Firm Size Distribution

741 Macroeconomics
Topic 1

Masao Fukui

Fall 2024

Course Logistics

- Lecture:
  - MonWed, 8:30-9:45am in SSW 315
- Instructor:
  - Masao Fukui (mfukui@bu.edu)
  - Office hours: MonTue 4:15-5:45pm in Room 400 (my office)
- Grades:
  - 80%: problem sets
  - 20%: research proposal or a final project
  - Bonus points if you catch coding errors in my code

Course Theme

- In macro, we often postulate a representative firm solving:

$$\max_L f_t(L) - wL$$

- This gives the (inverse) aggregate labor demand function

$$f_t'(L) = w$$

- Together with aggregate labor supply, it pins down wages and employment.

Course Theme

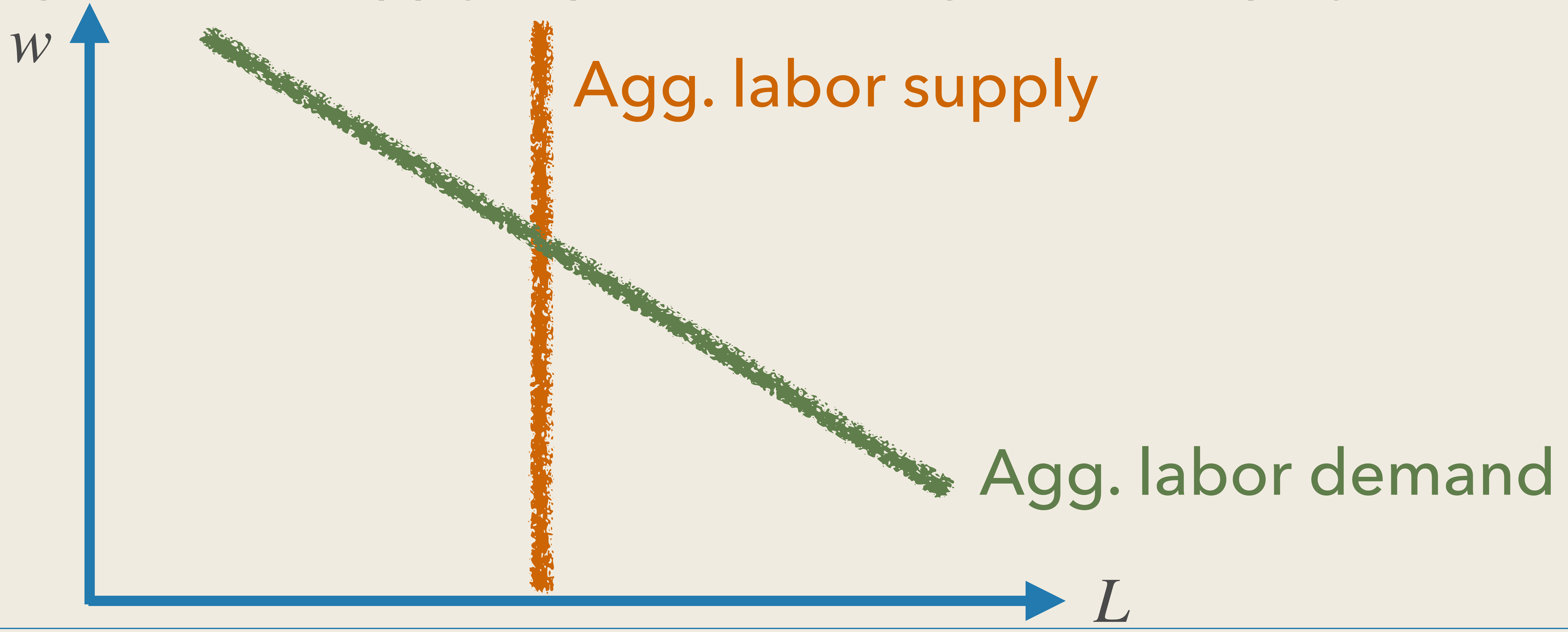
- In macro, we often postulate a representative firm solving:

$$\max_{L} f_t(L) - wL$$

- This gives the (inverse) aggregate labor demand function

$$f_t'(L) = w$$

- Together with aggregate labor supply, it pins down wages and employment.



Unpacking Aggregate Labor Demand

What is aggregate labor demand? – Two themes we highlight

1. There is no "representative firm"
    - The reality, of course, consists of heterogeneous firms
    - How does the heterogeneity shape the aggregate labor demand?

   First theme: heterogeneous firms

2. The labor market is not competitive
    - We assumed firms could hire any L taking w as given
    - It is hard to imagine there is any real firm that thinks in such a way

   Second theme: monopsony and frictional labor market

The Course is Not About

- The course is not about aggregate labor supply
  - We will mostly assume that the labor supply is fixed
  - There is a literature focusing on labor supply (see Rogerson (2024) for a survey)
- The course is not about investment/capital demand or innovation
  - We will mostly abstract from capital
  - Another big literature on heterogeneous firms focuses on investment/R&D

Technical Tools

Along the way, I put emphasis on two technical tools:

1. **Continuous-time techniques**
    - Increasingly becoming popular in macro
    - Superficially looks elegant & sometimes actually useful
    - At best, you will be able to use it after this course
    - At worst, you won't be scared of reading continuous-time papers

Technical Tools

Along the way, I put emphasis on two technical tools:

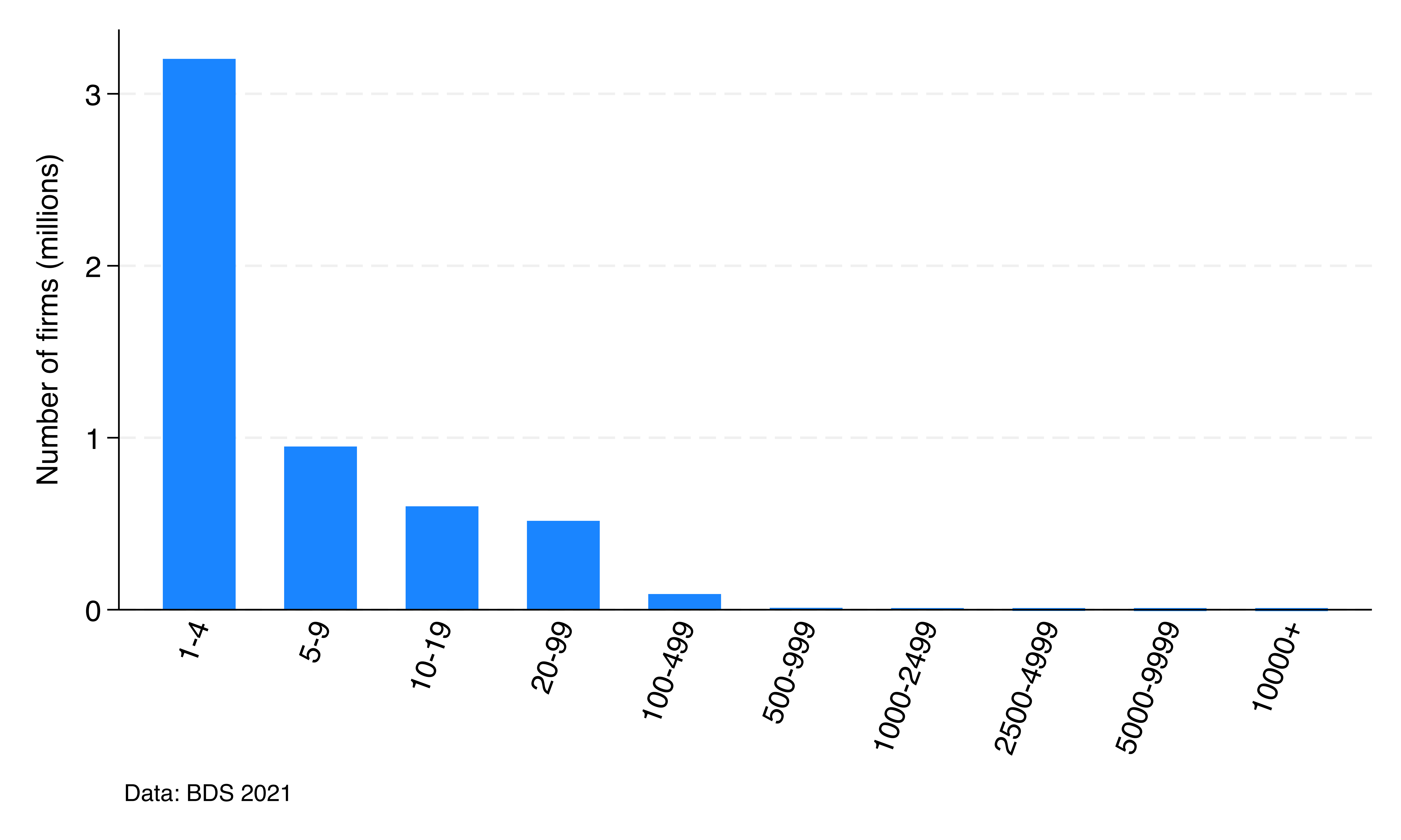
1. **Continuous-time techniques**
   - Increasingly becoming popular in macro
   - Superficially looks elegant & sometimes actually useful
   - At best, you will be able to use it after this course
   - At worst, you won't be scared of reading continuous-time papers
2. **Computational methods**
   - Extremely important in macro nowadays
   - Hard to write qualitative papers now, quantification is almost always necessary
   - The frontier expanded a lot in the past 5 years
   - Young generation's comparative advantage

Computation Tips

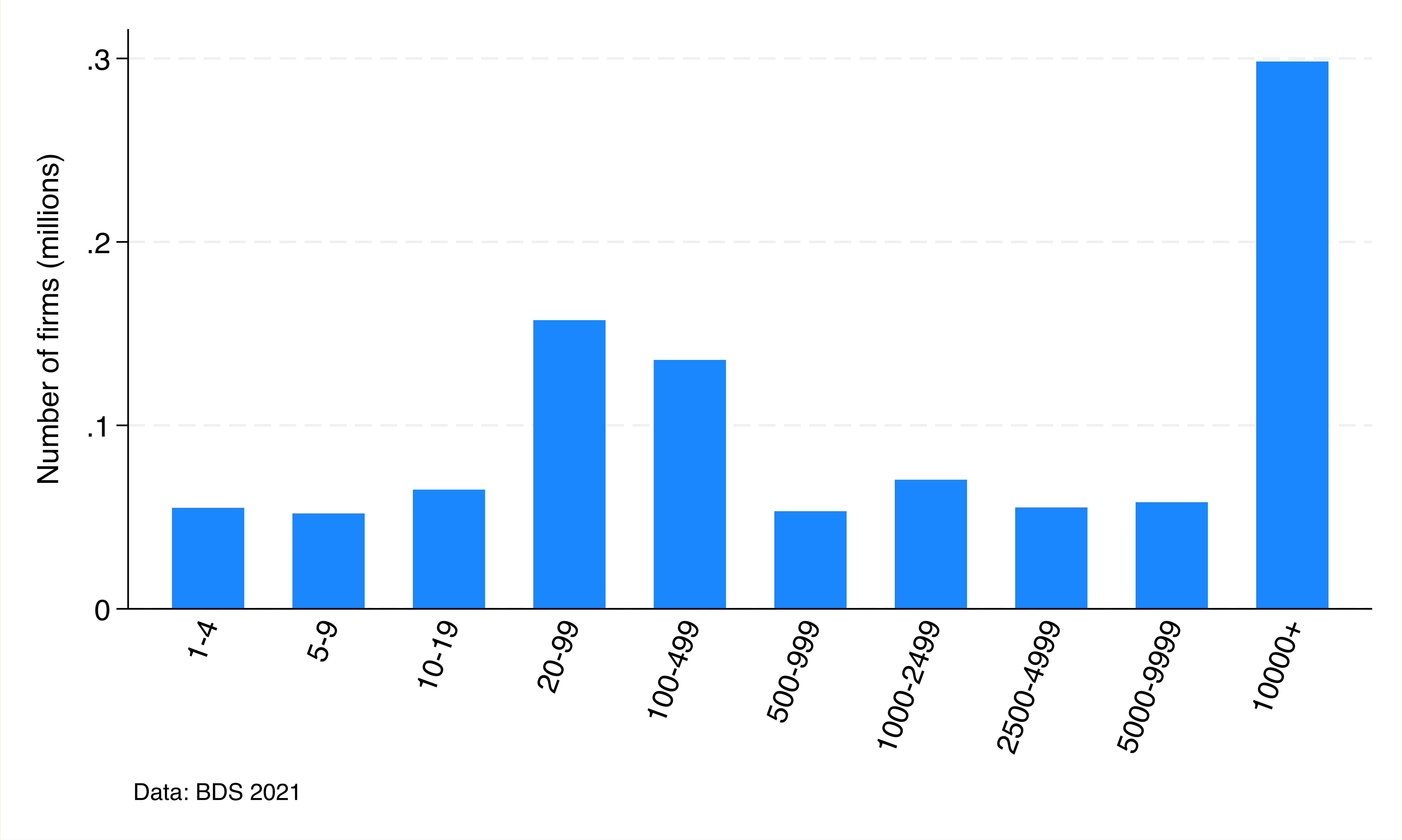
- I strongly recommend Julia as a computational language for quantitative macro
  - Very similar to Matlab in terms of syntax, but much faster
    - Matlab is a dying language in my view
  - Python is good for many purposes, but not for quantitative macro
    - needs a lot of work (JAX) to speed up & struggles to handle sparse matrices
  - Slightly slower than Fortran and C++, but much easier to code/debug
    - Remember: total time cost = time running + time coding/debugging
- I recommend VS Code + Github Copilot as an editor
  - Github copilot is a game changer for me (free for academia)
- I post all the codes at: https://github.com/masaofukui/741_Julia

Firm Size Distribution in the US 2021

Firm Size (Employment) Distribution



Employment Share of Each Size Category

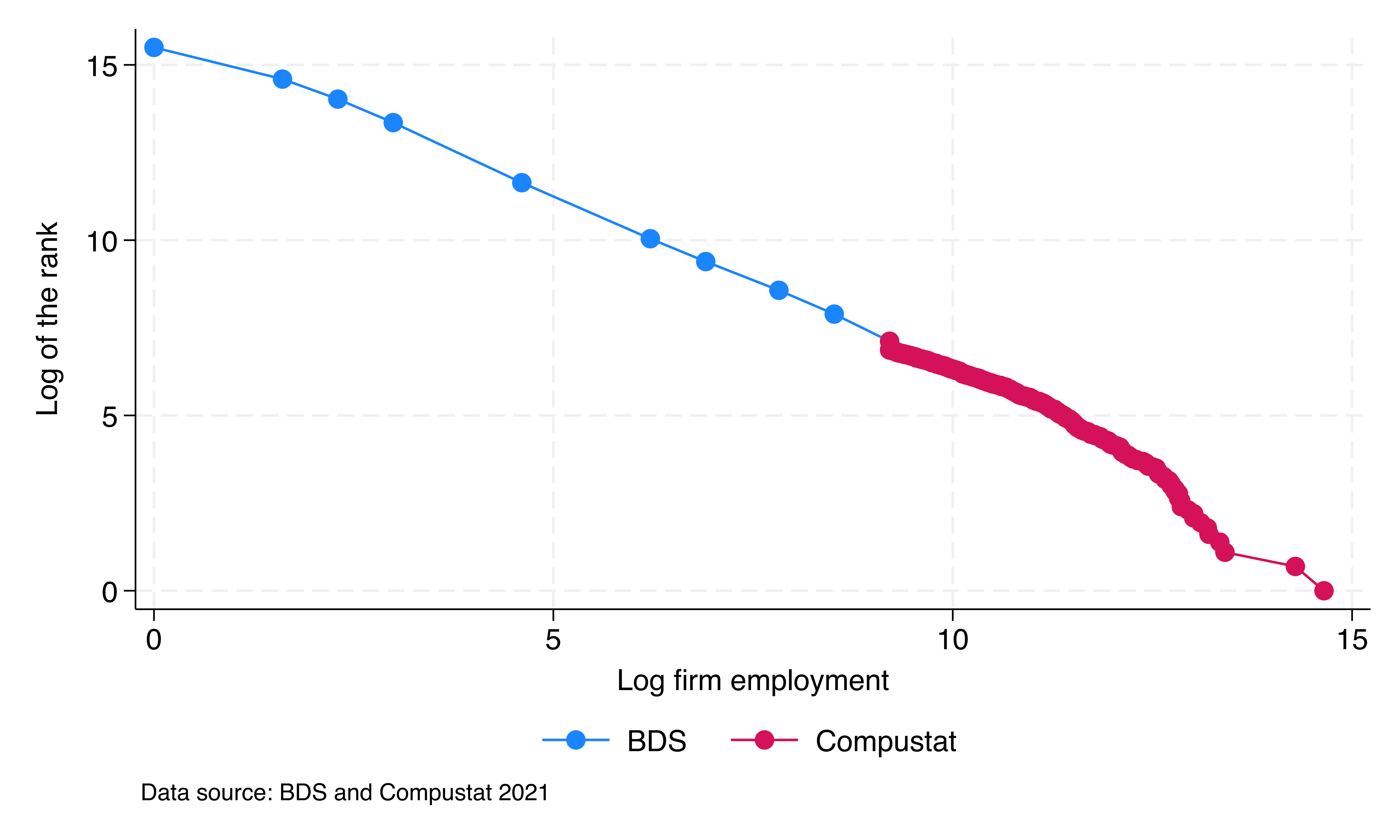


A Handful of Firms Hire Majority of Workers

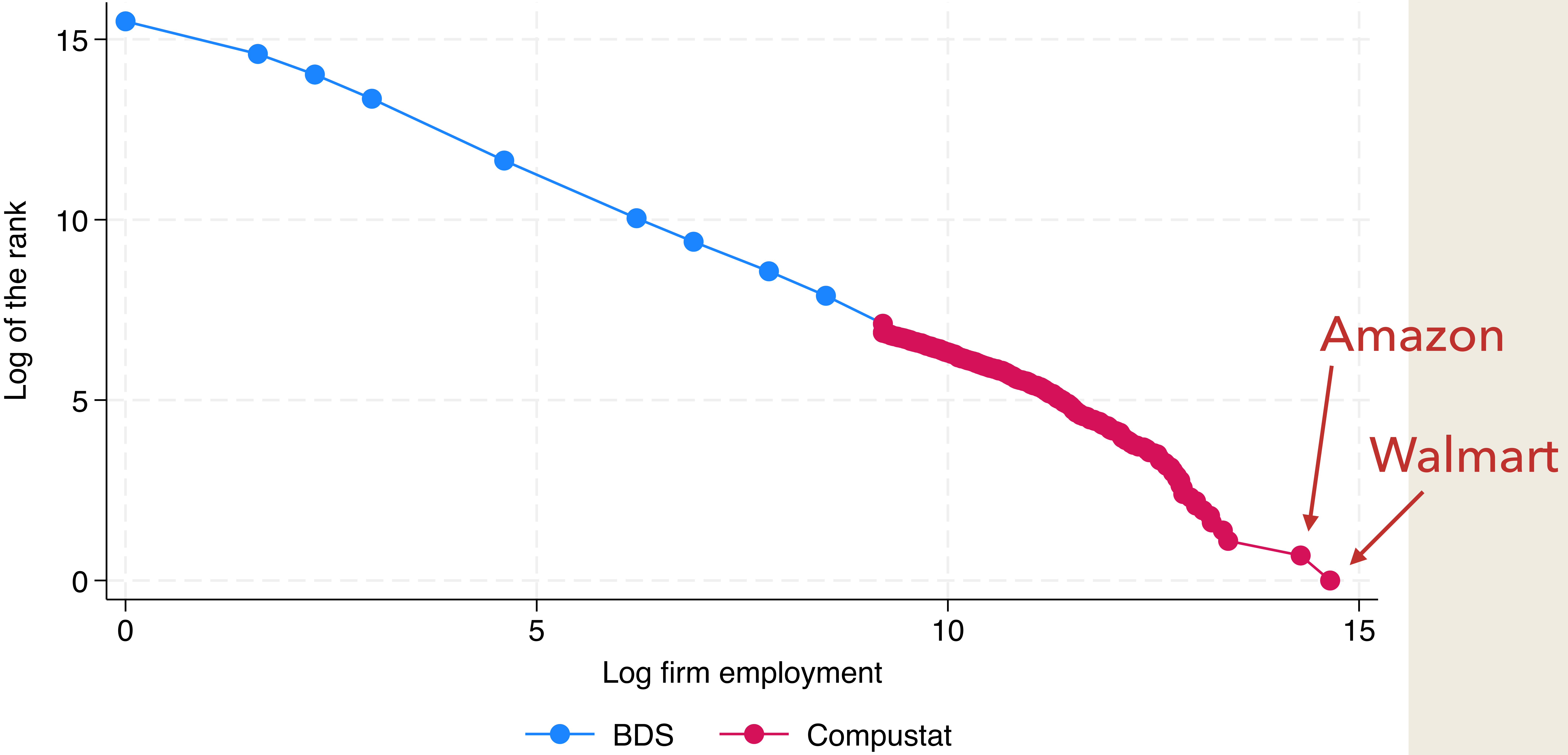
- Large firms in the US are extremely large
  - Top 0.02% of firms ($\approx$ 1,200 firms) account for 30% of employment in the US
  - Top 1% of firms ($\approx$ 60,000 firms) account for 60% of employment in the US

- What does the right tail of the firm size distribution look like?

