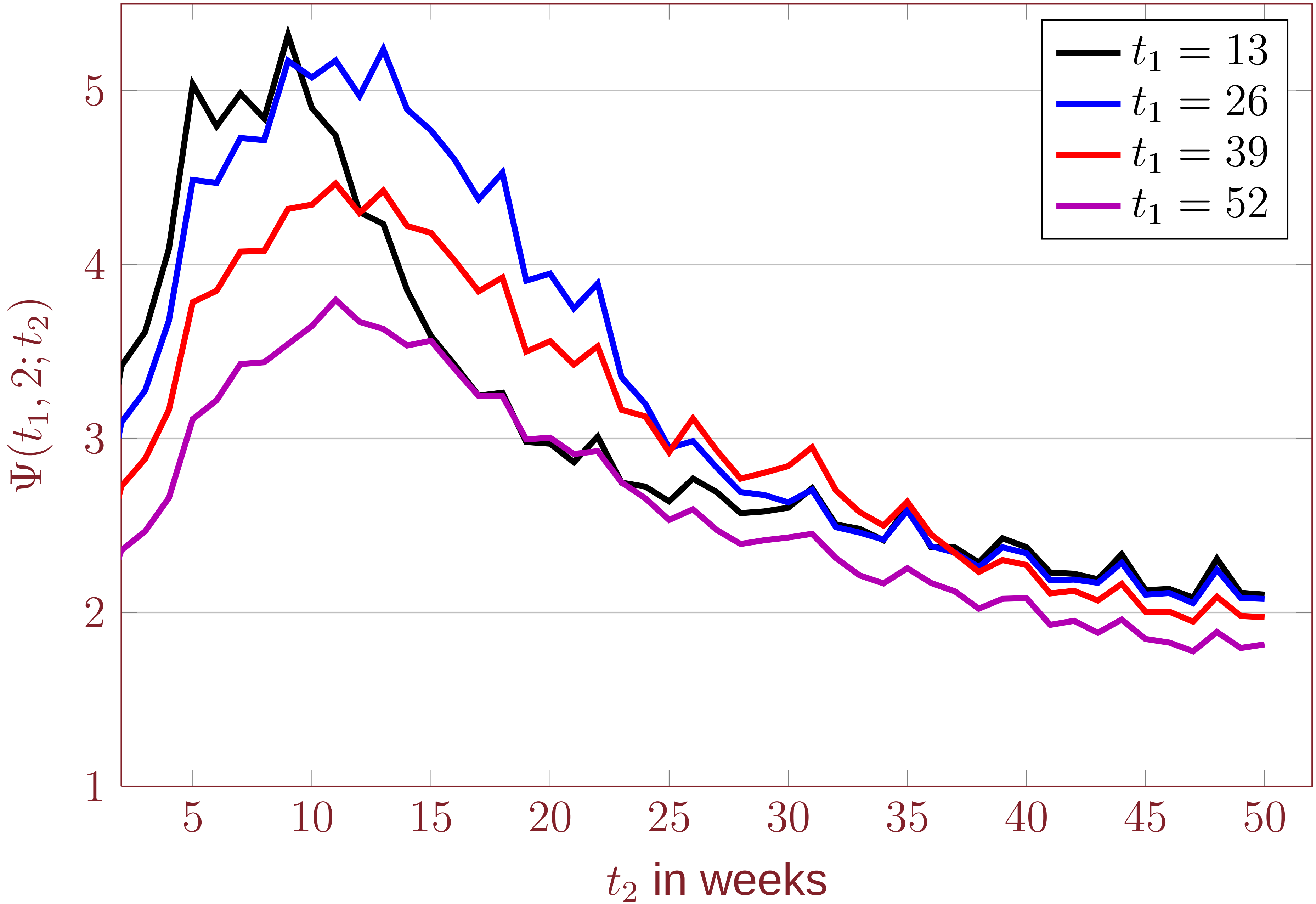
Non-employment Exit Rate



Non-employment Exit Rate



Non-employment Exit Rate

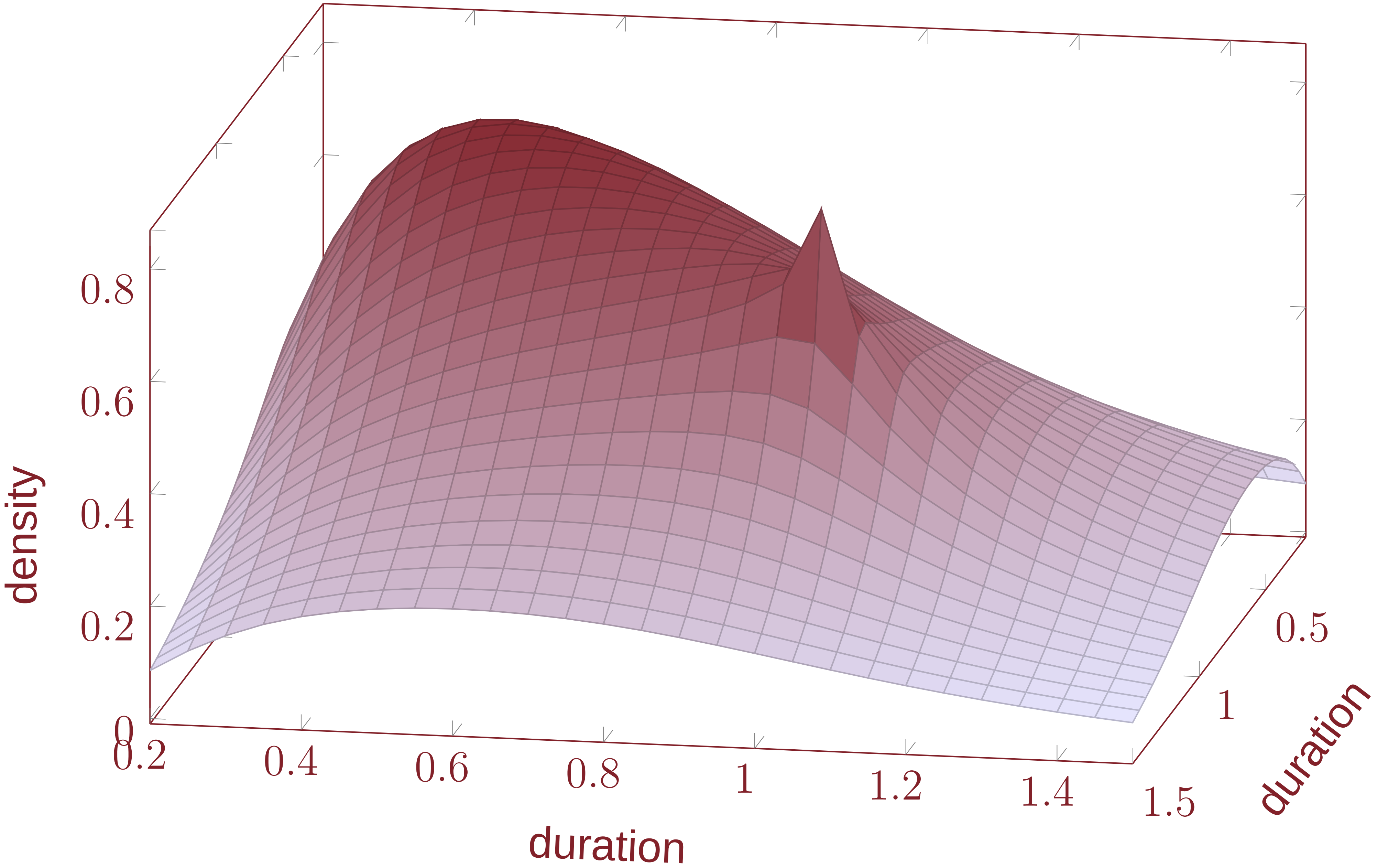


Details on Identification Proof

ϕ is Smooth Almost Everywhere

- our identification approach differentiates ϕ to construct moments of G
- for any G, ϕ is infinitely differentiable at any $t_1 \neq t_2$
  - proof applies Leibniz's formula
- easy to construct examples where ϕ is not differentiable at some (t, t)
  - $\alpha/\beta = \mu$ and β is Pareto distributed
  - mixture of these distributions
- consistent with the "ridge" in the empirical distribution of ϕ

Kinked ϕ



density
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.5
duration
0.5
1
duration

First Moments of $\hat{G}(\alpha^2, \beta^2 | t_1, t_2)$

- differentiate ϕ: $\frac{2t_i^2 \phi_i(t_1,t_2)}{\phi(t_1,t_2)} = \mathbb{E}(\beta^2|t_1,t_2) - 3t_i - \mathbb{E}(\alpha^2|t_1,t_2)t_i^2$, where