Power Law in Firm Size Distribution

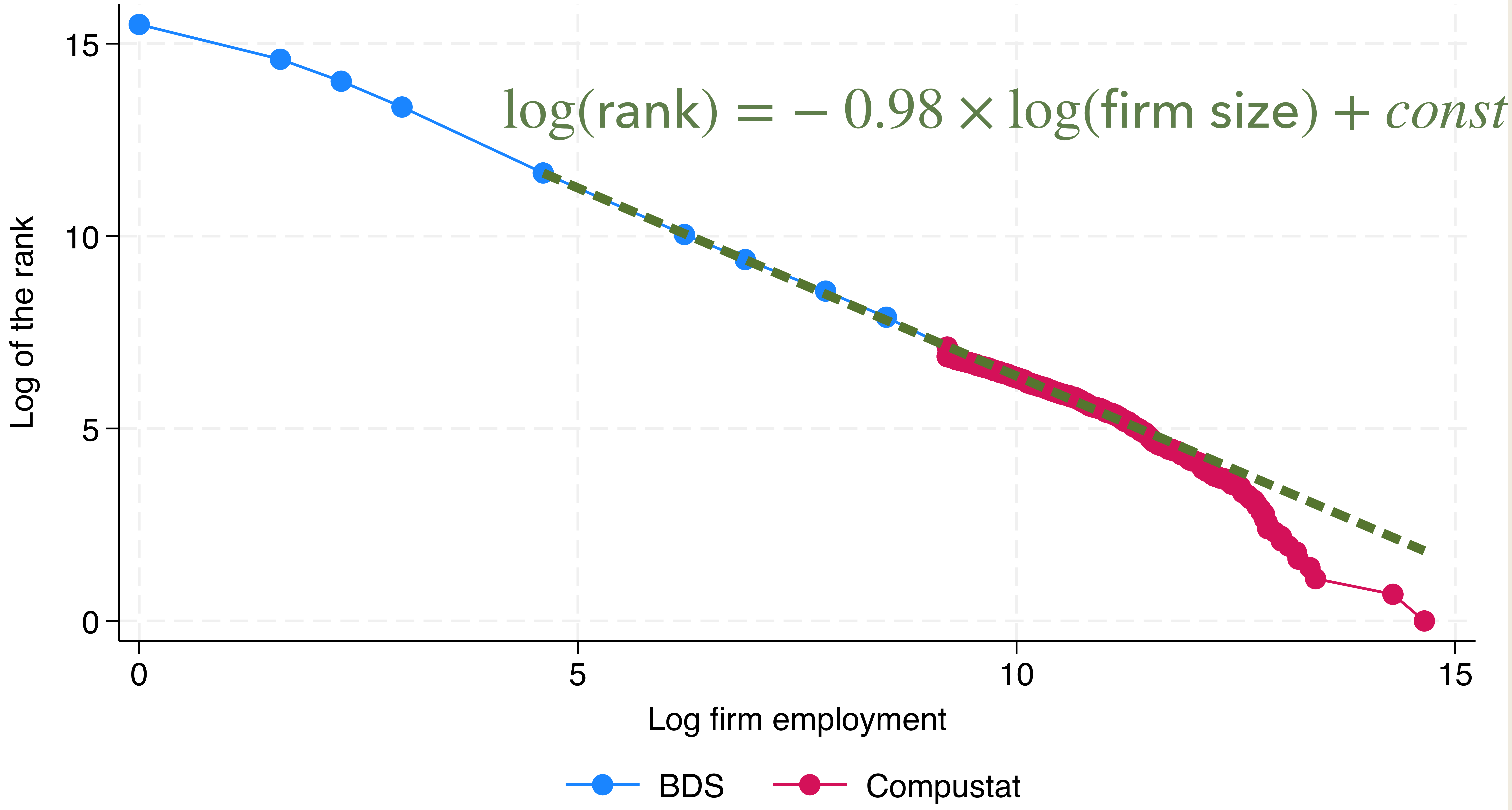


Power Law in Firm Size Distribution



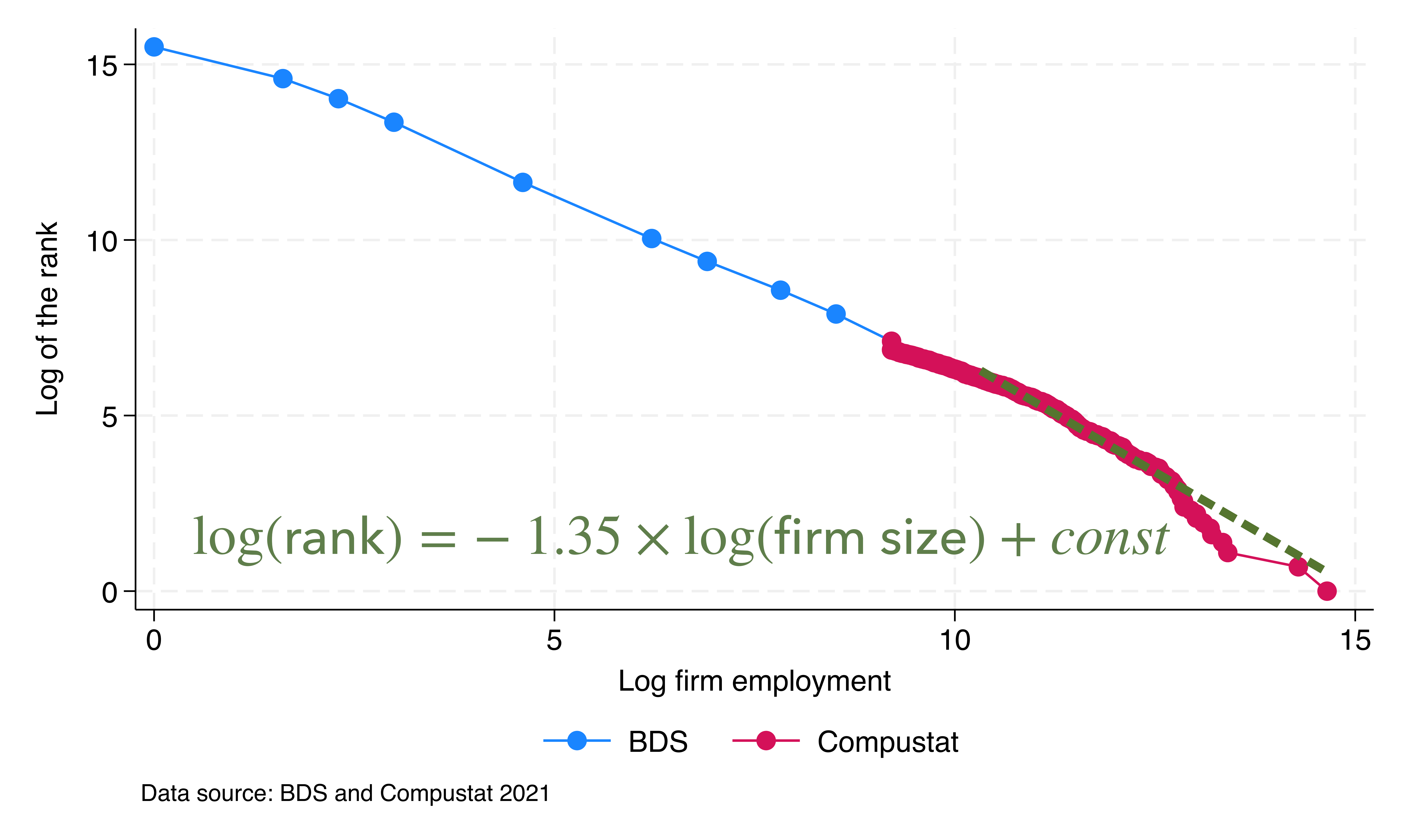
Data source: BDS and Compustat 2021

Power Law in Firm Size Distribution



Data source: BDS and Compustat 2021

Power Law in Firm Size Distribution



Two Facts in Firm Size Distribution

- Two surprises:
  1. The ranking of firm size is log-linear in firm size (**Power law**)
  2. The coefficient is close to one (**Zipf's law**)
- Mathematically,

$$\log \underbrace{\text{Pr}(\tilde{x} \geq x)}_{\text{ranking}} = -\zeta \log x + const, \qquad \zeta \approx 1$$

- What is this distribution?

  – Pareto: $\text{Pr}(\tilde{x} \geq x) = (x/\underline{x})^{-\zeta}$

Power Laws in Economics

"Paul Samuelson (1969) was once asked by a physicist for a law in economics that was both nontrivial and true… Samuelson answered, 'the law of comparative advantage.'

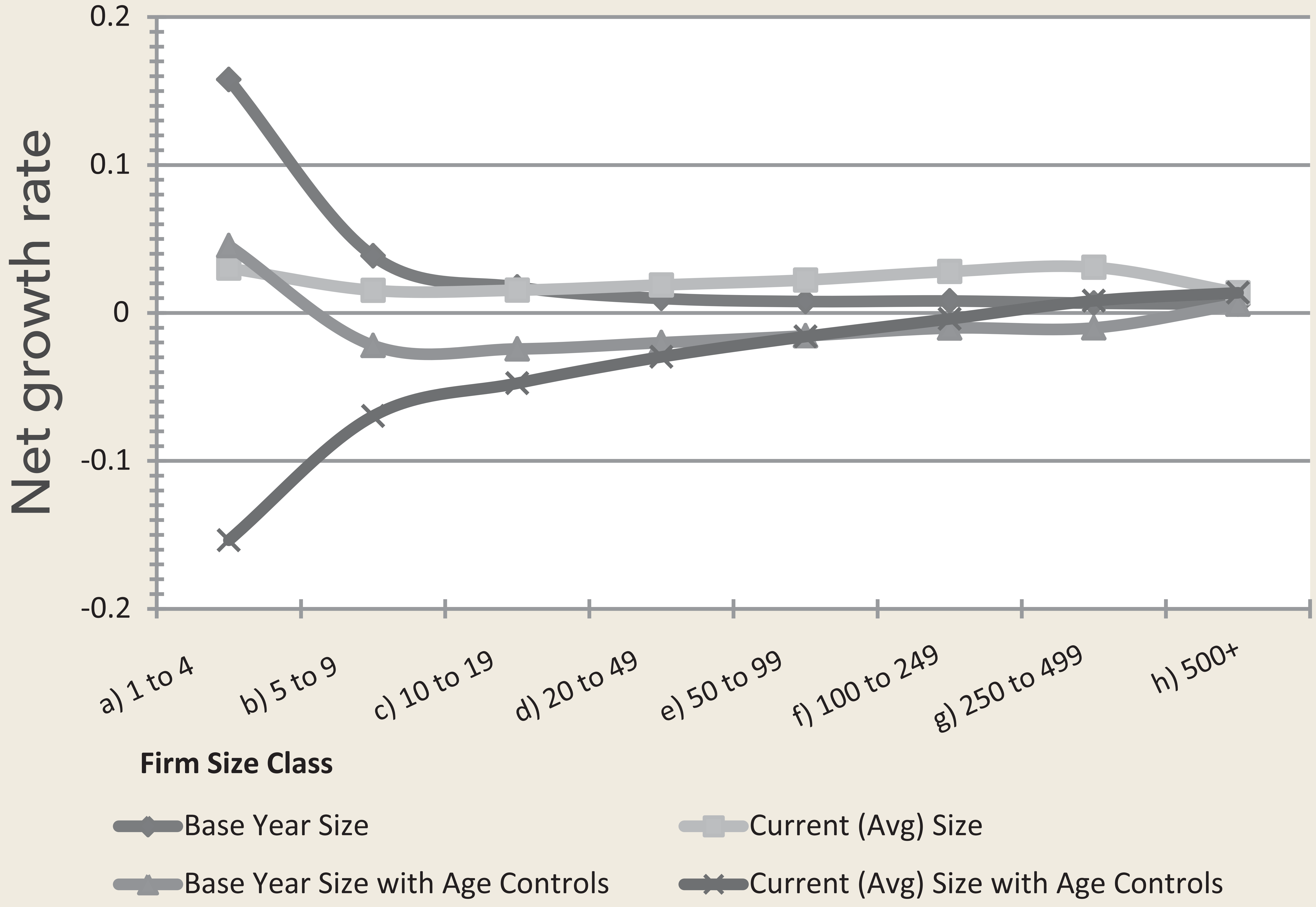
A modern answer to the question posed to Samuelson would be that a series of power laws count as actually nontrivial and true laws in economics."

— Gabaix (2016)

The Nature of Firm Growth

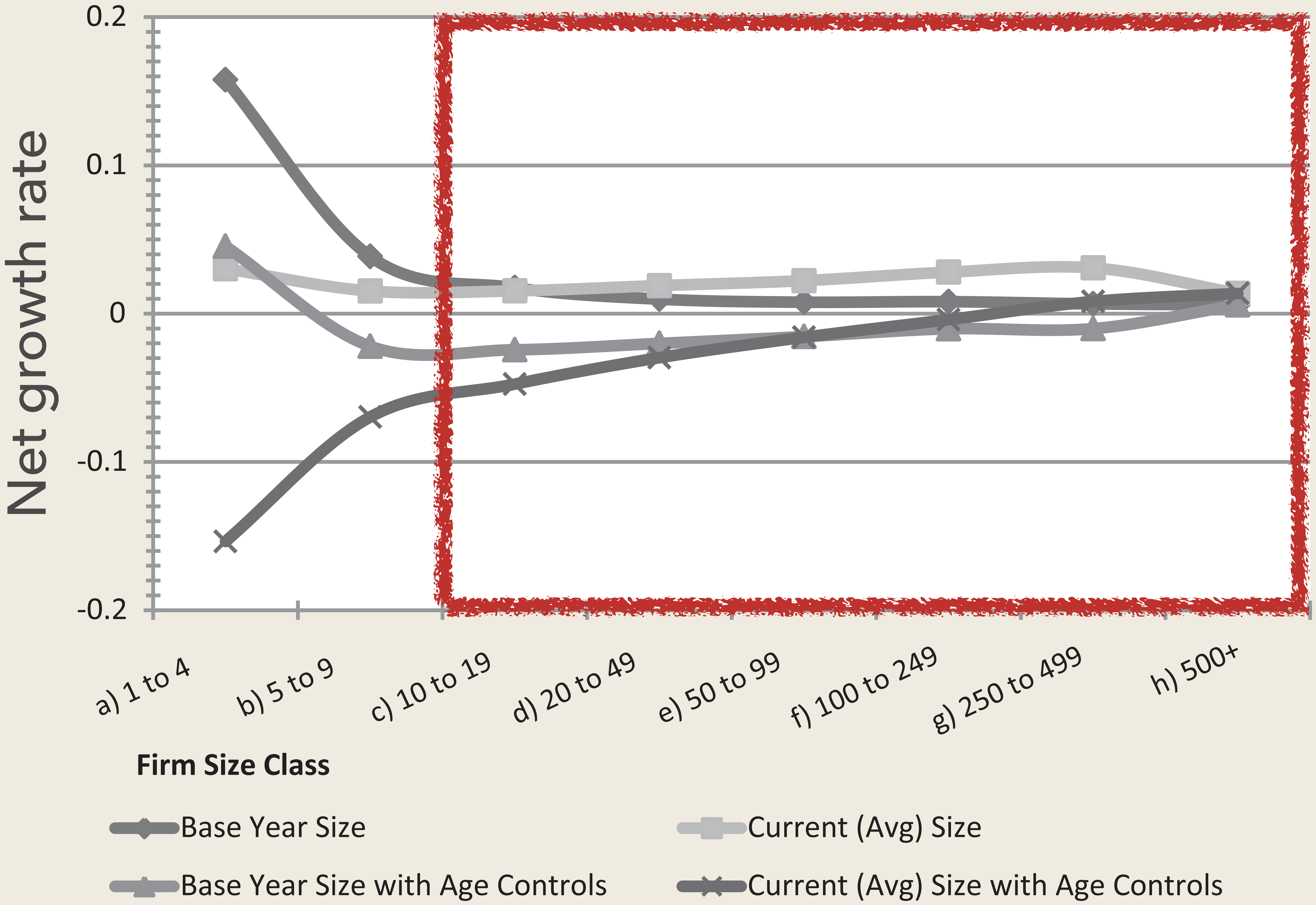
- How do large firms grow going forward?
  - Do they systematically shrink? (i.e., mean reversion in firm size)
  - Do they keep outperforming other smaller firms?
- Look at the relationship between firm growth and firm size

Firm Growth and Firm Size



Source: Haltiwanger, Jarmin & Mirand (2013)

Firm Growth and Firm Size



Source: Haltiwanger, Jarmin & Mirand (2013)

Gibrat's Law

- Firm growth rate is roughly independent of firm size…
  … if we exclude small firms
- This is called Gibrat's law

A Mechanical Model of Firm Size Distribution with Continuous-Time Toolkits

Connecting Two Laws

- Two robust features of the firm dynamics
  1. Power law
  2. Gibrat's law

- Gabaix (1999): Gibrat's law $\Rightarrow$ Power law

Continuous-Time Toolkits

– Diffusion and Kolmogorov Forward Equation

Brownian Motion

- **Definition:** a standard Brownian motion is a stochastic process Z_t with
  1. $Z_{t+s} - Z_t \sim N(0,s)$
  2. $Z_{t+s} - Z_t$ is independent of Z_t
- A continuous time version of (Gaussian) random walk: $Z_{t+1} = Z_t + \epsilon_t,\ \epsilon_t \sim N(0,1)$
- A Brownian motion with drift μ and variance σ^2 is given by

$$X_t = X_0 + \mu t + \sigma Z_t$$

  where Z_t is a standard Brownian motion

- Alternatively, we can write

$$dX_t = \mu dt + \sigma dZ_t$$

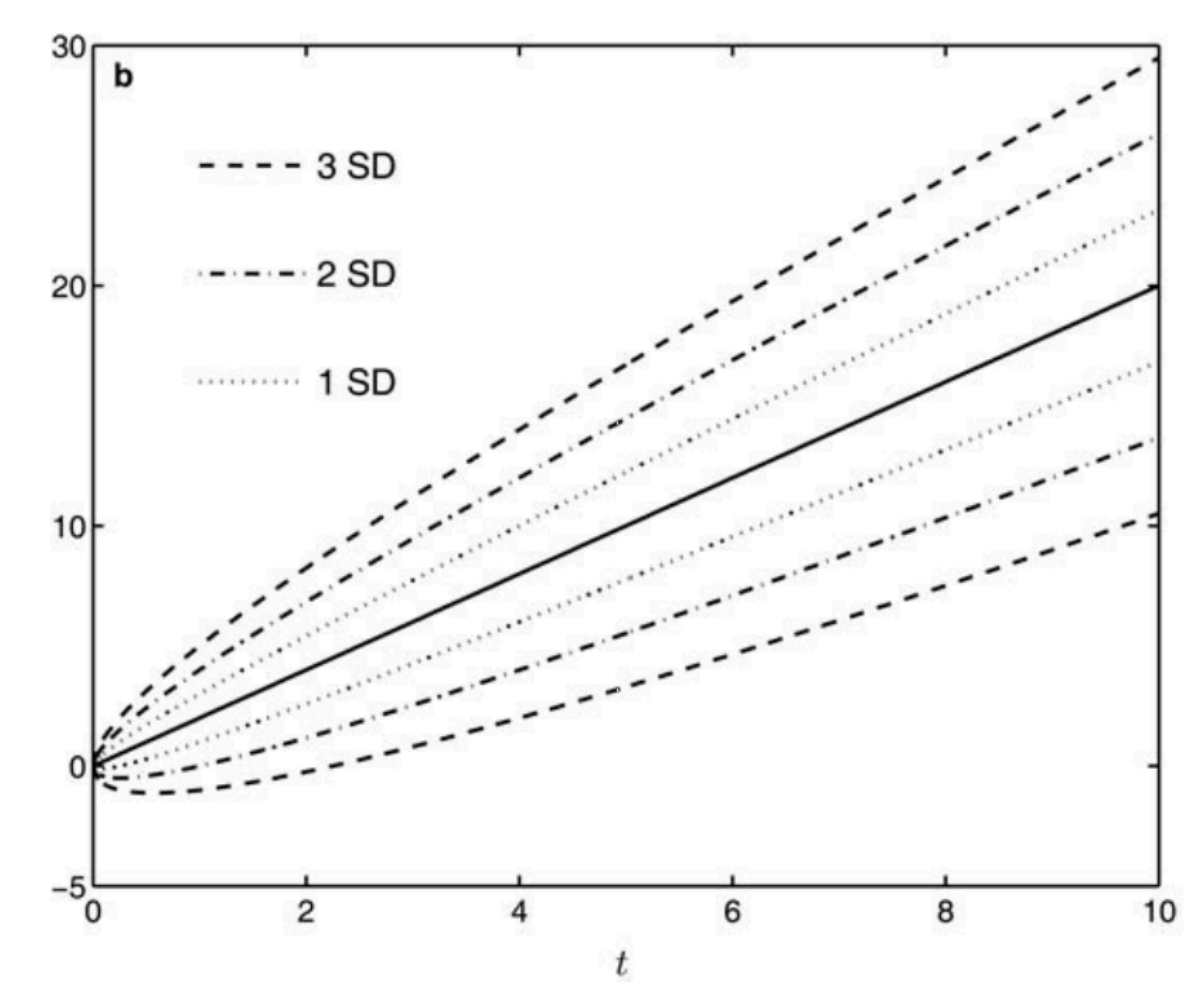
Visualizing Brownian Motion

- Mean and variance of Brownian motion:

$$\mathbb{E}[X_t - X_0] = \mu t, \quad \text{Var}[X_t - X_0] = \sigma^2 t$$

or

$$\mathbb{E}[dX_t] = \mu dt, \quad \text{Var}[dX_t] = \sigma^2 dt$$



Diffusion Process

- More generally, a diffusion process X_t is

$$dX_t = \mu(X_t)dt + \sigma(X_t)dZ_t$$

  - Brownian motion: $\mu(X_t) = \mu$, $\sigma(X_t) = \sigma$
  - Geometric Brownian motion: $\mu(X_t) = \mu X_t$, $\sigma(X_t) = \sigma X_t$
  - Ornstein-Uhlenbeck process: $\mu(X_t) = -\alpha X_t$, $\sigma(X_t) = \sigma$
    - Continuous time version of AR(1) process

- Note $\mathbb{E}[dX_t] = \mu(X_t)dt$ and $\text{Var}(dX_t) = \sigma^2(X_t)dt$

- A diffusion is a continuous-time version of a Markov process but rules out jumps

Discrete Time Approximation

- Discrete-time $t = \Delta t, 2\Delta t, \ldots$
- Consider

$$\Delta X_t \equiv X_{t+\Delta t} - X_t = \begin{cases} \mu(X_t)\Delta t + \sigma(X_t)\sqrt{\Delta t} & \text{with prob } 1/2 \\ \mu(X_t)\Delta t - \sigma(X_t)\sqrt{\Delta t} & \text{with prob } 1/2 \end{cases}$$

- Then

$$\mathbb{E}[\Delta X_t] = \mu(X_t)\Delta t, \quad \text{Var}(\Delta X) = \sigma^2(X_t)\Delta t$$

What is the Implied Distribution?

- Suppose X_t follows diffusion process
  - We will model firm growth through a diffusion process
- How does the distribution of X_t evolve?
  - This gives us the implied firm size distribution
  - Let $G_t(X) \equiv \text{Prob}(X_t \leq X)$ be the cdf and $g_t(X) = \partial_X G_t(X)$ be the pdf

Kolmogorov Forward Equation

- If X_t follows diffusion, $dX_t = \mu(X_t)dt + \sigma(X_t)dZ_t$, then $g_t(X) \equiv \partial_X G_t(X)$ follows

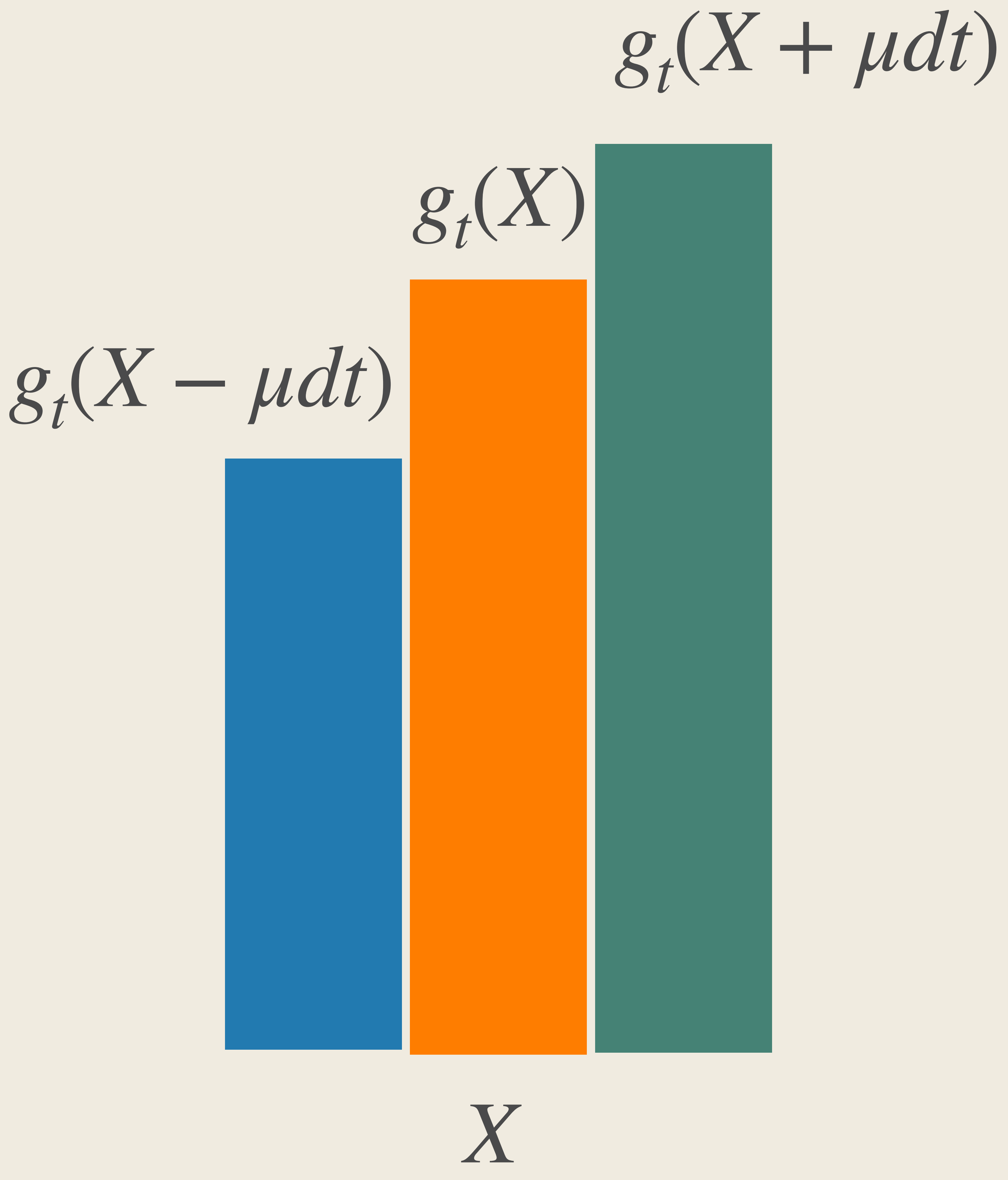
$$\partial_t g_t(X) = -\partial_X[\mu(X)g_t(X)] + \frac{1}{2}\partial^2_{XX}\left[\sigma(X)^2 g_t(X)\right]$$

  which is a partial differential equation called **Kolmogorov Forward equation**

- What is the intuition? Assume $\mu(X) = \mu > 0$ and $\sigma(X) = \sigma$ for simplicity.

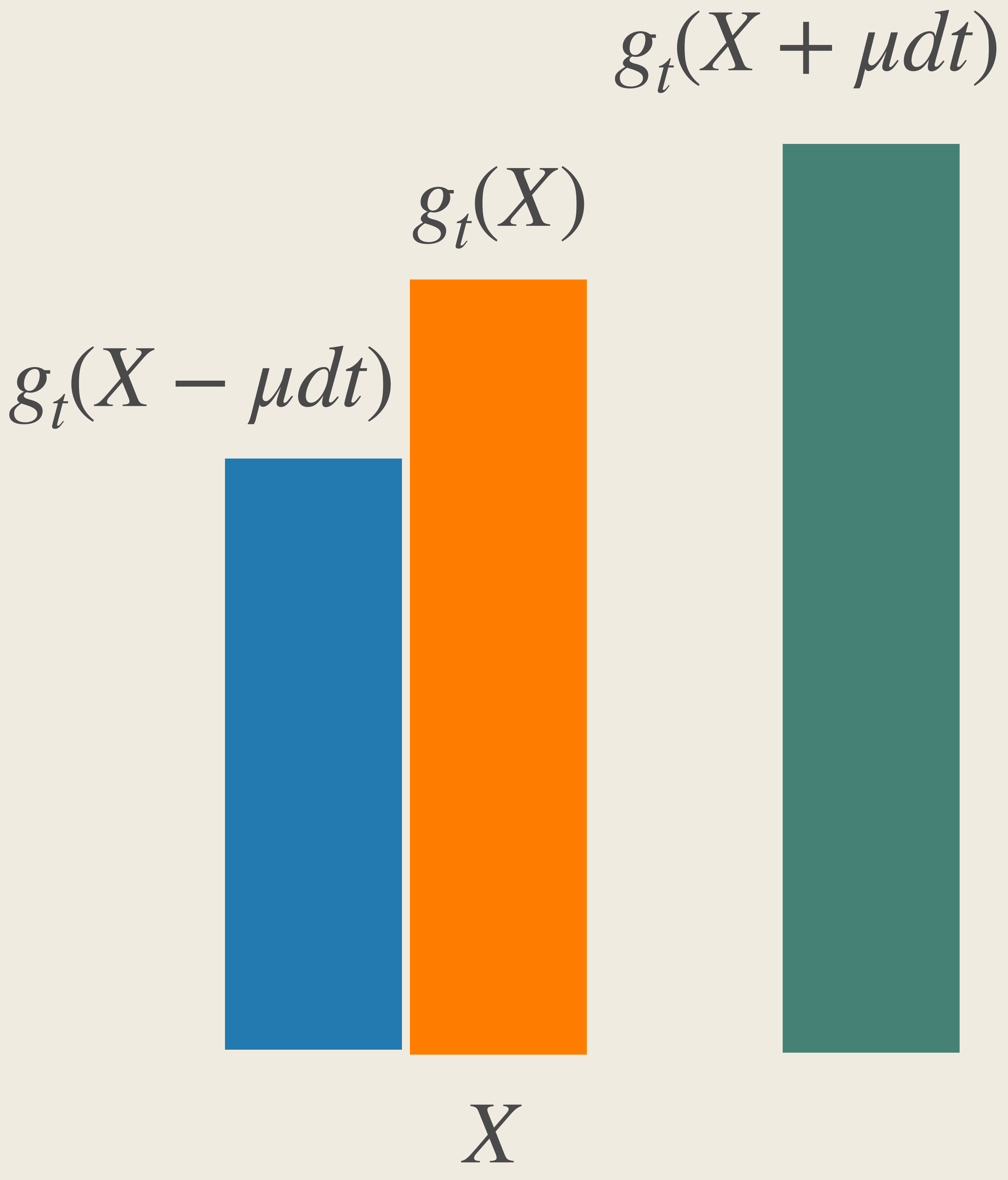
Intuition for Drift Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X) \right]$$



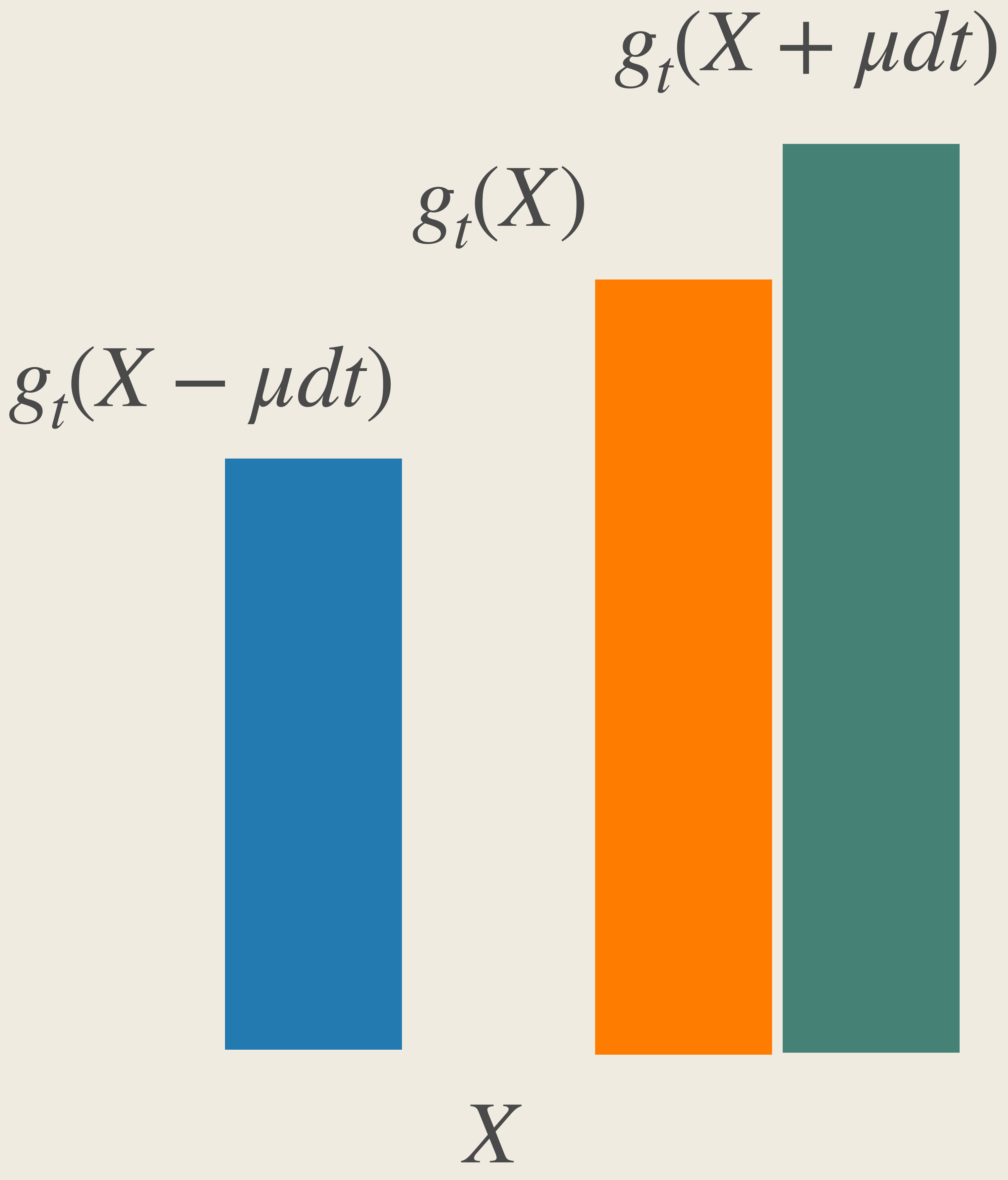
Intuition for Drift Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X) \right]$$



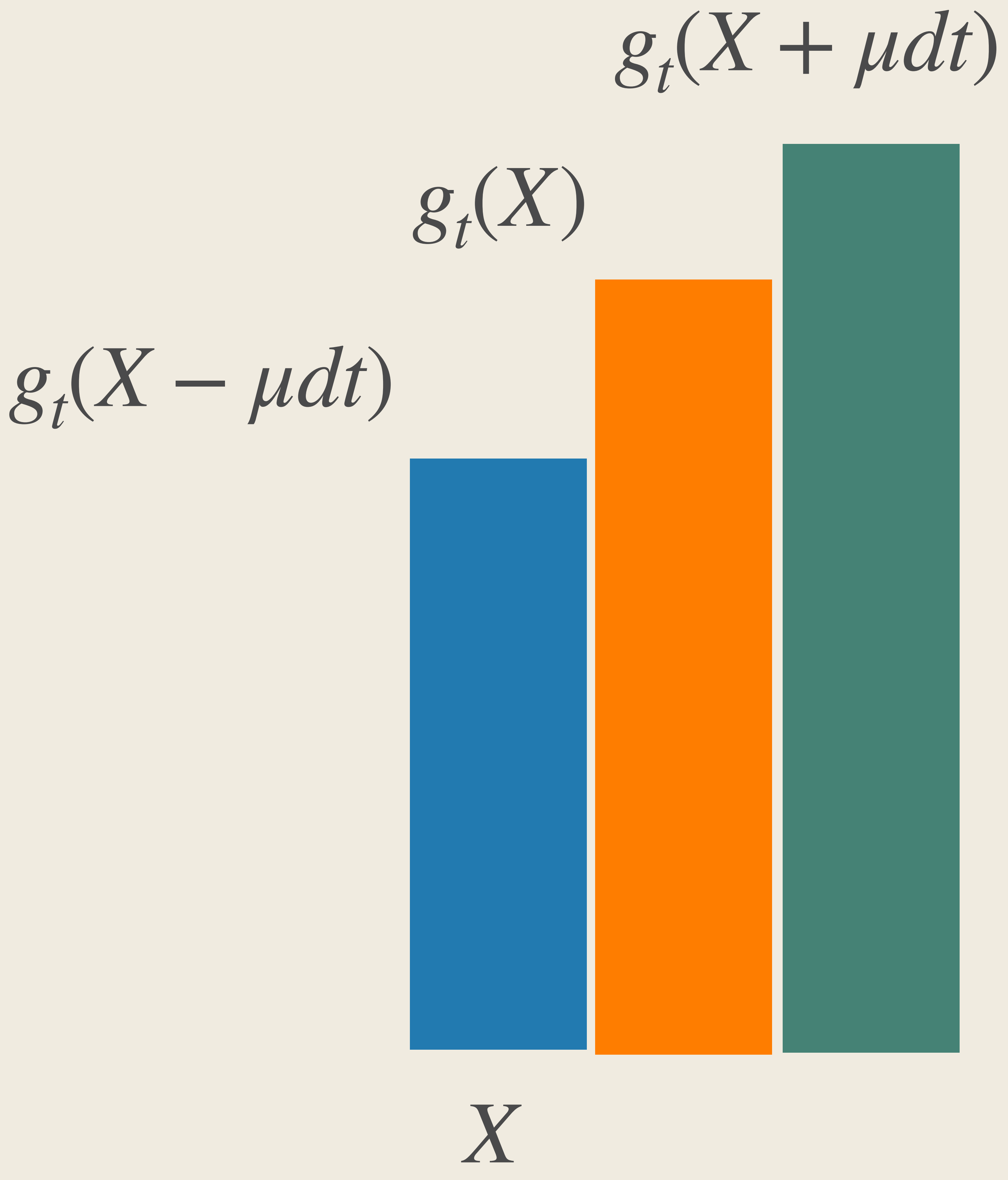
Intuition for Drift Term

$$\partial_t g_t(X) = -\mu\partial_X[g_t(X)] + \frac{\sigma^2}{2}\partial^2_{XX}\left[g_t(X)\right]$$



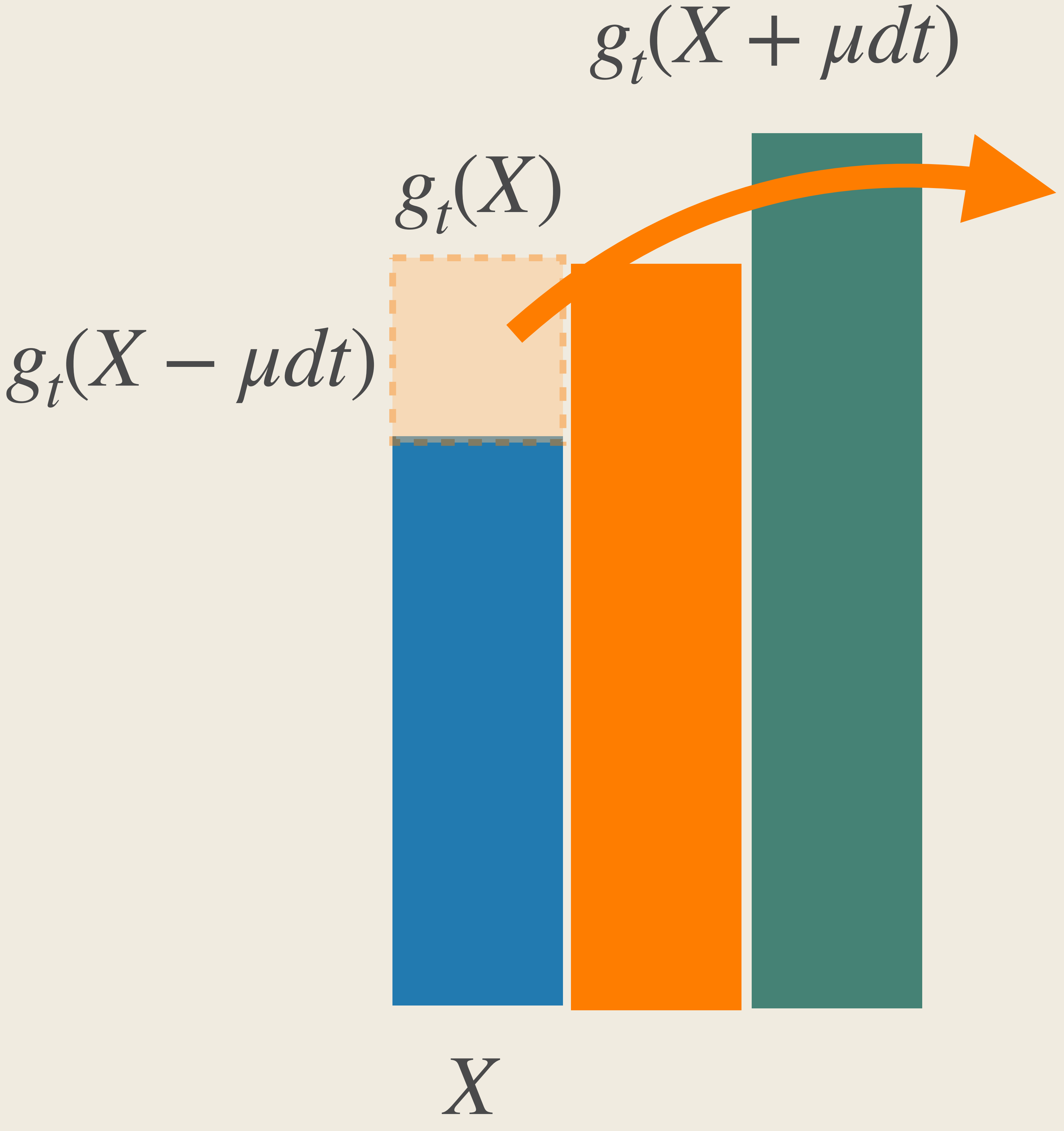
Intuition for Drift Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X)\right]$$



Intuition for Drift Term

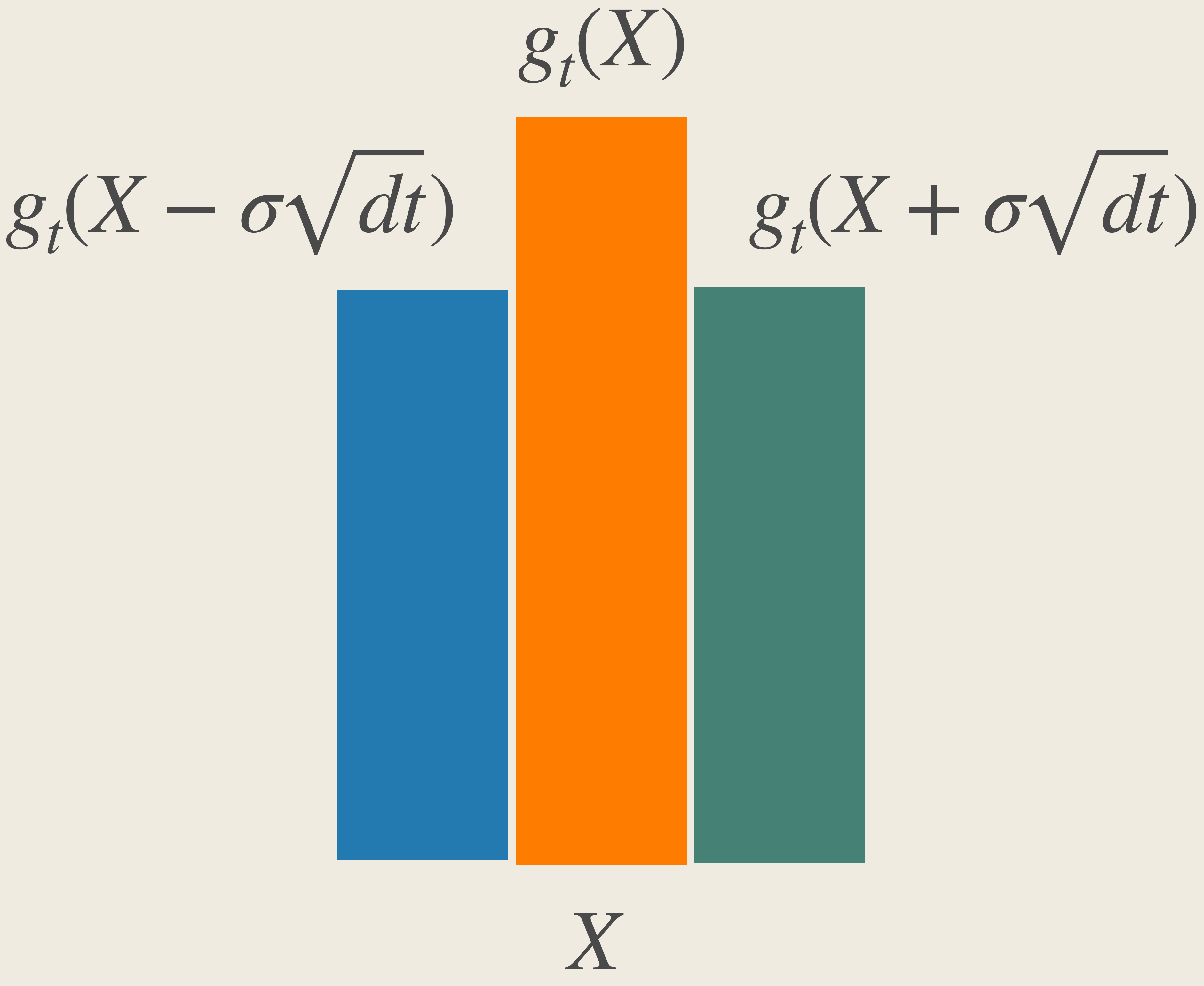
$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X)\right]$$



$$\Delta g_t(x) = g_t(X) - g_t(X - \mu dt)$$
$$= -\mu \partial_X g_t(X) dt \quad \text{as} \quad dt \to 0$$