$$\mathbb{E}(\alpha^2|t_1,t_2) = \frac{\iint \alpha^2 f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}{\iint f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}$$

and $$\mathbb{E}(\beta^2|t_1,t_2) = \frac{\iint \beta^2 f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}{\iint f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)}$$

- if $t_1 \neq t_2$, this gives two invertible linear equations in two unknowns

Second Moments of $\hat{G}(\alpha^2, \beta^2 | t_1, t_2)$

- differentiate ϕ a second time:

$$\frac{4t_i^4\phi_{ii}(t_1,t_2)}{\phi(t_1,t_2)} = \mathbb{E}(\alpha^4|t_1,t_2)t_i^4 + \mathbb{E}(\beta^4|t_1,t_2) - 2\mathbb{E}(\alpha^2\beta^2|t_1,t_2)t_i^2 + 6\mathbb{E}(\alpha^2|t_1,t_2)t_i^3 - 10\mathbb{E}(\beta^2|t_1,t_2)t_i + 15t_i^2, \; i = 1,2,$$

$$\frac{4t_1^2t_2^2\phi_{12}(t_1,t_2)}{\phi(t_1,t_2)} = \mathbb{E}(\alpha^4|t_1,t_2)t_1^2t_2^2 + \mathbb{E}(\beta^4|t_1,t_2) - \mathbb{E}(\alpha^2\beta^2|t_1,t_2)(t_1^2+t_2^2) + 3\mathbb{E}(\alpha^2|t_1,t_2)t_1t_2(t_1+t_2) - 3\mathbb{E}(\beta^2|t_1,t_2)(t_1+t_2) + 9t_1t_2.$$