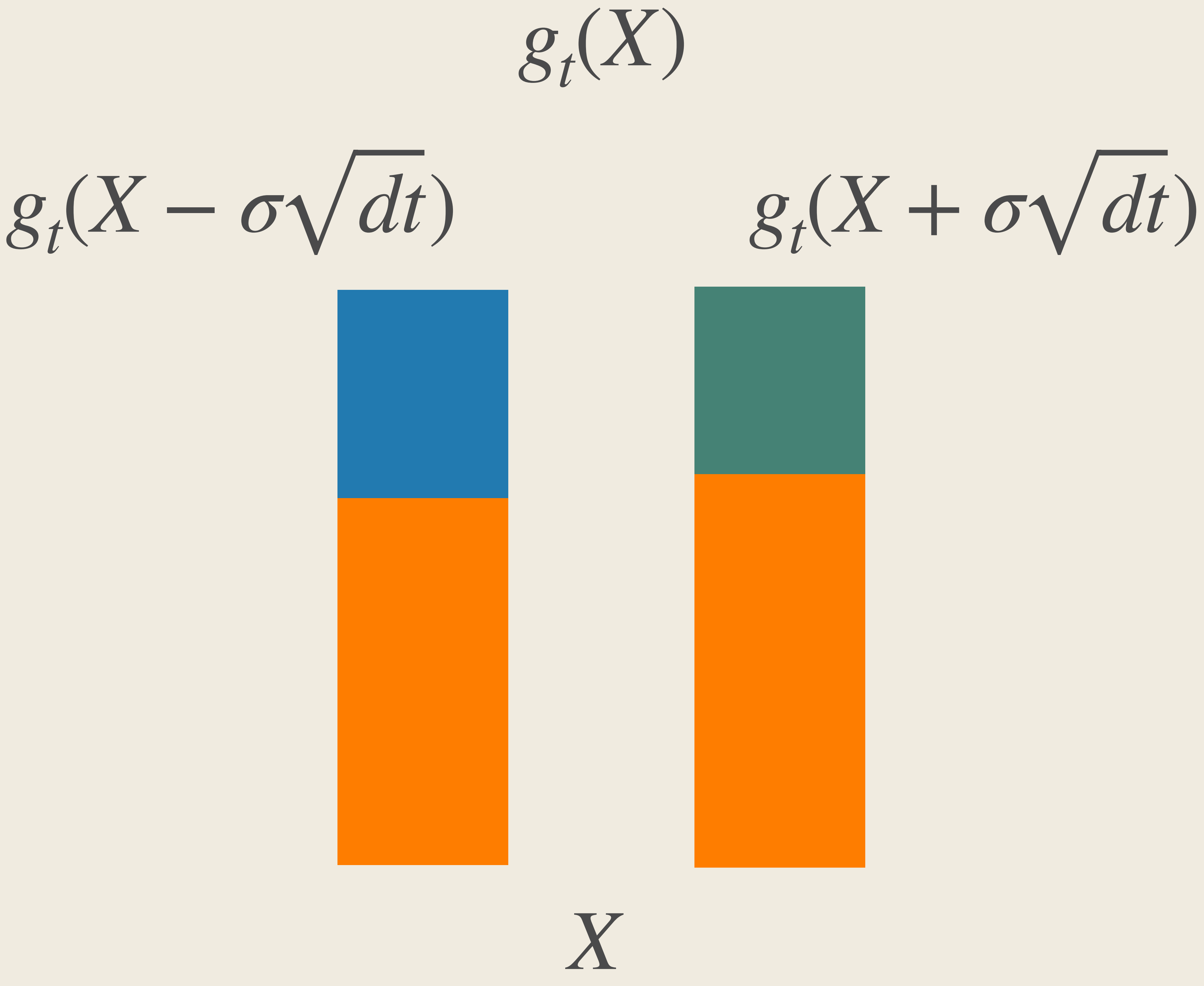
Intuition for Variance Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X) \right]$$



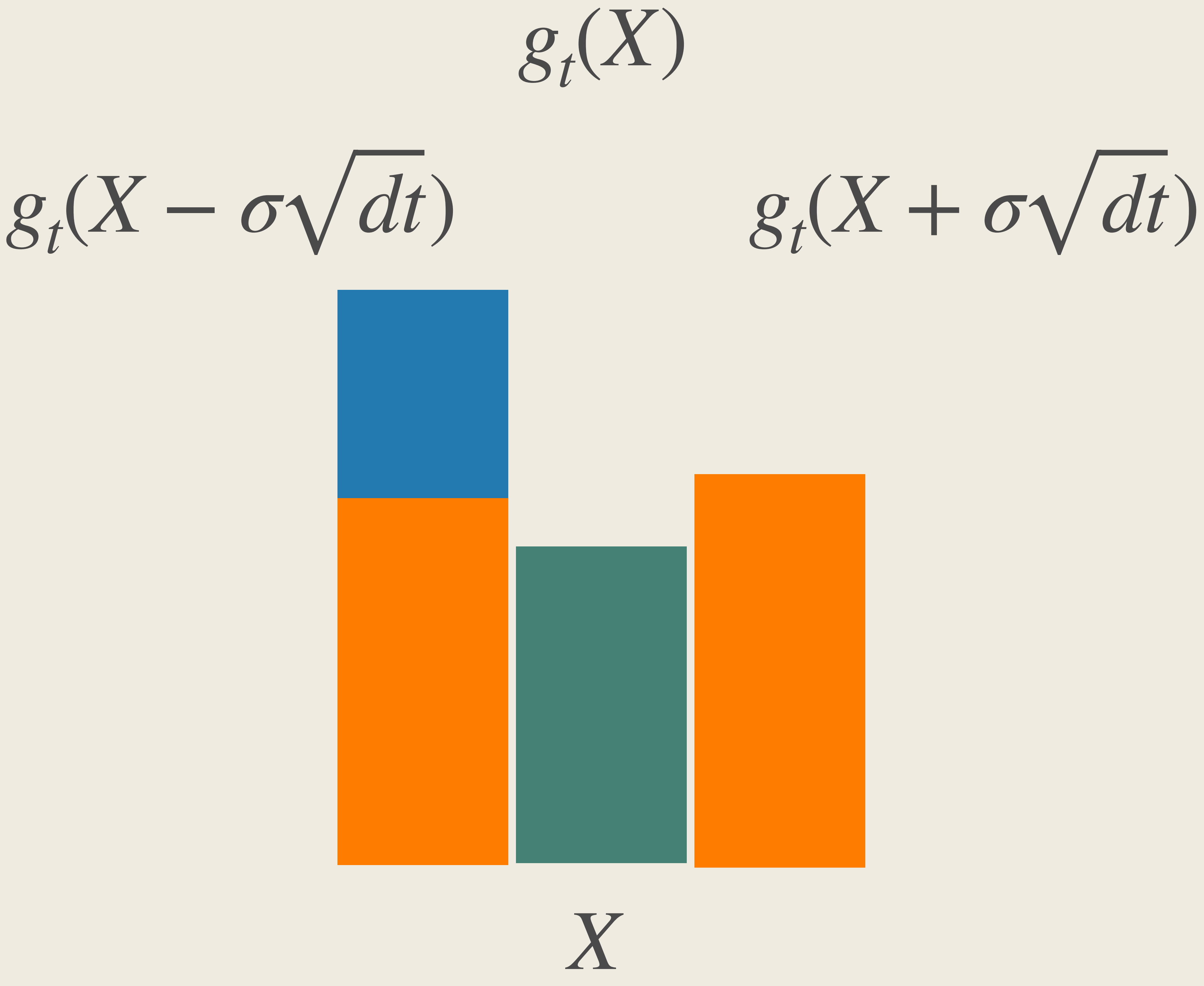
Intuition for Variance Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X) \right]$$



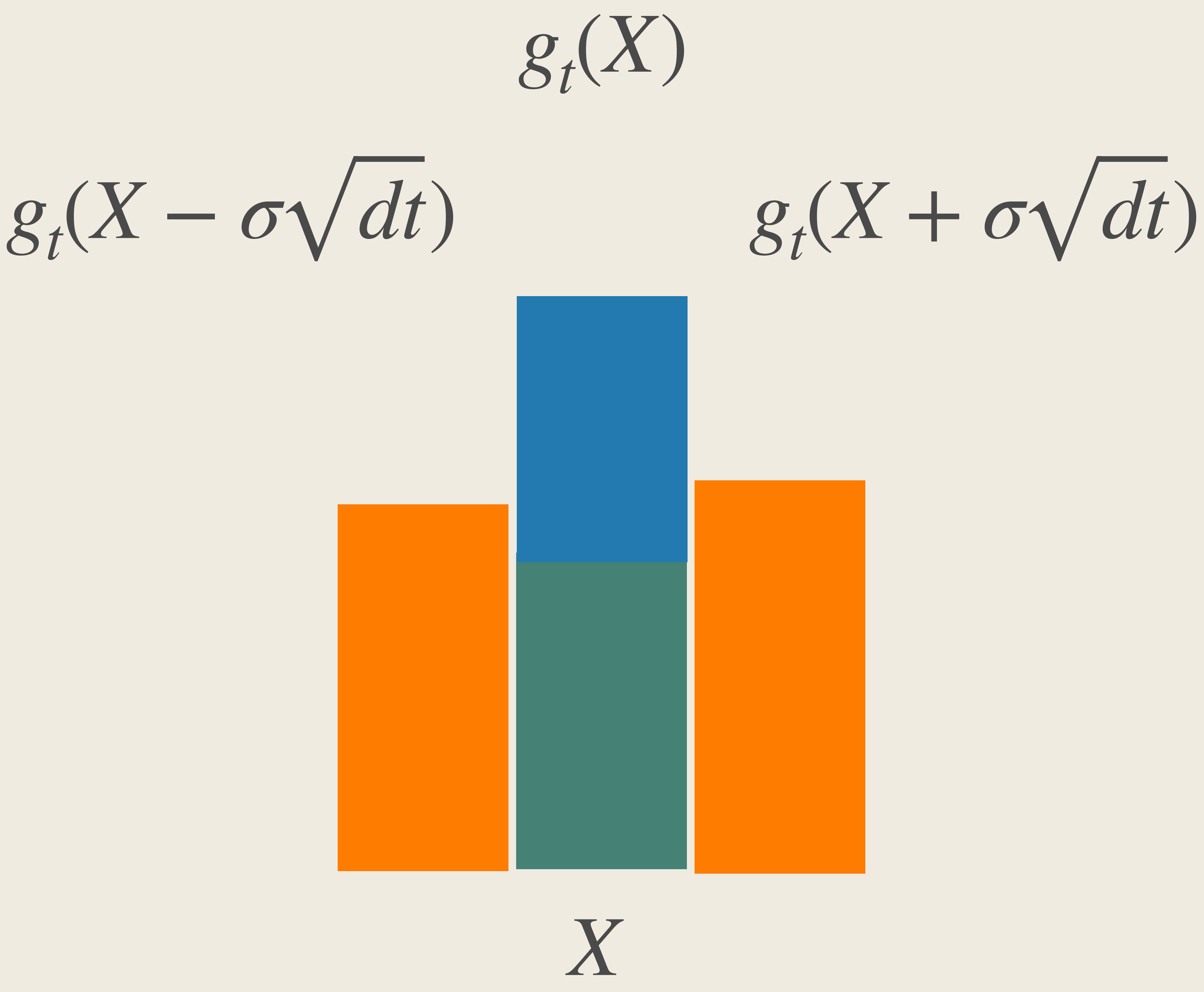
Intuition for Variance Term

$$\partial_t g_t(X) = -\mu \partial_X [g_t(X)] + \frac{\sigma^2}{2} \partial^2_{XX} \left[g_t(X) \right]$$



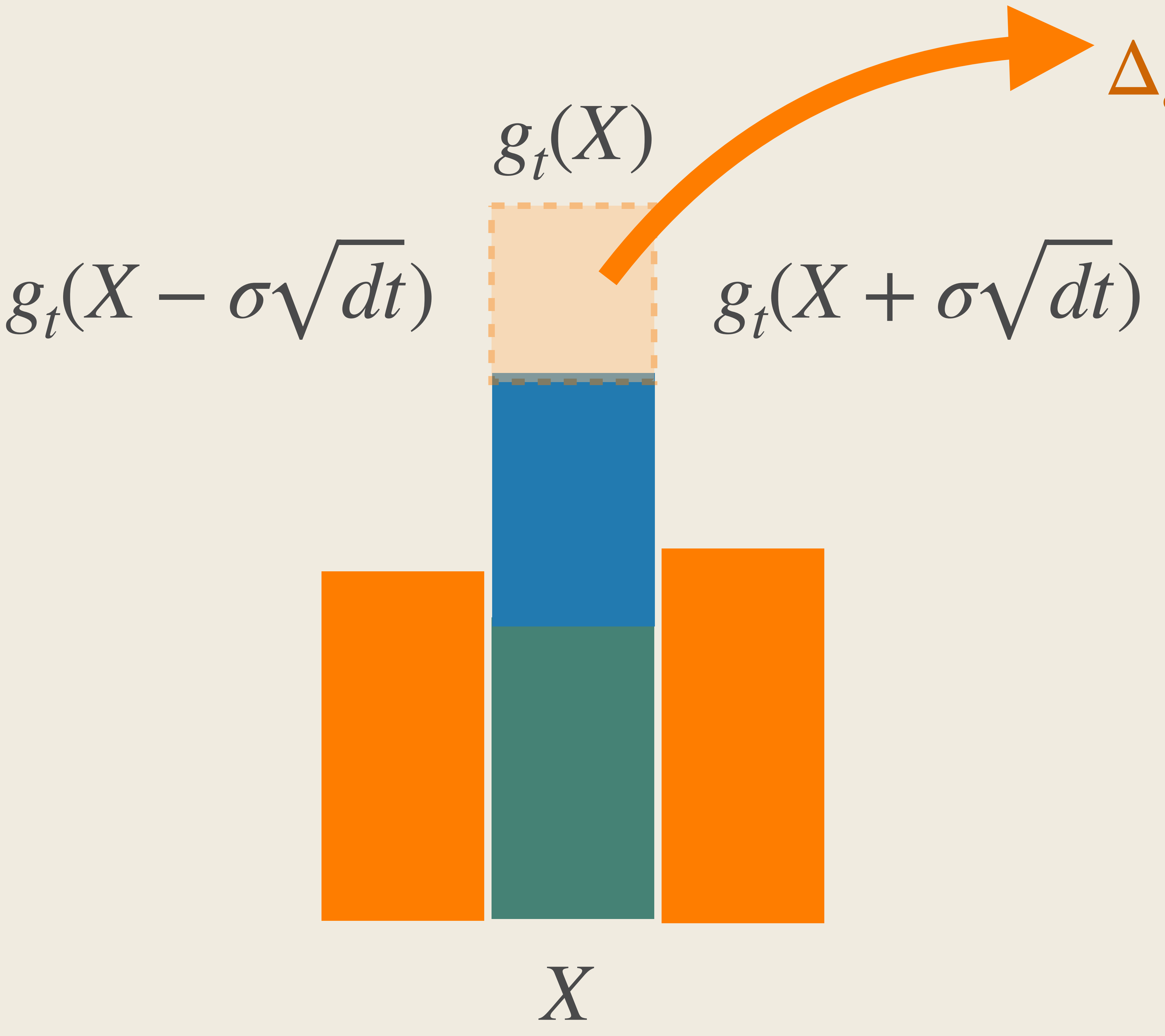
Intuition for Variance Term

$$\partial_t g_t(X) = -\mu\partial_X[g_t(X)] + \frac{\sigma^2}{2}\partial^2_{XX}\left[g_t(X)\right]$$



Intuition for Variance Term

$$\partial_t g_t(X) = -\mu\partial_X[g_t(X)] + \frac{\sigma^2}{2}\partial^2_{XX}\left[g_t(X)\right]$$



$$\Delta g_t(x) = \frac{1}{2}g_t(X - \sigma\sqrt{dt}) + \frac{1}{2}g_t(X + \sigma\sqrt{dt}) - g_t(X)$$

$$\approx \frac{1}{2}\left[g_t(X) - \sigma\partial_X g_t(X)\sqrt{dt} + \frac{\sigma^2}{2}\partial^2_{XX}g_t(X)dt\right]$$

$$+\frac{1}{2}\left[g_t(X) + \sigma\partial_X g_t(X)\sqrt{dt} + \frac{\sigma^2}{2}\partial^2_{XX}g_t(X)dt\right] - g_t(X)$$

$$= \frac{\sigma^2}{2}\partial^2_{XX}g_t(X)dt$$

Heuristic Proof (1/2)

- Let dX_t be the change in X_t over a time interval dt
- Let $p(dX_t, X_t)$ be density over dX_t
- The changes in density $g_t(X_t)$ over a time interval dt is

$$\Delta g_t(X_t) = \int \left(\underbrace{-p(dX_t, X_t)g_t(X_t)}_{\text{outflow}} + \underbrace{p(dX_t, X_t - dX_t)g(X_t - dX_t)}_{\text{inflow}} \right) d(dX_t) \tag{1}$$

- Taylor-expand the inflow around $dX_t = 0$:

$$\begin{aligned} p(dX_t, X_t - dX_t)g(X_t - dX_t) \approx\ & p(dX_t, X_t)g(X_t) - \partial_X[p(dX_t, X_t)g(X_t)]dX_t \\ & + \frac{1}{2}\partial^2_{XX}[p(dX_t, X_t)g(X_t)](dX_t)^2 \end{aligned} \tag{2}$$

Heuristic Proof (2/2)

- Substitute back (2) into (1):

$$\Delta g_t(X_t) = \int \left(-\partial_X [p(dX_t, X_t) g_t(X_t)] dX_t + \frac{1}{2}\partial^2_{XX}[p(dX_t, X_t) g_t(X_t)](dX_t)^2 \right) d(dX_t)$$

$$= -\partial_X \left[\underbrace{\int \left(p(dX_t, X_t) dX_t \right) d(dX_t)}_{=\mu(X_t)dt} g_t(X_t) \right] + \frac{1}{2}\partial^2_{XX} \left[\underbrace{\int \left(p(dX_t, X_t)(dX_t)^2 \right) d(dX_t)}_{=\sigma(X_t)^2 dt} g_t(X_t) \right]$$

$$= -\partial_X \left[\mu(X_t) g_t(X_t) \right] dt + \frac{1}{2}\partial^2_{XX} \left[\sigma(X_t)^2 g_t(X_t) \right] dt$$

Steady State Distribution

- Corollary: Steady-state distribution, $g_t(X) = g(X)$, if it exists, solves

$$0 = -\partial_X[\mu(X)g(X)] + \frac{1}{2}\partial^2_{XX}\left[\sigma(X)^2 g(X)\right]$$

  - (Inflow into X) = (outflow from X)

- Steady-state distribution is characterized by a 2nd-order ODE

- This is a beauty of continuous time

A Mechanical Model of Firm Size Distribution

Firm Growth as a Stochastic Process

- Let n_t denote the firm size and n_t follows diffusion process
- Gibrat's law suggests n_t follows a geometric Brownian motion:

$$dn_t = \mu n_t dt + \sigma n_t dZ_t$$

$$\Leftrightarrow \quad \frac{dn_t}{n_t} = \mu dt + \sigma dZ_t$$

- One can show $\text{Var}(\log n_t) = \sigma^2 t$
  $\Rightarrow$ Distribution explodes as $t \to \infty \Rightarrow$ no steady-state distribution
- Gabaix's (1999) insight:
  Gibrat's law + stabilizing force $\Rightarrow$ SS distribution exists and features power law

Stabilizing Forces

- A particular approach undertaken by Gabaix (1999):
  - Minimum fim size requirement, $\underline{n}$:
    - ✓ If firms hit $\underline{n}$, they exit
    - ✓ The same mass of new firms with size $\underline{n}$ enter at the same time

- Stationary firm size distribution $g(n)$ solves

$$0 = -\partial_n[\mu n g(n)] + \frac{1}{2}\partial^2_{nn}\left[\sigma^2 n^2 g(n)\right] \quad \text{for } n > \underline{n}$$

with boundary conditions such that $\int_{\underline{n}}^{\infty} g(n)dn = 1$ and $g(n) \geq 0$ for all n

Power Law in Firm Size Distribution

Result: The solution is Pareto: $g(n) = \zeta \underline{n}^{\zeta} n^{-\zeta-1}$ with $\zeta = 1 - \frac{\mu}{2\sigma^2} > 0$

1. Integrate the ODE once to obtain (c_1, c_2 are integration constants)

$$c_1 = -2\mu n g(n) + \partial_n[\sigma^2 n^2 g(n)]$$
$$\Leftrightarrow \quad n^{\frac{-2\mu}{\sigma^2}} c_1 = \partial_n \left[n^{\frac{-2\mu}{\sigma^2}} \sigma^2 n^2 g(n) \right]$$

2. Integrate one more time

$$c_1 \int^n m^{\frac{-2\mu}{\sigma^2}} dm = n^{\frac{-2\mu}{\sigma^2}} \sigma^2 n^2 g(n) + c_2$$
$$\Leftrightarrow \quad g(n) = \tilde{c}_1 n^{-1} - \tilde{c}_2 n^{-\zeta-1},$$

   where $\tilde{c}_1 \equiv c_1/(\sigma^2 - 2\mu)$, $\tilde{c}_2 \equiv c_2/\sigma^2$.

3. Since $g(n)$ is pdf, $\int_{\underline{n}}^{\infty} g(n) dn = 1 \Rightarrow \tilde{c}_1 = 0$ and $\tilde{c}_2 = \zeta \underline{n}^{\zeta}$

Power Law and Zipf's Law

- The cdf is $G(n) = 1 - (n/\underline{n})^{-\zeta}$, so power law holds:

$$\log \Pr(\tilde{n} \geq n) = \log(1 - G(n)) = -\zeta \log n + const$$

- The existence of mean requires $\zeta > 1 \Leftrightarrow \mu < 0$
- What about Zipf's law? It holds if $\zeta = 1 - \frac{\mu}{2\sigma^2} \approx 1 \Leftrightarrow \mu \approx 0$
- The result is much more general than presented here:
  - random growth + stabilizing force
    $\Rightarrow$ asymptotic power law: $\Pr(\tilde{n} \geq n) \to cn^{-\zeta}$ as $n \to \infty$
  - stabilizing force $\approx 0 \Rightarrow$ Zipf's law

Numerically Computing Stationary Firm Size Distribution

How to Solve ODE on a Computer?

- Gabaix's (1999) case admits analytical solutions
- Easy to come up with variations that prevent analytical characterizations
  - For example, what if firm size follows a general diffusion with $\mu(n)$ and $\sigma(n)$?
- Even in these cases, one can always solve the following ODE numerically:

$$0 = -\partial_n[\mu(n)g(n)] + \frac{1}{2}\partial^2_{nn}\left[\sigma(n)^2 g(n)\right] \quad \text{for } n > \underline{n}$$

- How do we do that?

Discretization and Derivatives

- Discretize the firm-size space: $n \in \{n_1, n_2, \ldots, n_J\}$ with $n_1 = \underline{n}$ and equispaced grids:

$$\Delta n \equiv n_j - n_{j-1}$$

- We discretize the derivative $-\partial_n[\mu(n)g(n)]$ as well. Two-ways:

1. Forward difference approximation:

$$-\partial_n[\mu(n_i)g(n_i)] \approx -\frac{\mu(n_{i+1})g(n_{i+1}) - \mu(n_i)g(n_i)}{\Delta n}$$

2. Backward difference approximation:

$$-\partial_n[\mu(n_i)g(n_i)] \approx -\frac{\mu(n_i)g(n_i) - \mu(n_{i-1})g(n_{i-1})}{\Delta n}$$

- Use forward when $-\mu(n_i) > 0$ and backward when $-\mu(n_i) < 0$

- The second derivative is

$$\partial^2_{nn}\left[\sigma(n_i)^2 g(n_i)\right] \approx \frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2}$$

Discretized KFE

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

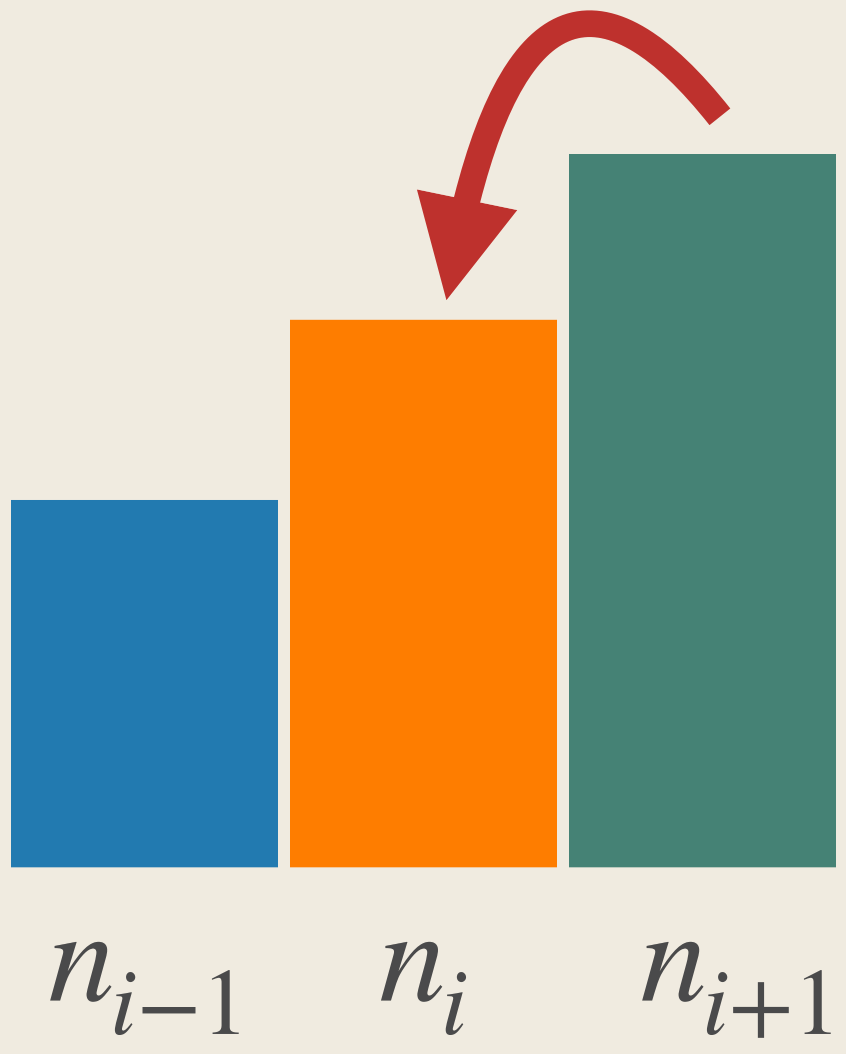
Discretized KFE

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

Inflow from $i + 1$ due to drift

Inflow from $i + 1$ due to variance



Discretized KFE

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$

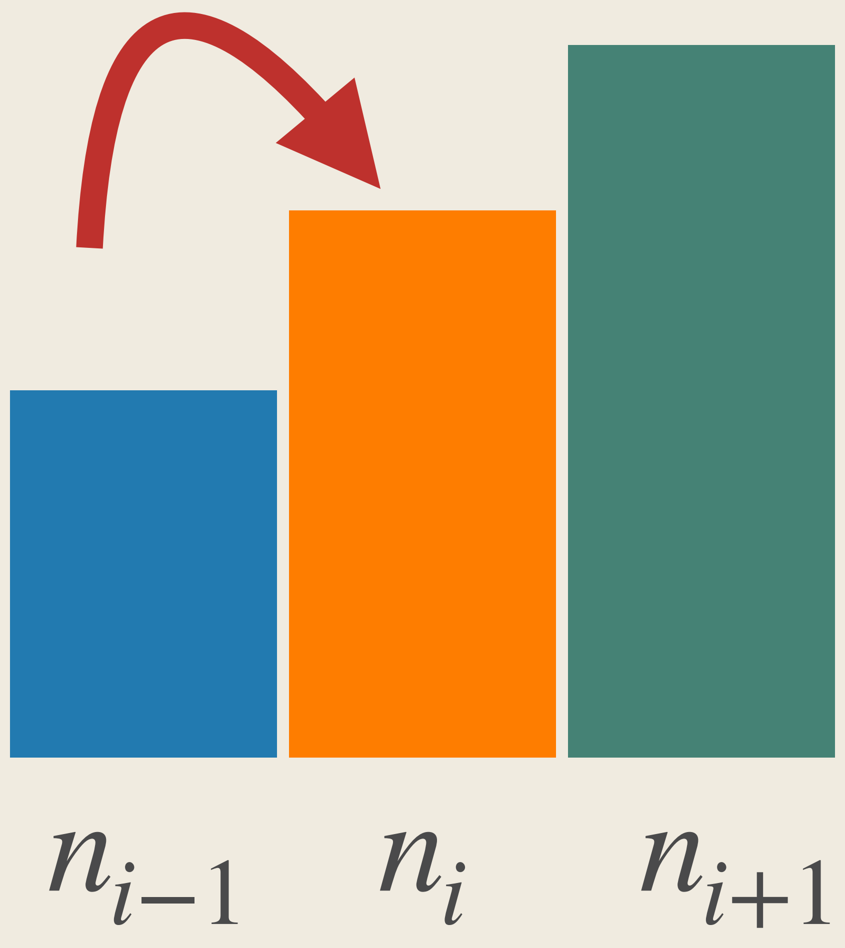
Discretize

Inflow from $i-1$ due to variance

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$



Discretized KFE

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

Discretiz

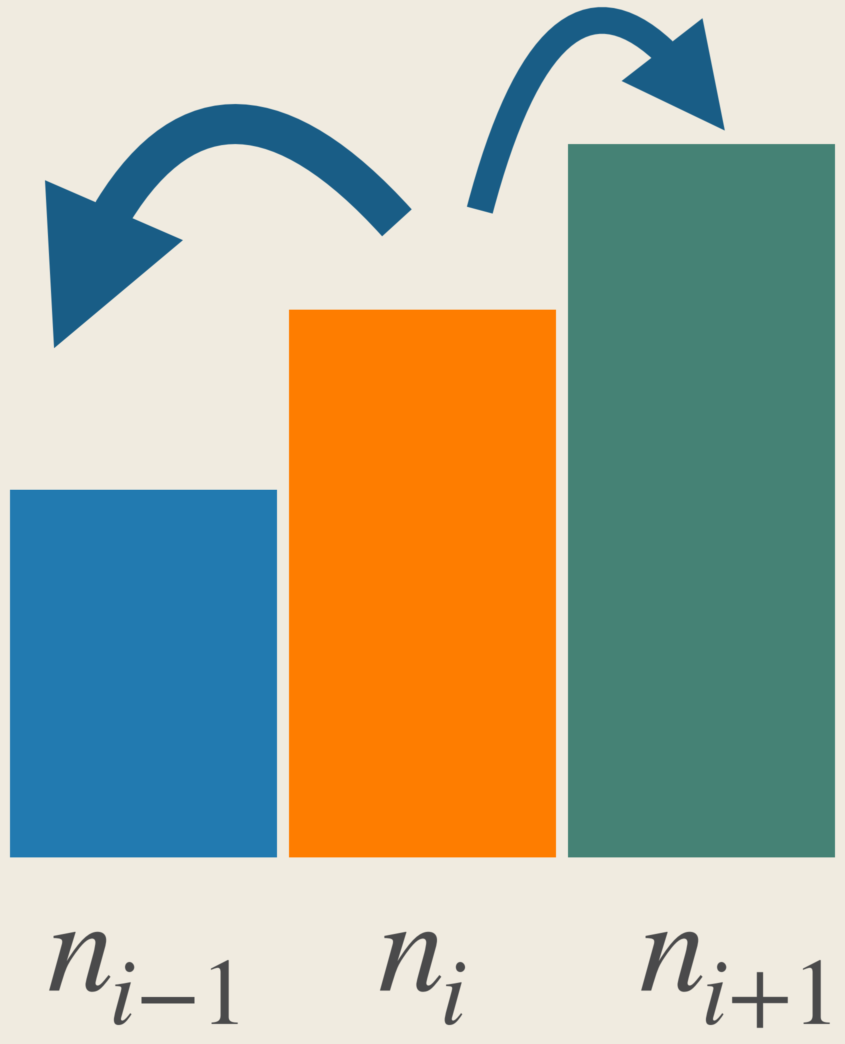
outflow from i due to variance

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J -$

outflow from i due to drift



Discretized KFE

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

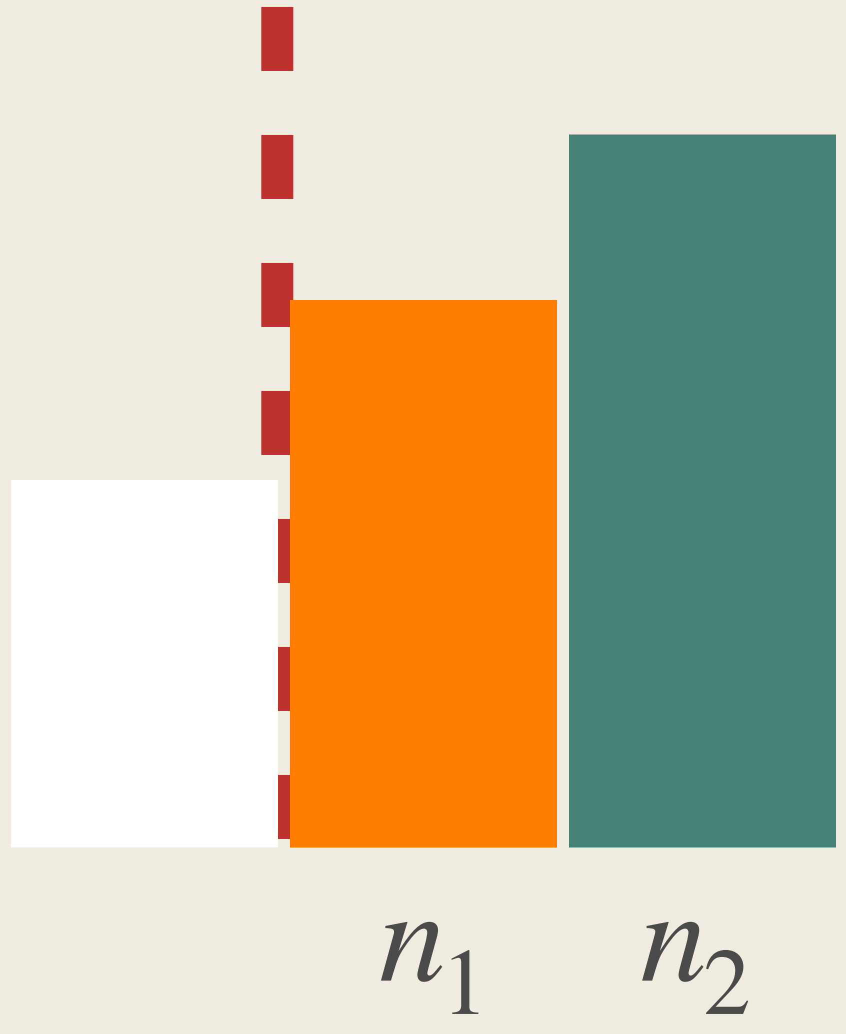
for $i = 1, \ldots, J-1$

Entry & Exit at Lower Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1})-2\sigma(n_i)^2 g(n_i)+\sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$



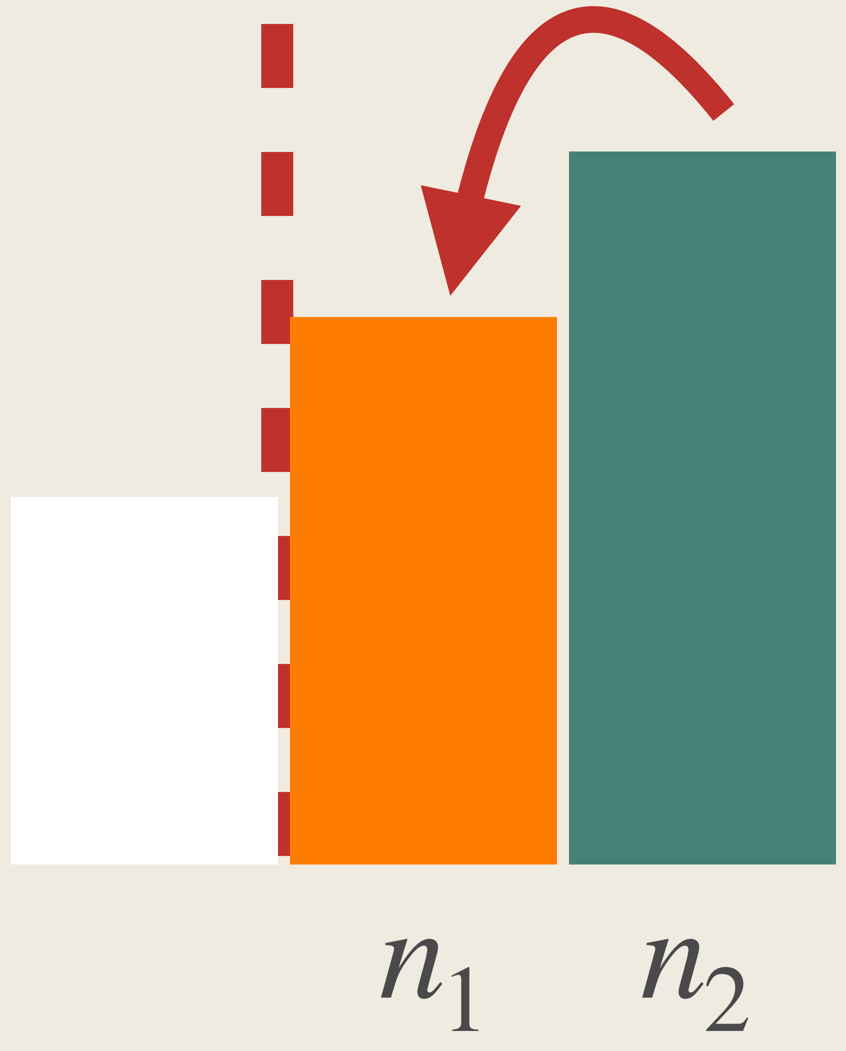
Entry & Exit at Lower Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

Inflow from $i+1$ due to drift

Inflow from $i+1$ due to variance

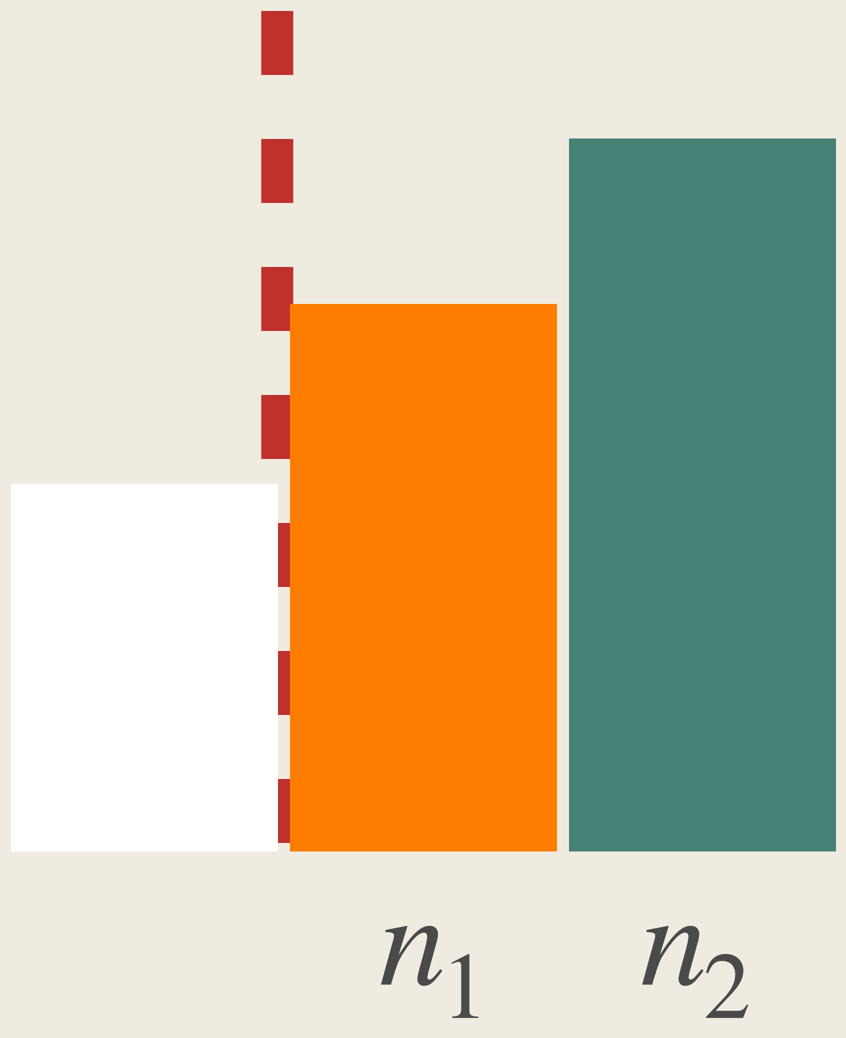


Entry & Exit at Lower Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n}+\frac{1}{2}\frac{\sigma(n_{i+1})^2g(n_{i+1})-2\sigma(n_i)^2g(n_i)+\sigma(n_{i-1})^2g(n_{i-1})}{(\Delta n)^2}=0$$

for $i = 1,\ldots,J-1$



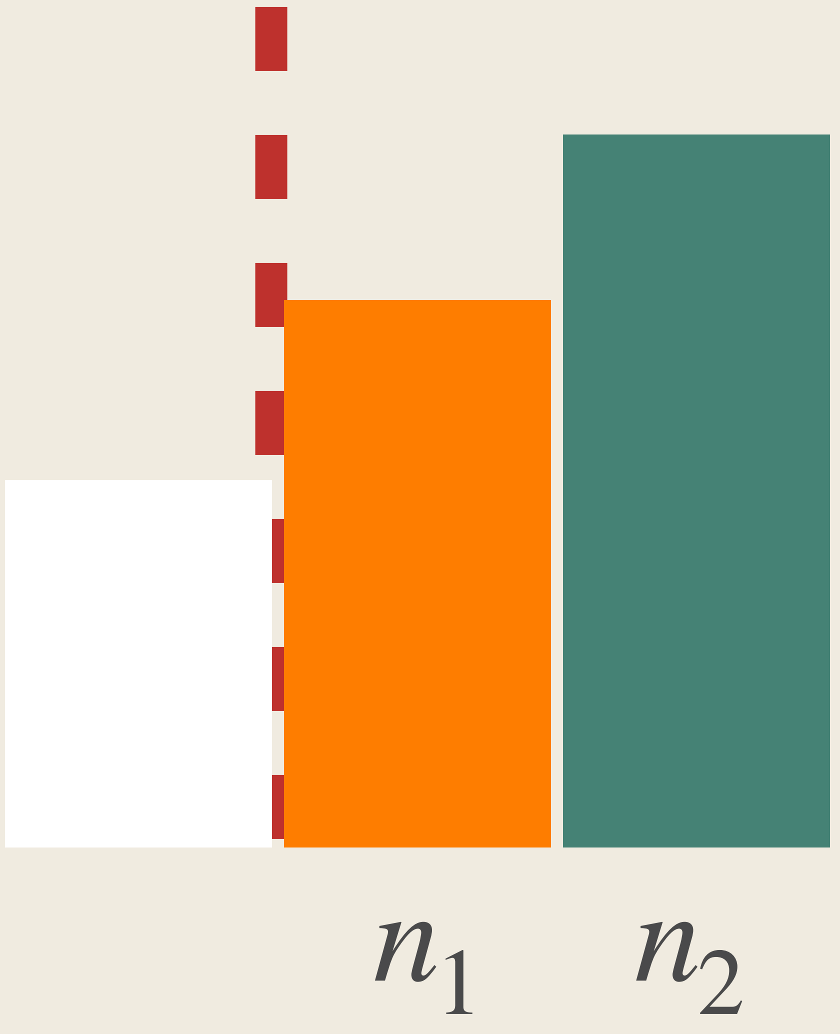
Entry & Exit at Lo

Inflow from $i - 1$ due to variance

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$

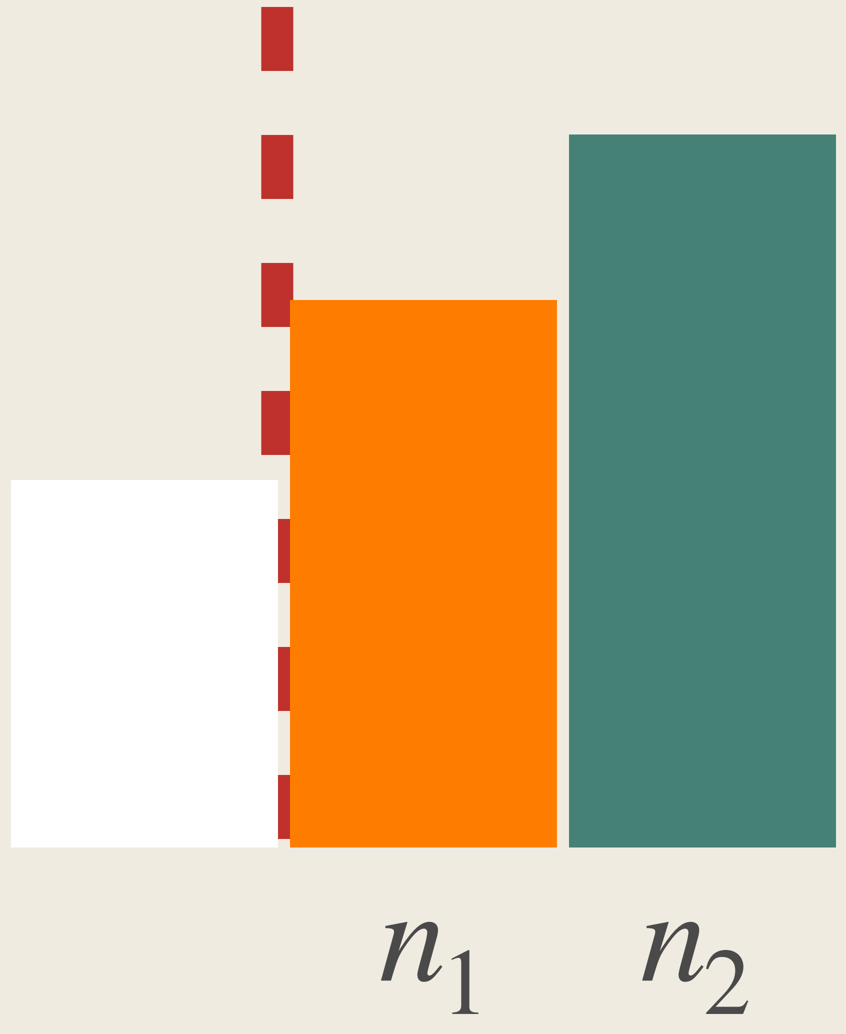


Entry & Exit at Lower Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n}+\frac{1}{2}\frac{\sigma(n_{i+1})^2g(n_{i+1})-2\sigma(n_i)^2g(n_i)+\sigma(n_{i-1})^2g(n_{i-1})}{(\Delta n)^2}=0$$

for $i = 1, \ldots, J-1$



Entry & Exit at L

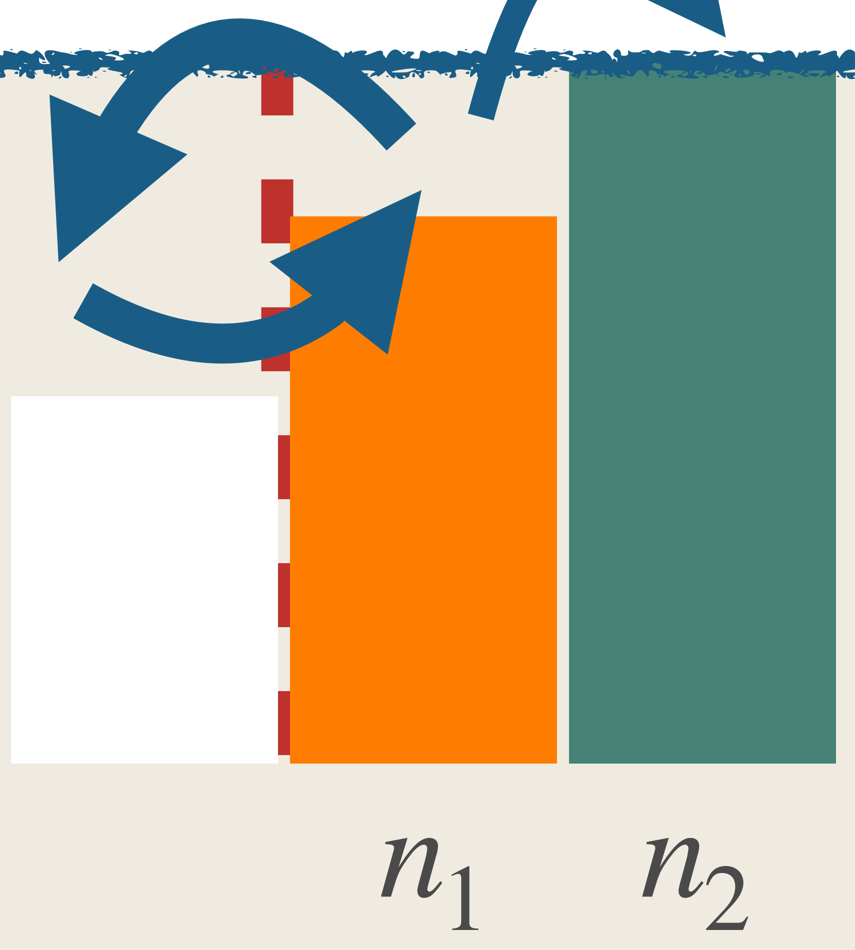
outflow from i due to variance
+ inflow from entry