- if $t_1 \neq t_2$, this gives three invertible linear equations in three unknowns

m^{th} **moments of** $\hat{G}(\alpha^2, \beta^2 | t_1, t_2)$

$$\Phi_m(t_1,t_2) = L_m(t_1,t_2) \cdot U_m(t_1,t_2) \cdot M_m(t_1,t_2) + v_m(t_1,t_2)$$

- $(m+1) \times 1$ data vector: $\Phi_m(t_1,t_2) = \left[\frac{2^m t_1^{2(m-i)} t_2^{2i}}{\phi(t_1,t_2)} \frac{\partial^m \phi(t_1,t_2)}{\partial t_1^{m-i} \partial t_2^i} \right]_{0 \le i \le m}$
- $(m+1) \times (m+1)$ lower-triangular matrix, nonsingular if $t_1 \neq t_2$:

$$L_m(t_1,t_2) = \left[\frac{(-1)^i (m-j)!}{(m-i)!(i-j)!} t_2^{2(i-j)} (t_1^2 - t_2^2)^{j/2} \right]_{0 \le j \le i \le m}$$

- $(m+1) \times (m+1)$ upper-triangular matrix, nonsingular if $t_1 \neq t_2$:

$$U_m(t_1,t_2) = \left[\frac{j!}{i!(j-i)!} (t_1^2 - t_2^2)^{i/2} \right]_{0 \le i \le j \le m}$$

- $(m+1) \times 1$ moment vector $M_m(t_1,t_2) = \left[\mathbb{E}(\alpha^{2(m-i)} \beta^{2i} | t_1, t_2) \right]_{0 \le i \le m}$
- $(m+1) \times 1$ known lower moment vector $v_m(t_1,t_2)$

From Conditional Moments to Distributions

- all moments of distribution of $\hat{G}(\alpha^2, \beta^2 | t_1, t_2)$ are identified

- under regularity conditions, this identifies $\hat{G}(\alpha^2, \beta^2 | t_1, t_2)$
  - Carleman's sufficient condition for a one dimensional problem
  - apply Cramér-Wold theorem to extend it to two dimensions
  - the conditions hold in our environment

From Conditional to Unconditional Distributions

- using Bayes rule,

$$d\hat{G}(\alpha^2, \beta^2 | t_1, t_2) = \frac{f(t_1; \alpha, \beta)\, f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\int_0^\infty \int_0^\infty f(t_1; \alpha', \beta')\, f(t_2; \alpha', \beta')\, dG(\alpha', \beta')}.$$

- take ratios and invert to get

$$\frac{dG(\alpha, \beta)}{dG(\alpha', \beta')} = \frac{d\hat{G}(\alpha^2, \beta^2 | t_1, t_2)}{d\hat{G}(\alpha'^2, \beta'^2 | t_1, t_2)} \frac{f(t_1; \alpha', \beta') f(t_2; \alpha', \beta')}{f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)}$$

- theorem: ϕ near any $t_1 \neq t_2$ identifies G

Pareto Example Revisited

- Pareto distribution has infinite higher moments
- yet we prove the conditional moments is always finite
- can't we integrate conditional moments to get unconditional ones?

$$\mathbb{E}(\beta^{2m}) = \iint_{T^2} \mathbb{E}(\beta^{2m}|t_1, t_2)\phi(t_1, t_2)\, dt_1\, dt_2$$

  - $\mathbb{E}(\beta^{2m}|t_1, t_2) < \infty$ at all $t_1 \neq t_2$
  - integral does not converge exactly when the moment is infinite

Multidimensional HP Filter

Nonparametric Filter

- data: $\psi(t_1, t_2)$ for $t_1 \in \{1, \ldots, t_2\}$ and $t_2 \in \{1, \ldots, T\}$
  - smooth ψ on each side of the diagonal
- trend: $\bar{\psi}(t_1, t_2)$

$$\min_{\{\bar{\psi}(t_1,t_2)\}} \Bigg(\sum_{t_2=1}^{T} \sum_{t_1=1}^{t_2} (\psi(t_1, t_2) - \bar{\psi}(t_1, t_2))^2 + \lambda \sum_{t_2=3}^{T} \sum_{t_1=2}^{t_2-1} (\bar{\psi}(t_1+1, t_2) - 2\bar{\psi}(t_1, t_2) + \bar{\psi}(t_1-1, t_2))^2 + \lambda \sum_{t_2=2}^{T-1} \sum_{t_1=1}^{t_2-1} (\bar{\psi}(t_1, t_2+1) - 2\bar{\psi}(t_1, t_2) + \bar{\psi}(t_1, t_2-1))^2 \Bigg)$$