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\dots, J -$

outflow from i due to drift
+ inflow from entry

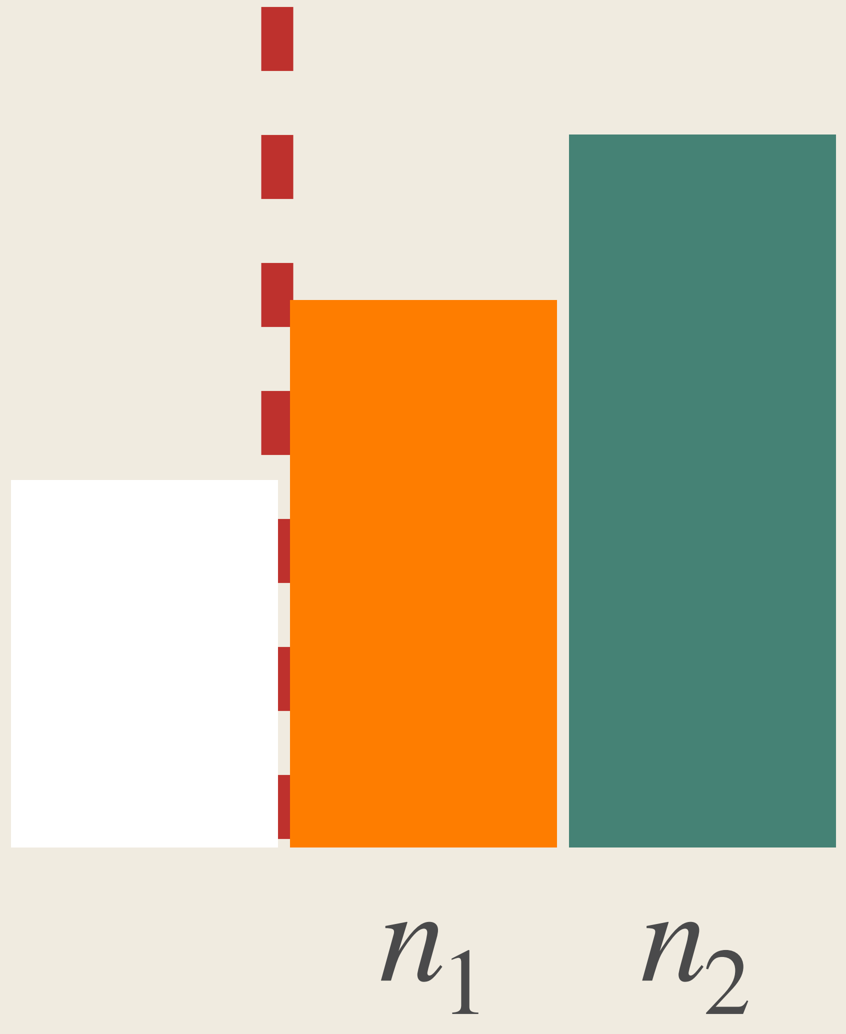


Entry & Exit at Lower Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n}+\frac{1}{2}\frac{\sigma(n_{i+1})^2g(n_{i+1})-2\sigma(n_i)^2g(n_i)+\sigma(n_{i-1})^2g(n_{i-1})}{(\Delta n)^2}=0$$

for $i = 1,\ldots,J-1$

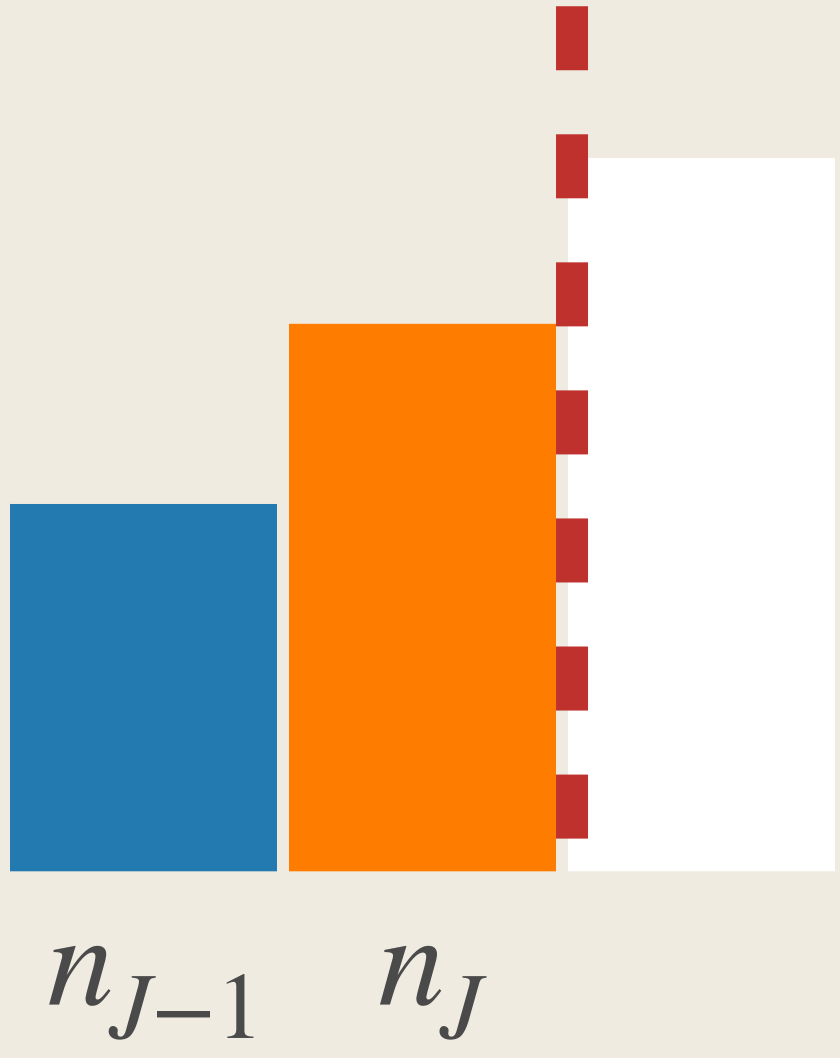


Reflection at Upper Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$



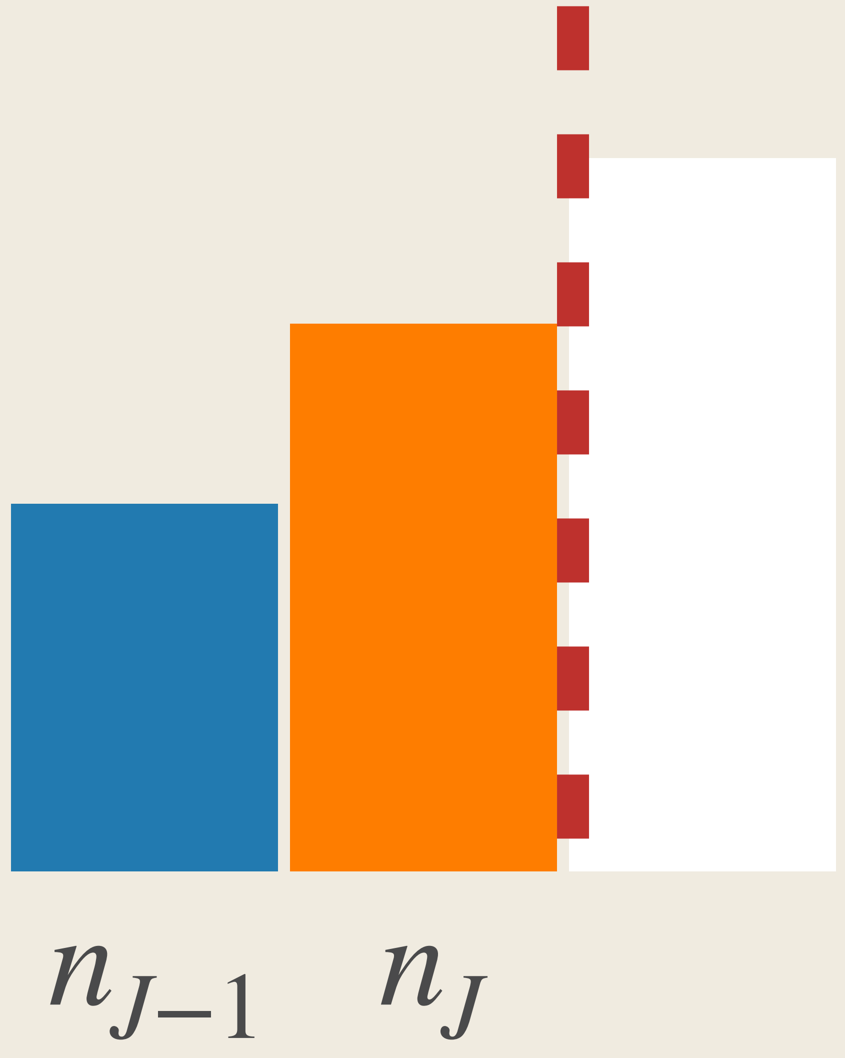
Reflection at Upper Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

Inflow from $i+1$ due to drift

Inflow from $i+1$ due to variance

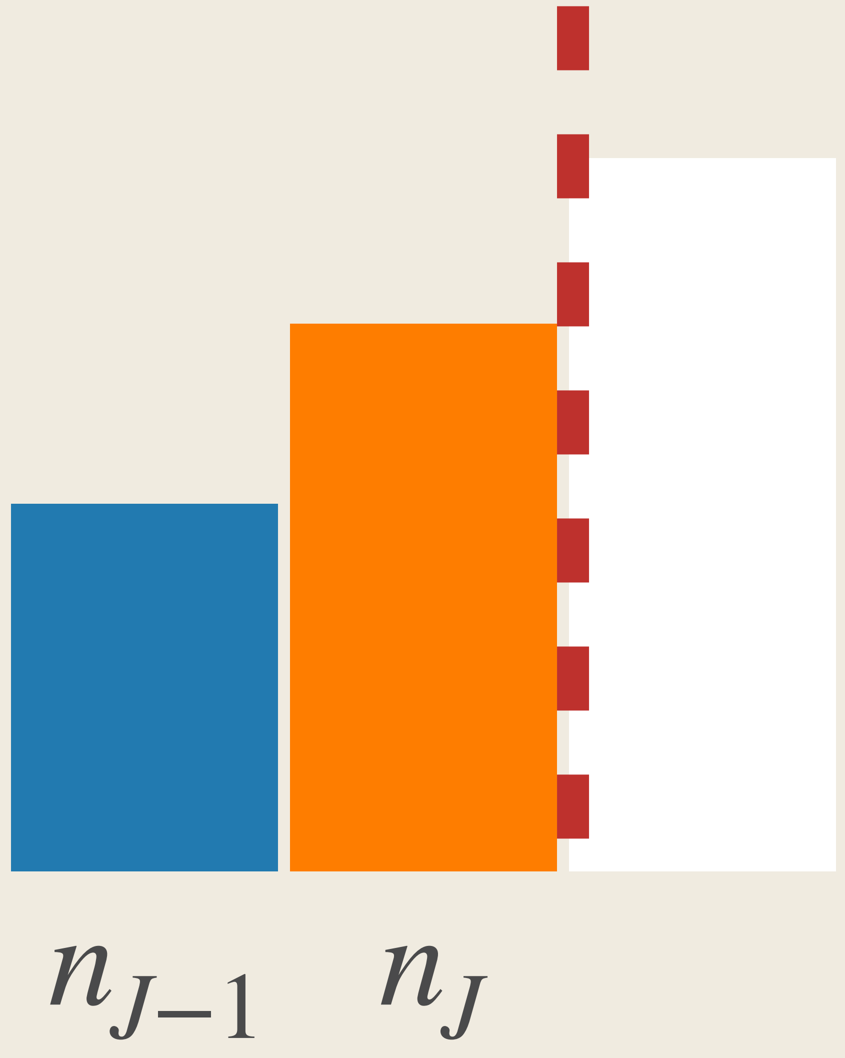


Reflection at Upper Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$



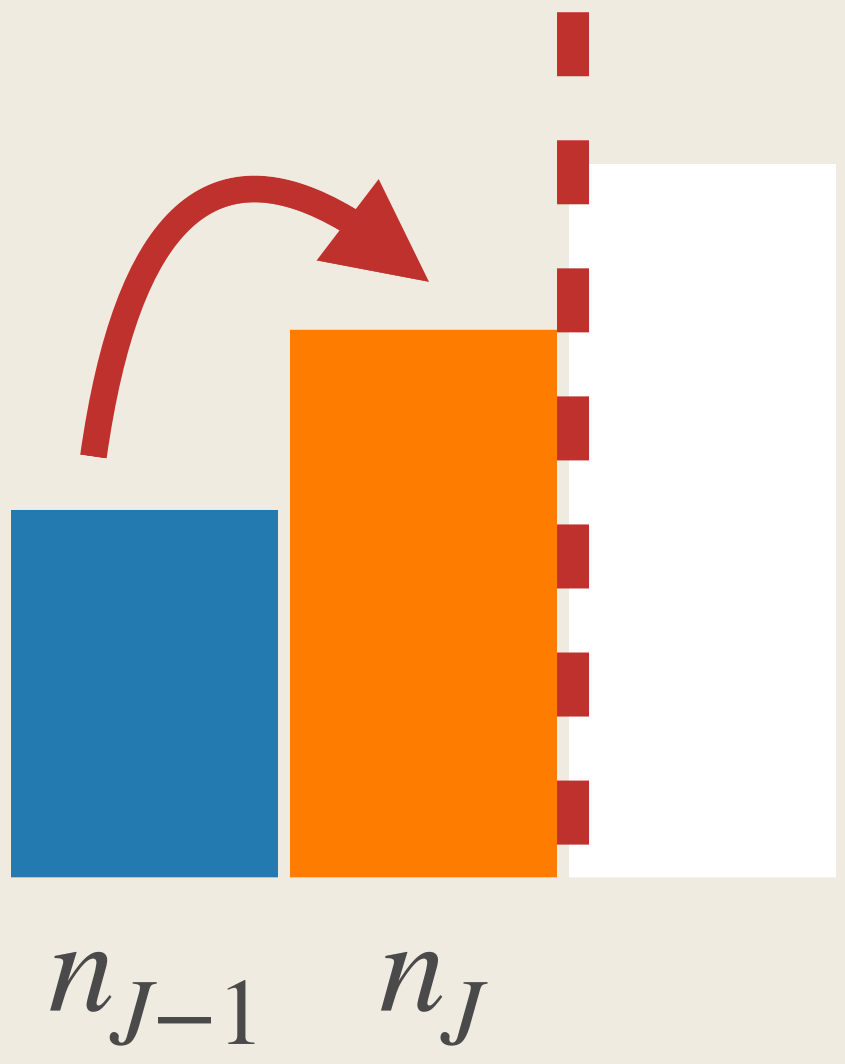
Reflection at Up

Inflow from $i-1$ due to variance

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \dots, J-1$

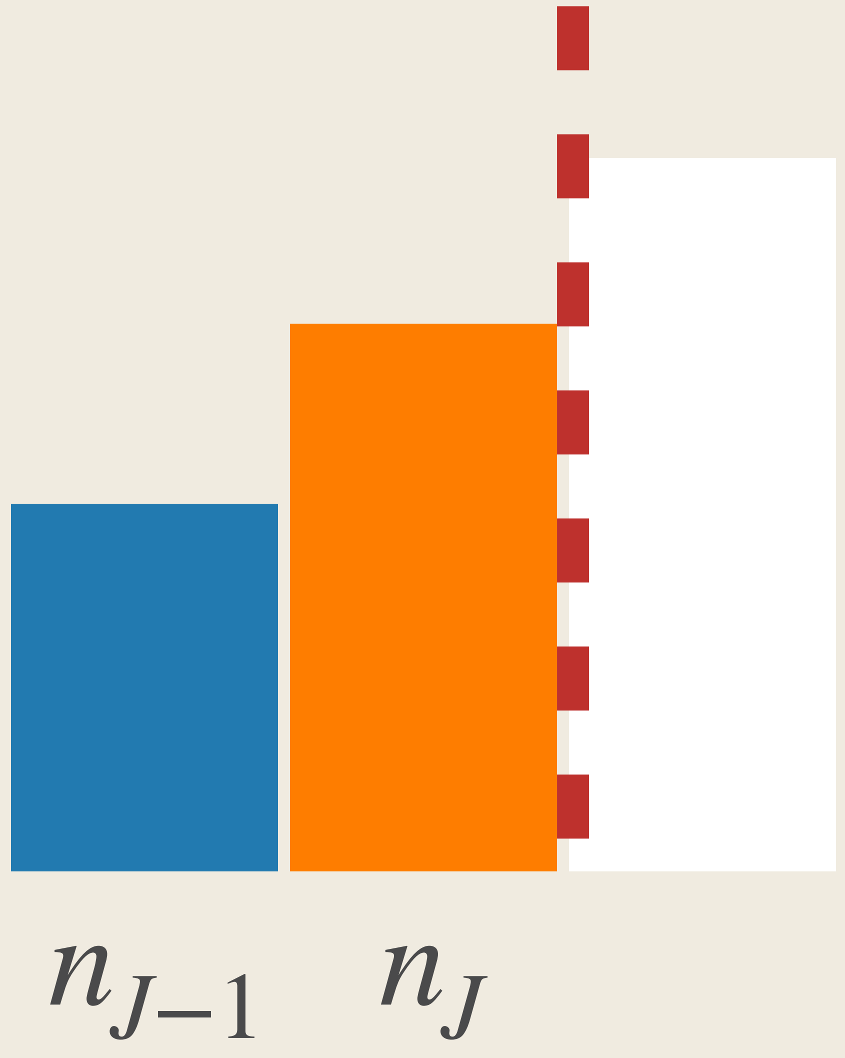


Reflection at Upper Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1,\ldots,J-1$



Reflection at Up

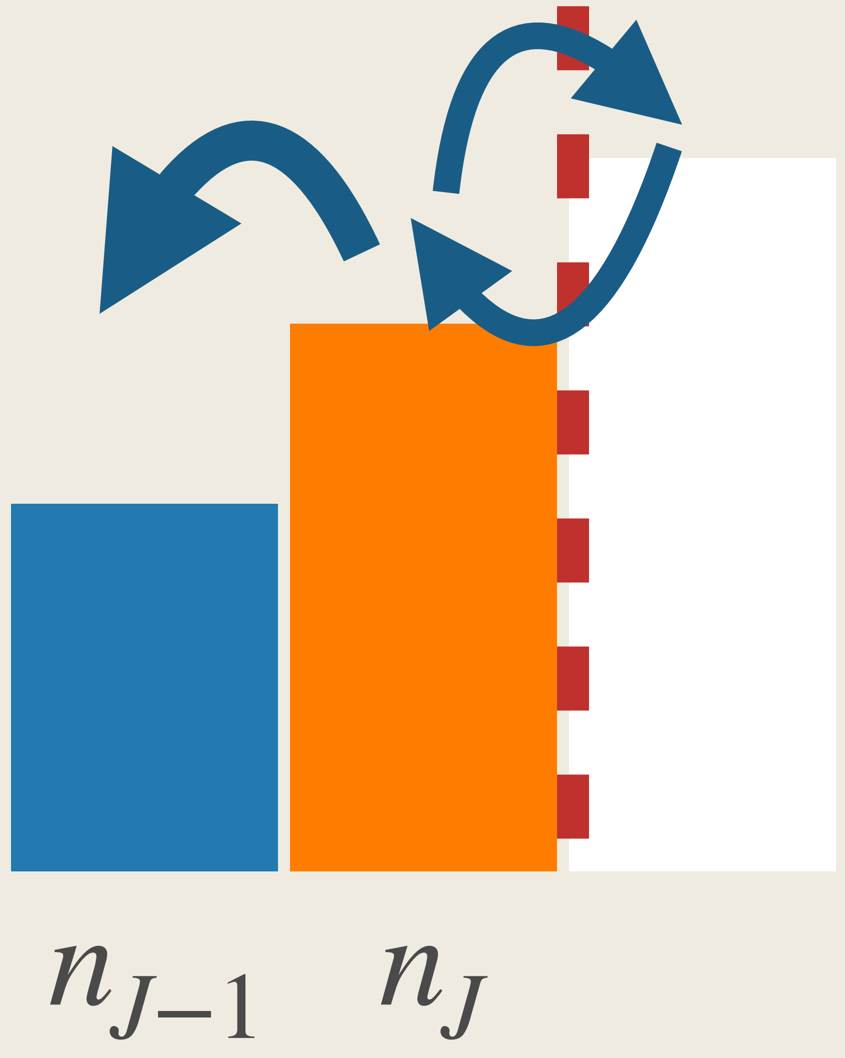
- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

outflow from i due to variance + reflection

$$\frac{-\mu(n_{i+1})g(n_{i+1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J -$

outflow from i due to drift

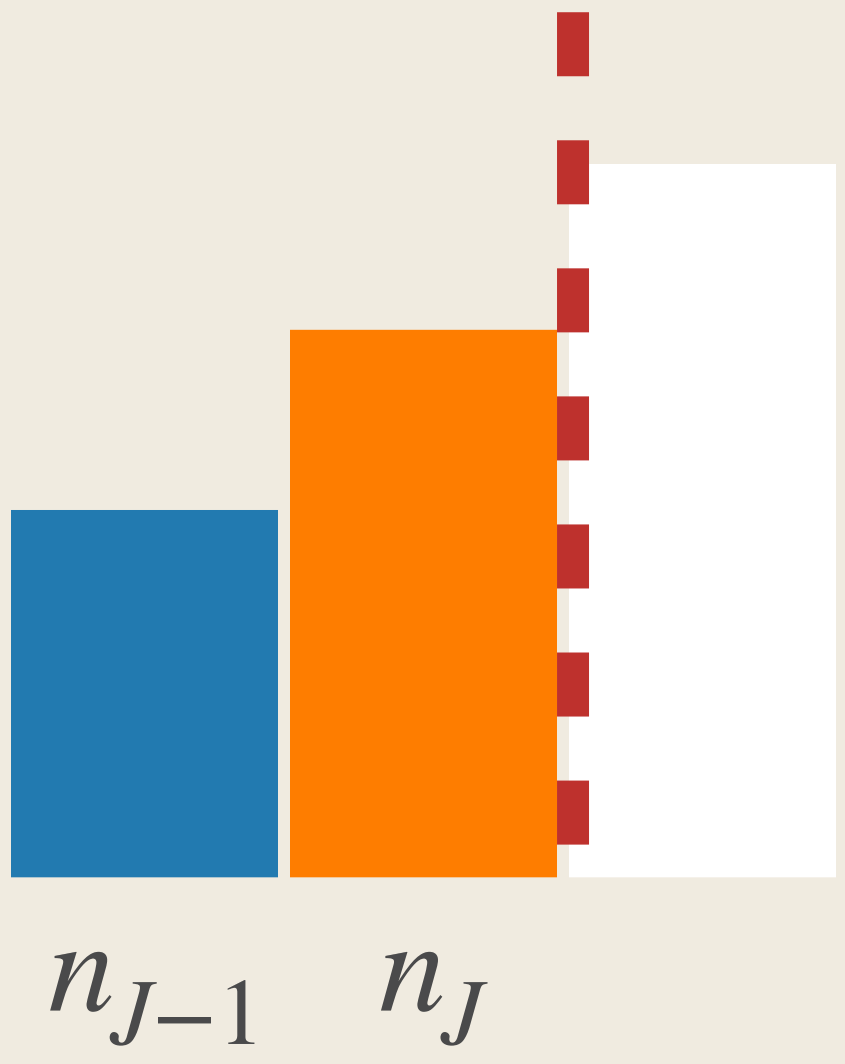


Reflection at Upper Boundary

- Suppose $\mu(n_i) < 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_{i+1})g(n_{i+1})+\mu(n_i)g(n_i)}{\Delta n}+\frac{1}{2}\frac{\sigma(n_{i+1})^2g(n_{i+1})-2\sigma(n_i)^2g(n_i)+\sigma(n_{i-1})^2g(n_{i-1})}{(\Delta n)^2}=0$$

for $i = 1,\ldots,J-1$



Linear System

- Realize that discretized KFE is a linear system of $\boldsymbol{g} \equiv [g(n_i)]_i$
- Since g is a density,

$$\sum_{j=1}^{J} g(n_j)\Delta n = 1$$

  which is also linear in $\boldsymbol{g}$

- Letting $\mu_i \equiv \mu(n_i)$ and $\sigma_i \equiv \sigma(n_i)$, the system can simply written in a matrix form

Linear System when $\mu(n) < 0$

$$\boldsymbol{A}^T \boldsymbol{g} = \boldsymbol{0} \qquad \textbf{(A)}$$

$$\Delta n \times \boldsymbol{1}' \boldsymbol{g} = 1 \qquad \textbf{(B)}$$

where $\boldsymbol{A} \equiv [A_{i,j}]_{i,j}$, and

$$A_{i,i} = \frac{\mu_j}{\Delta n} - \frac{\sigma_i^2}{(\Delta n)^2}, \quad A_{i,i-1} = -\frac{\mu_i}{\Delta n} + \frac{1}{2}\frac{\sigma_i^2}{(\Delta n)^2}, \quad A_{i,i+1} = \frac{1}{2}\frac{\sigma_i^2}{(\Delta n)^2}$$

All the other elements are 0.

- Intuitively, $A_{i,j}$ is the net transition rate from i to j. In fact, $\sum_j A_{i,j} = 0$

Matrix A when $\mu(n) < 0$

$$
\boldsymbol{A} \equiv \begin{bmatrix}
-\frac{1}{2(\Delta n)^2}(\sigma_1)^2 & \frac{1}{2(\Delta n)^2}(\sigma_1)^2 & 0 & 0 & \cdots & 0 & 0 \\
-\frac{1}{\Delta n}\mu_2 + \frac{1}{2(\Delta n)^2}(\sigma_2)^2 & \frac{1}{\Delta n}\mu_2 - \frac{1}{(\Delta n)^2}(\sigma_2)^2 & \frac{1}{2(\Delta n)^2}(\sigma_2)^2 & 0 & \cdots & 0 & 0 \\
0 & -\frac{1}{\Delta n}\mu_3 + \frac{1}{2(\Delta n)^2}(\sigma_3)^2 & \frac{1}{\Delta n}\mu_3 - \frac{1}{(\Delta n)^2}(\sigma_3)^2 & \frac{1}{2(\Delta n)^2}(\sigma_3)^2 & 0 & \cdots & 0 \\
0 & \cdots & \ddots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & & -\frac{1}{\Delta n}\mu_{J-1} - \frac{1}{(\Delta n)^2}(\sigma_{J-1})^2 & \frac{1}{\Delta n}\mu_{J-1} + \frac{1}{2(\Delta n)^2}(\sigma_{J-1})^2 \\
0 & 0 & 0 & \cdots & & -\frac{1}{\Delta n}\mu_J + \frac{1}{2(\Delta n)^2}(\sigma_J)^2 & \frac{1}{\Delta n}\mu_J - \frac{1}{2(\Delta n)^2}(\sigma_J)^2
\end{bmatrix}
$$

Matrix Inversion to solve g

- One of the rows in **(A)** is colinear (implied by **(B)**)
- Replace one of the rows in **(A)** with **(B)** to write

$$\tilde{A}g = \tilde{B} \quad \Rightarrow \quad g = \tilde{A}^{-1}\tilde{B}$$

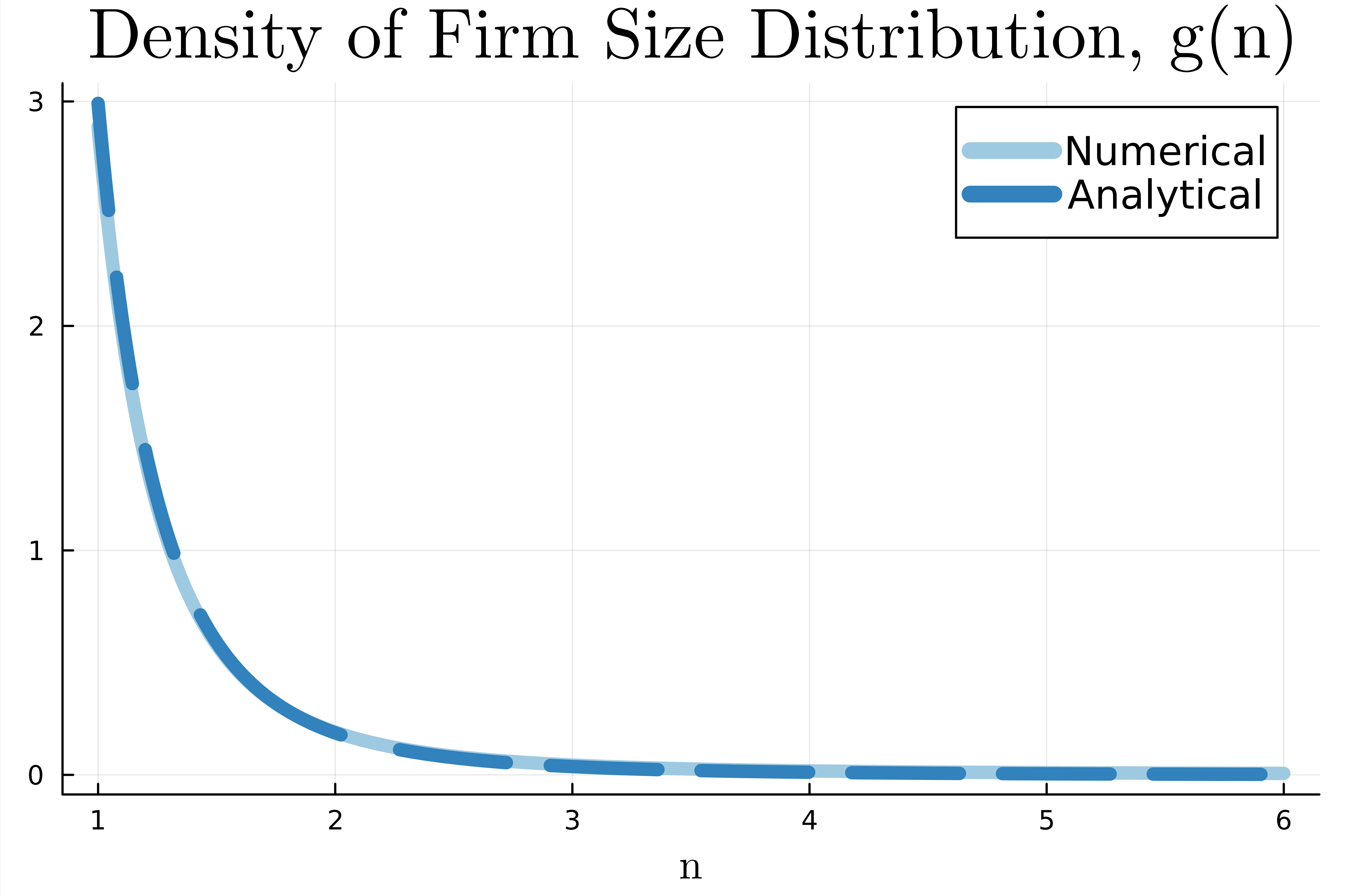
$\tilde{A}$: one row in A is replaced with $\Delta n \mathbf{1}'$, and the same row in $\tilde{B}$ is 1 and 0 elsewhere

- Inverting a big matrix like $\tilde{A}$ is typically expensive
- But, $\tilde{A}$ is sparse (many zero entries)
- Always work with a sparse matrix whenever the matrix has many zero entries
- Inverting a sparse matrix is cheap even when the matrix is big

Julia Code for Solving KFE

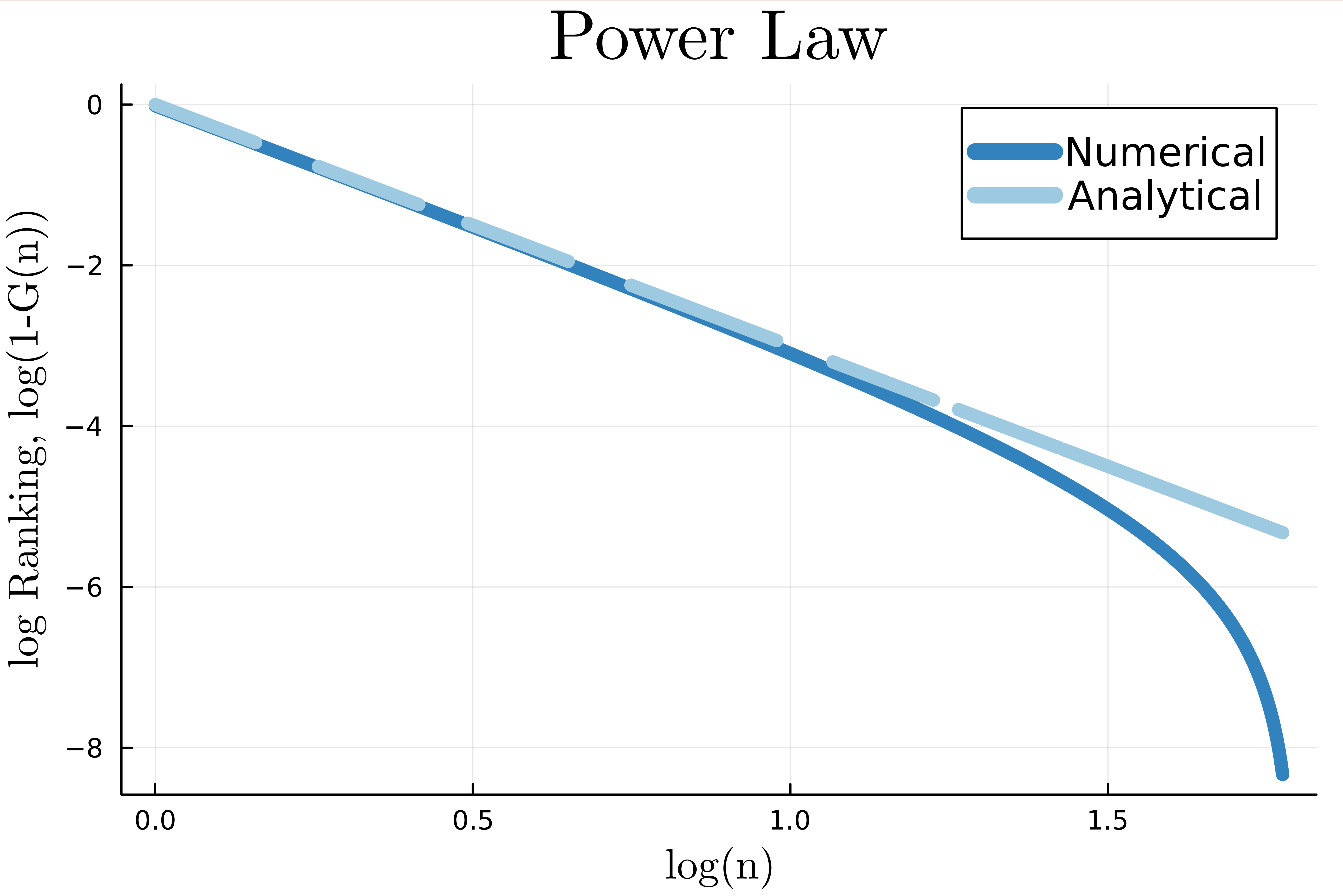
```julia
using SparseArrays
using Parameters
@with_kw mutable struct model
    J = 1000
    sig = 0.1
    mu = -0.01
    ng = range(1.0,6,length=J)
    dn = ng[2] - ng[1]
end
function populate_A(param)
    @unpack_model param
    A = spzeros(length(ng),length(ng))
    for (i,n) in enumerate(ng)
        A[i,i] += -(sig*n)^2/dn^2;
        A[i,min(i+1,J)] += 1/2*(sig*n)^2/dn^2;
        A[i,max(i-1,1)] += 1/2*(sig*n)^2/dn^2;
        if mu > 0
            A[i,i] += -mu*n/dn;
            A[i,min(i+1,J)] += mu*n/dn;
        else
            A[i,i] += mu*n/dn;
            A[i,max(i-1,1)] += -mu*n/dn;
        end
    end
    return A
end
function solve_stationary_distribution(param)
    @unpack_model param
    A = populate_A(param)
    B = zeros(length(ng));
    B[end] = 1;
    A[end,:] = ones(1,length(ng))*dn;
    g = A'\B;
    return g
end
param = model()
g = solve_stationary_distribution(param)
```

Solution



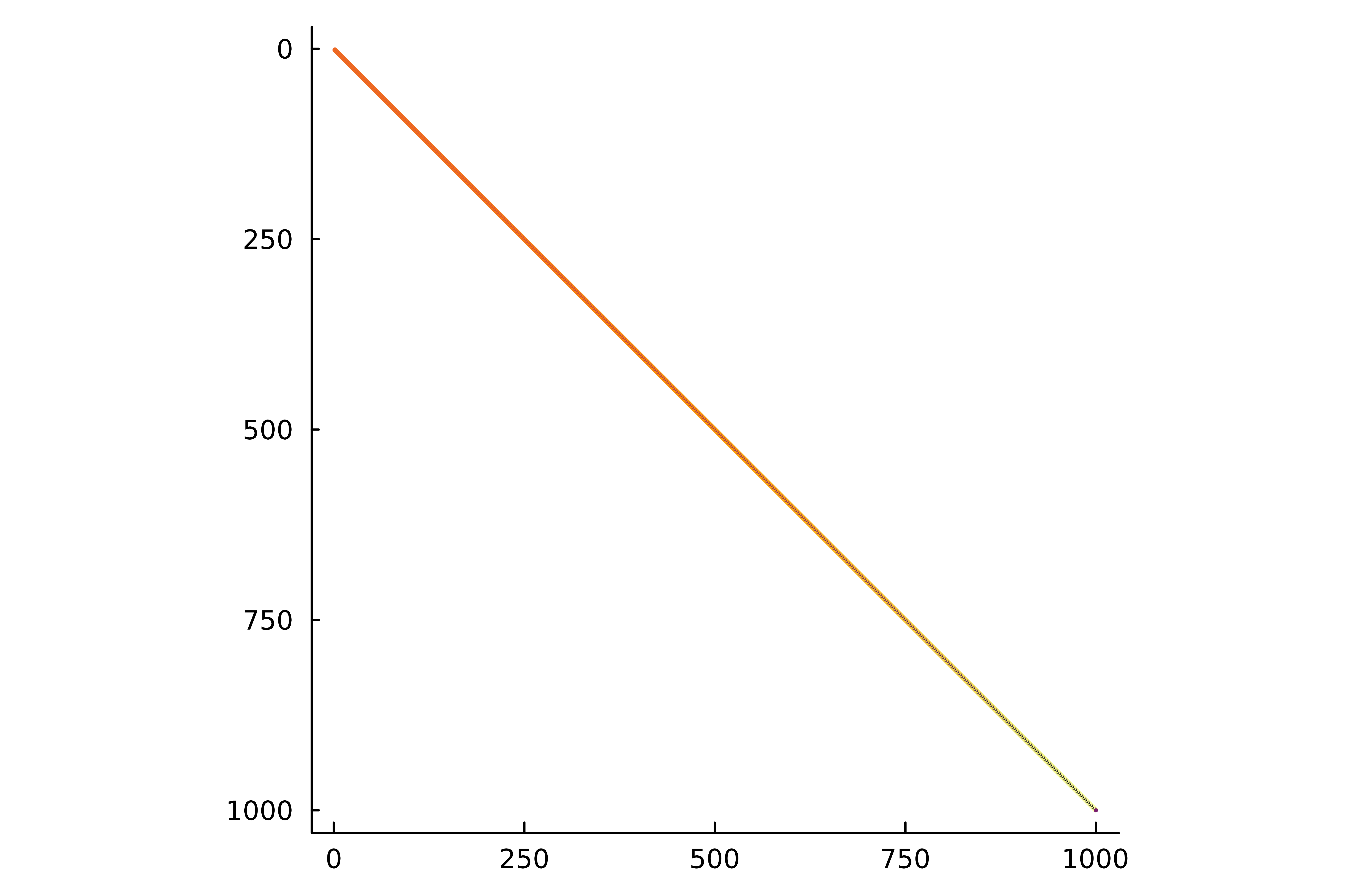
Density of Firm Size Distribution, g(n)
Numerical
Analytical
3
2
1
0
1
2
3
4
5
6
n

Power Law



- Bias in the upper tail due to truncation

spy(A)



- The advantage of continuous time with diffusion lies in the sparsity of A
- In discrete time, A is unlikely to be sparse in many applications

Numerically Computing Transition of Firm Size Distribution

Solving Transition Dynamics

- How do we numerically compute the transition path of $\{g_t(n)\}$ given $g_0(n)$?
- Recall the evolution of distribution is characterized by

$$\partial_t g_t(n) = -\partial_n[\mu(n)g_t(n)] + \frac{1}{2}\partial_{nn}^2\left[\sigma(n)^2 g_t(n)\right]$$

- We have to discretize time as well: $t \in [t_0, t_1, \ldots, t_N]$ and $\Delta t \equiv t_j - t_{j-1}$
- Approximate the time derivative using backward difference:

$$\partial_t g_t(X) \approx \frac{g_t(n) - g_{t-\Delta t}(n)}{\Delta t}$$

  - Can use forward difference but requires Δt to be small

Back to Markov Chain

- For any given $\boldsymbol{g}_{t-\Delta t} \equiv [g_{t-\Delta t}(n_i)]_i$, one can compute $\boldsymbol{g}_t$ by solving

$$\frac{\boldsymbol{g}_t - \boldsymbol{g}_{t-\Delta t}}{\Delta t} = \boldsymbol{A}^T \boldsymbol{g}_t$$

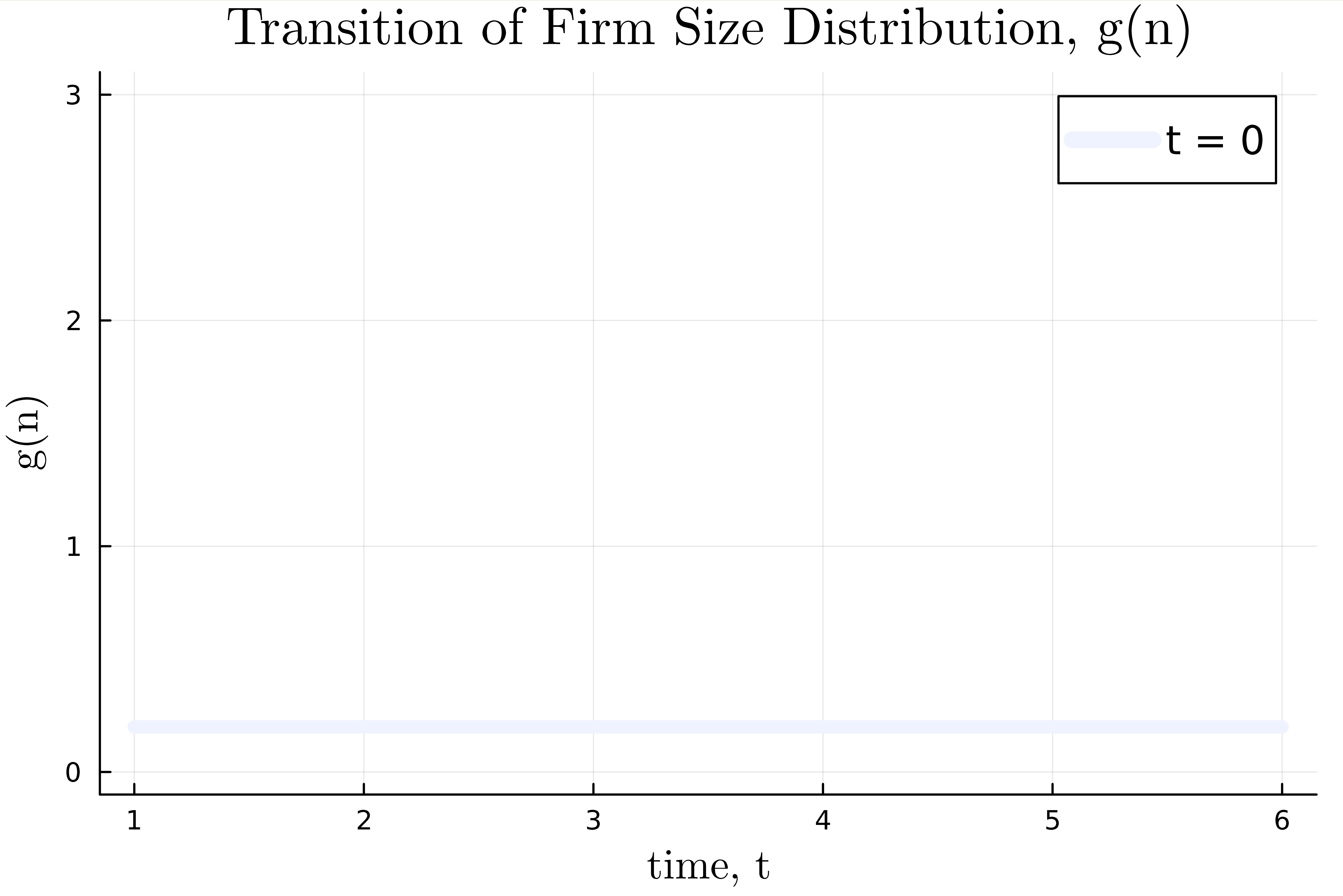
$$\Leftrightarrow \quad \boldsymbol{g}_t = \underbrace{\left[\boldsymbol{I} - \Delta t \times \boldsymbol{A}^T\right]^{-1}}_{\equiv \boldsymbol{P}} \boldsymbol{g}_{t-\Delta t}$$

- The matrix $\boldsymbol{P}$ corresponds to Markov Chain transition matrix in a time interval Δt

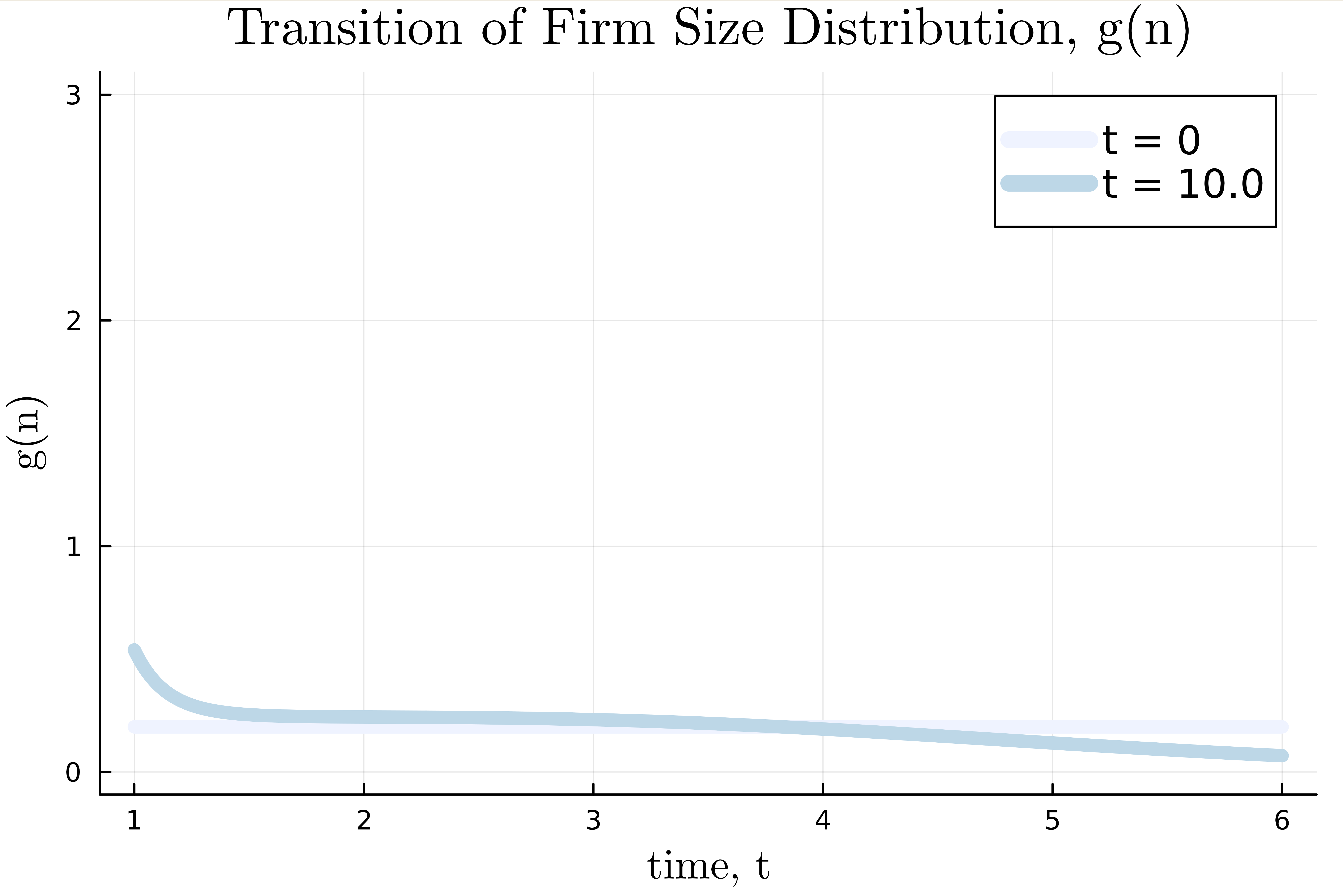
Julia Code for Transition

```
using LinearAlgebra
dt = 0.1;
T = 5000;
A = populate_A(param);
gpath = zeros(J,T);
gpath[:,1] = ones(J)./(J*dn);
for t = 2:T
    gpath[:,t] = (I - dt*A')\gpath[:,t-1]
end
```

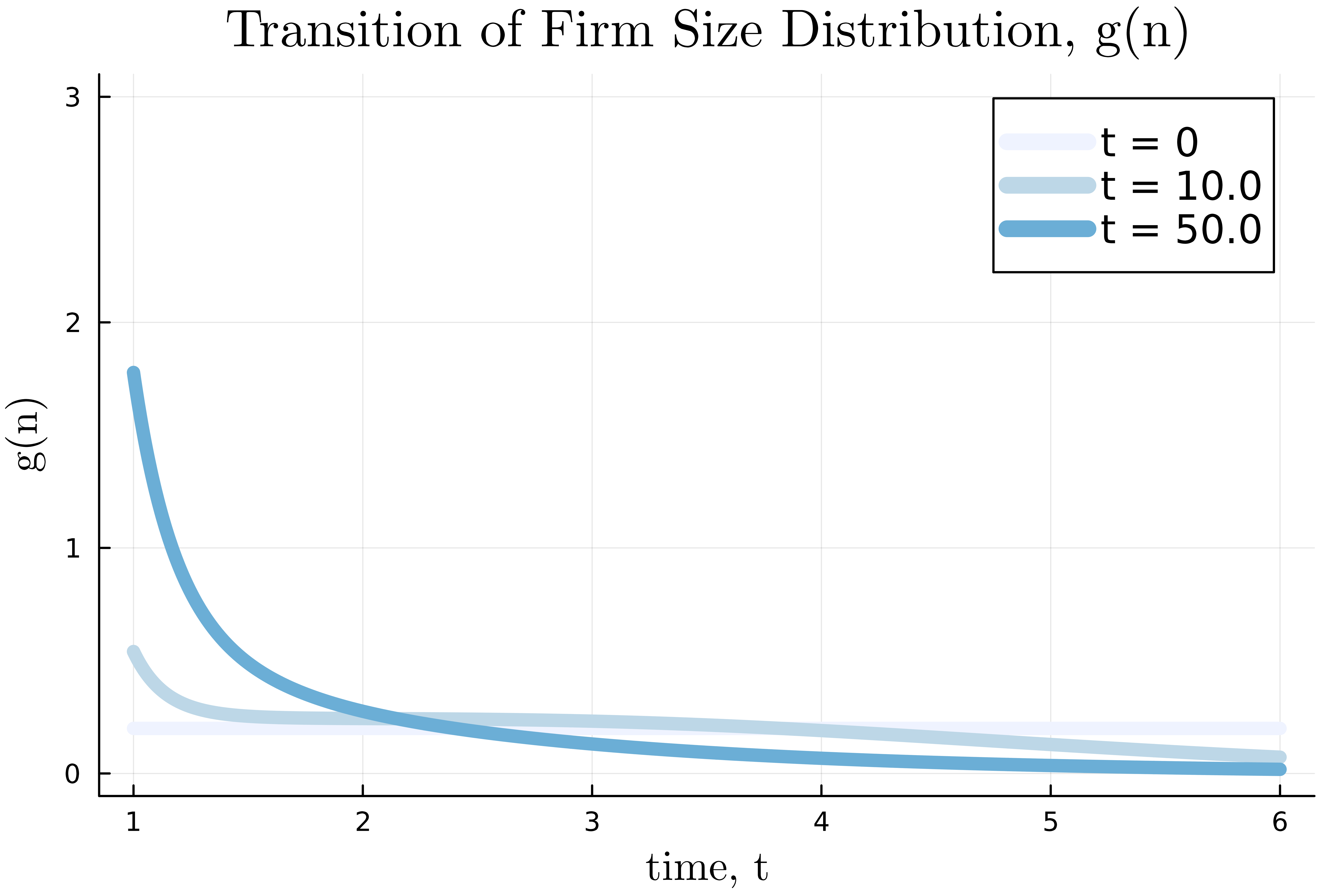
Transition Dynamics



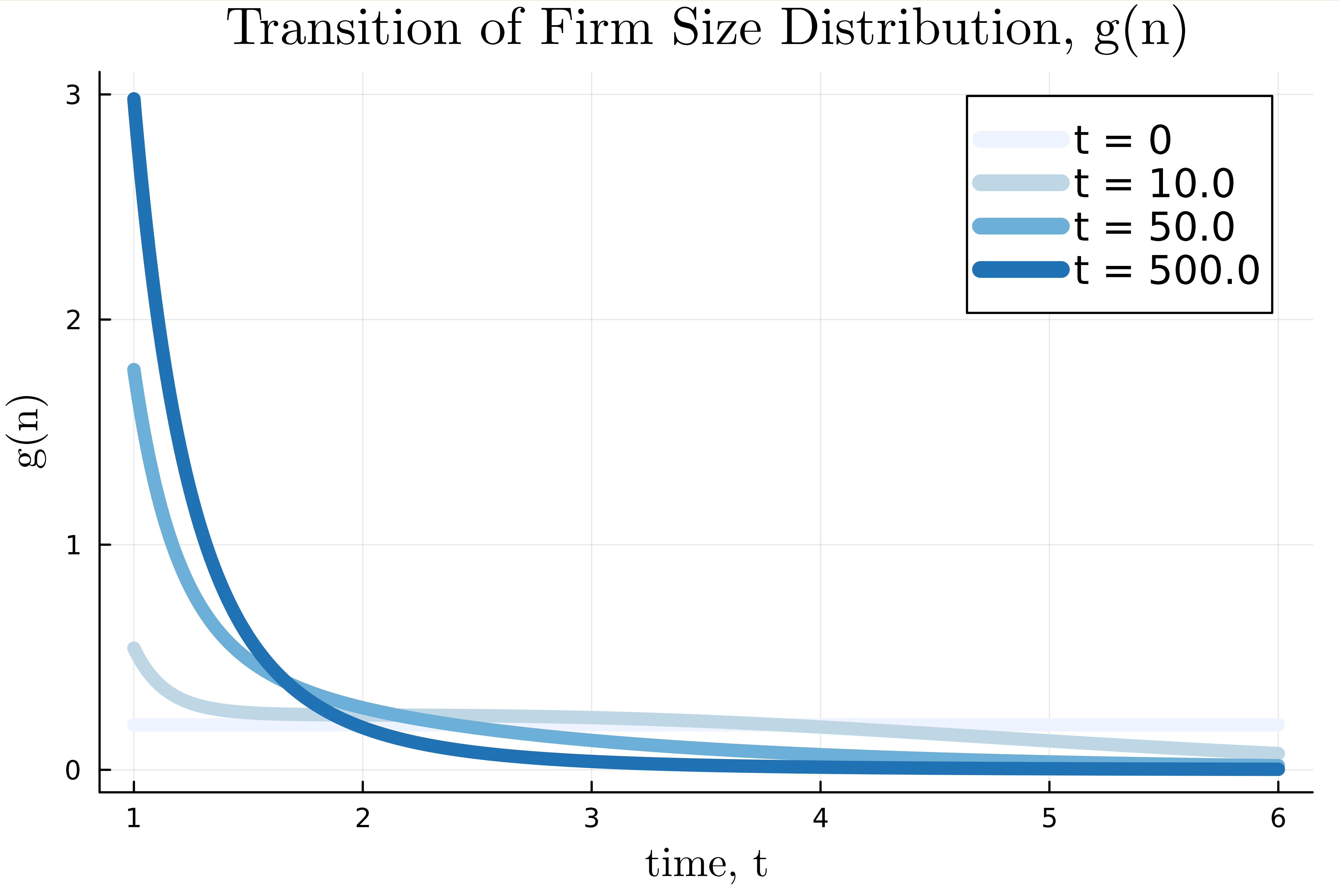
Transition Dynamics



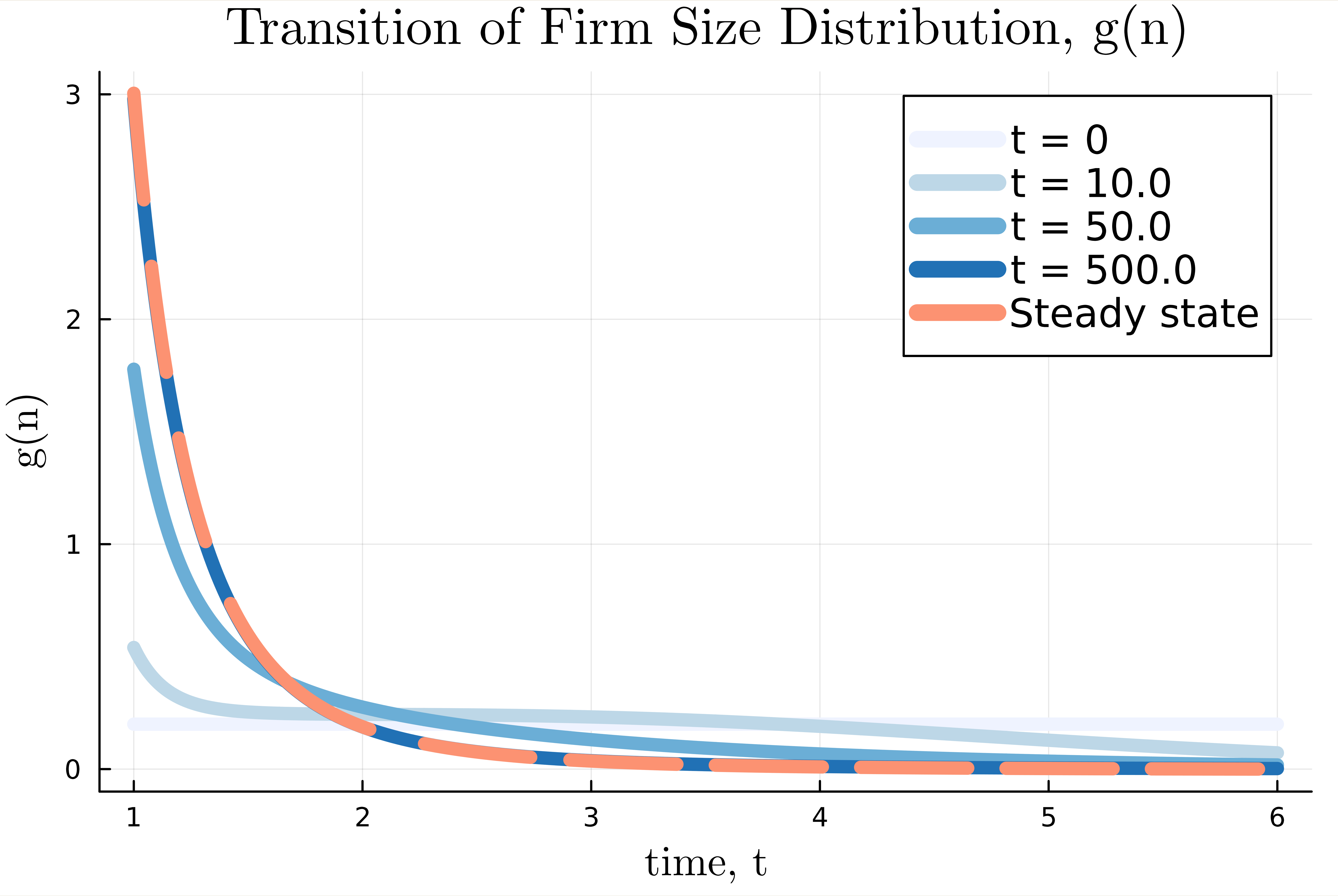
Transition Dynamics



Transition Dynamics



Transition Dynamics



Taking Stock

Taking Stock

- **Fact:** A handful of extremely large firms hire a large share of workers
  1. The firm size distribution is fat-tailed, Zipf's law
  2. Firm growth is roughly unrelated to firm size, Gibrat's law
- **Theory:** A mechanical model of firm growth as in Gabaix (1999)
  1. Gibrat's law + stabilizing force $\Rightarrow$ power law
  2. stabilizing force $\downarrow 0$ $\Rightarrow$ Zipf's law
- **Techniques:** We have covered important continuous-time tools
  1. Diffusion process, Kolmogorov forward equation (KFE)
  2. How to solve KFE on your computer

Appendix A: Non-Uniform Grid

Why Non-Uniform Grid?

- So far, we have considered equi-spaced grid:

$$\Delta n_j \equiv n_j - n_{j-1} = \Delta n$$

- In many applications, we would like to achieve the followings:
  1. We want the upper bound of the grid to be large enough
     - Walmart employs 2.3 million workers in 2021
  2. We want to accurately compute especially at the lower end of the grid
     - This is where exit decisions matter
  3. We do not want to take too many gridpoints
- We can achieve the above goal with non-uniform grid
  - Take many fine grids at lower ends and coarse grids at upper ends
  - log-spaced grid is a good example

Discretization with Non-Uniform Grid

- Suppose grids are non-uniform: $\boldsymbol{n} \equiv [n_1, n_2, \ldots, n_J]'$ with

$$\Delta n_{j,+} = n_{j+1} - n_j, \quad \Delta n_{j,-} = n_j - n_{j-1}$$

- Approximating first-derivative with non-uniform grid:

1. Forward difference approximation:

$$-\partial_n[\mu(n_i)g(n_i)] \approx -\frac{\mu(n_{i+1})g(n_{i+1}) - \mu(n_i)g(n_i)}{\Delta n_{j,+}}$$

2. Backward difference approximation:

$$-\partial_n[\mu(n_i)g(n_i)] \approx -\frac{\mu(n_i)g(n_i) - \mu(n_{i-1})g(n_{i-1})}{\Delta n_{j,-}}$$

- Approximating second-derivative with non-uniform grid:

$$\partial^2_{nn}\left[\sigma(n_i)^2 g(n_i)\right] \approx \frac{\Delta n_{j,-}\sigma(n_{i+1})^2 g(n_{i+1}) - (\Delta n_{j,+} + \Delta n_{j,-})\sigma(n_i)^2 g(n_i) + \Delta n_{j,+}\sigma(n_{i-1})^2 g(n_{i-1})}{\frac{1}{2}(\Delta n_{j,+} + \Delta n_{j,-})\Delta n_{j,+}\Delta n_{j,-}}$$

KFE in a Matrix Form when $\mu(n) < 0$

- Let $\boldsymbol{A} \equiv [A_{i,j}]_{i,j}$ with

$$A_{j,j-1} = -\frac{\mu_j}{\Delta n_{j,-}} + \frac{\Delta n_{j,+}\sigma_j^2}{(\Delta n_{j,+} + \Delta n_{j,-})\Delta n_{j,+}\Delta n_{j,-}}$$
$$A_{j,j} = \frac{\mu_j}{\Delta n_{j,-}} - \frac{(\Delta n_{j,+} + \Delta n_{j,-})\sigma_j^2}{(\Delta n_{j,+} + \Delta n_{j,-})\Delta n_{j,+}\Delta n_{j,-}}$$
$$A_{j,j+1} = \frac{\Delta n_{j,-}\sigma_j^2}{(\Delta n_{j,+} + \Delta n_{j,-})\Delta n_{j,+}\Delta n_{j,-}}$$

- If $\Delta n_{j,+} = \Delta n_{j,-} = \Delta n$, we go back to the uniform grid case

KFE with Non-Uniform Grid

- The density is $\boldsymbol{g} \equiv [g(n_j)]_j$. We work with the transformed density:

$$\tilde{\boldsymbol{g}} \equiv [\tilde{g}_j]_j, \quad \tilde{g}_j = g_j \tilde{\Delta} n_j$$

$$\tilde{\Delta} n_j = \begin{cases} \frac{1}{2}\Delta n_{j,+} & j = 1 \\ \frac{1}{2}(\Delta n_{j,+} + \Delta n_{j,-}) & j = 2,\ldots,J-1 \\ \frac{1}{2}\Delta n_{j,-} & j = J \end{cases}$$

- The KFE in a matrix form is

$$\boldsymbol{A}^T \tilde{g}_j = \boldsymbol{0}$$

Appendix B: Numerically Solving KFE when $\mu > 0$

Discretized KFE when $\mu(n_i) > 0$

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

D n $\mu(n_i) > 0$

Inflow from $i+1$ due to drift

- Suppose $\mu(n_i) > 0$ we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

Inflow from $i+1$ due to variance

Discretized KFE when $\mu(n_i) > 0$

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

Discretized KFE w

Inflow from $i-1$ due to variance

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

Discretized KFE when $\mu(n_i) > 0$

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

n_{i-1} n_i n_{i+1}

Discretized KFE when $\mu(n_i) > 0$

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

outflow from i due to drift

outflow from i due to variance

Discretized KFE when $\mu(n_i) > 0$

- Suppose $\mu(n_i) > 0$, we use backward difference and discretized KFE is

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1})}{\Delta n} + \frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1})}{(\Delta n)^2} = 0$$

for $i = 1, \ldots, J-1$

KFE at the Boundary when $\mu(n_i) > 0$

- At the boundary $i = 1$,

$$\frac{-\mu(n_i)g(n_i) \cancel{+\mu(n_{i-1})g(n_{i-1})}}{\Delta n}$$

$$+\frac{1}{2}\frac{\sigma(n_{i+1})^2 g(n_{i+1}) - 2\sigma(n_i)^2 g(n_i) \cancel{+\sigma(n_{i-1})^2 g(n_{i-1})} + \sigma(n_i)^2 g(n_i)}{(\Delta n)^2} = 0$$

  - Since $g(n_{i-1}) = 0$, inflow from $i - 1$ is absent
  - Since mass $\sigma(n_i)^2 g(n_i)\frac{1}{(\Delta n)^2}$ exits, the same mass enters at $n_i = \underline{n}$

- At $i = J$, assume reflecting barrier so that

$$\frac{-\mu(n_i)g(n_i) + \mu(n_{i-1})g(n_{i-1}) + \mu(n_i)g(n_i)}{\Delta n} + \frac{1}{2}\frac{-2\sigma(n_i)^2 g(n_i) + \sigma(n_{i-1})^2 g(n_{i-1}) + \sigma(n_i)^2 g(n_i)}{(\Delta n)^2} = 0$$