Knowledge Diffusion Through Networks

Treb Allen Kamran Bilir Zhang Chen Christopher Tonetti

April 2019

How is Knowledge Diffusion Shaped by Social Networks?

- Frictions that impede knowledge flows have the strong potential to affect efficiency and inequality— for example,
  - Slow diffusion of cost-reducing innovation (e.g. Griliches 1957)
  - Deviations from the law of one price (e.g. Jensen 2007)
  - Possible herding on inefficient choice (Banerjee 1992)

  $\longrightarrow$ Motivates policies that facilitate flows of knowledge

- Understanding the *structure* of knowledge frictions is key to implementing such 'information' policies well

- This paper: study a medical context. Doctors are positioned within a network and make prescription choices based on current knowledge
  - Provide evidence of doctor scripting behavior converging over time
  - Develop a social learning model that rationalizes this fact (no causal identification strategy)
  - Compare estimated vs. model-generated knowledge paths, inspect the mechanism, and perform counterfactuals

Main Findings (so far)

1. Two Descriptive Statistics:
   - Fact #1: Prescription shares are converging across doctors
   - Fact #2: Convergence rates are increasing in network centrality
   - These facts are also evident in generic substitution context
     - Not just about advertising
   - Both facts are robust to different measures of centrality and convergence
2. Imposing some additional structure, we develop an econometric approach to estimate doctor-specific knowledge paths
3. We develop a dynamic model of learning on a network
   - We estimate initial knowledge stocks and network structure
   - Model-generated learning matches estimated paths

Related Literature

1. Theory papers on dynamics of social learning
   - Banerjee (1992), Smith and Sorensen (2008), Acemoglu et al (2011), etc.
   - Our contribution:
     1. A quantitative paper
     2. Transition dynamics
     3. Social learning on a fairly general estimated network
2. Idea diffusion papers
   - Grossman-Helpman (1991), Luttmer (2007), Lucas (2009), Lucas-Moll (2013), Perla-Tonetti (2013), Buera-Oberfield (2017), etc.
   - Our contribution:
     1. Empirical focus with micro data and specific tangible technologies
     2. Incorporate geography and other elements of networks (not random search)
     3. Bayesian learning instead of $\max(z, z')$

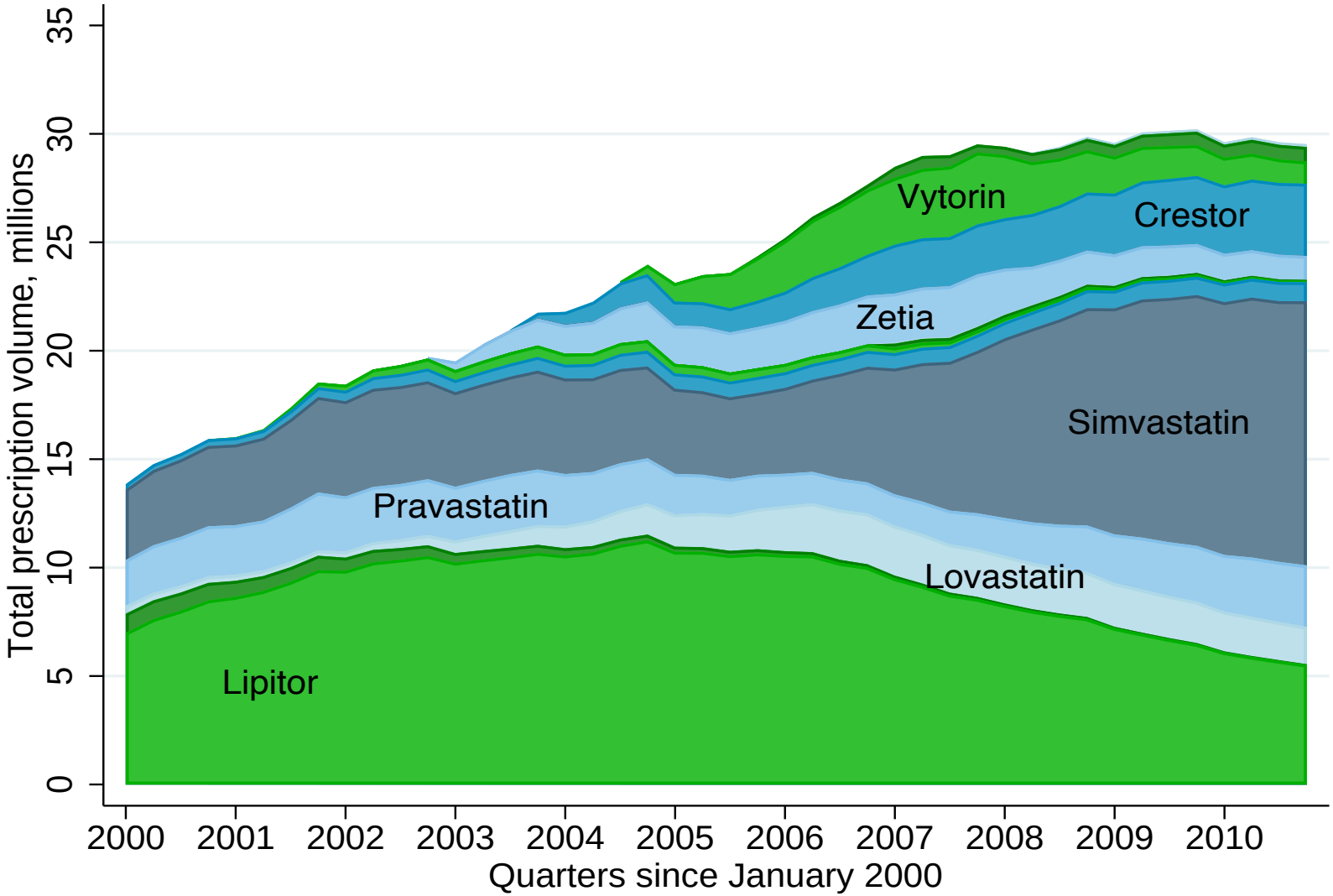
Related Literature

3. Learning in pharmaceutical markets
    - Erdem and Keane (1996), Ackerberg (2003), Crawford and Shum (2005), Arrow, Bilir and Sorensen (2017), Dickstein (2018)
        - Approach: dynamic discrete choice estimation with Bayesian learning
        - Estimate importance of 'signals' from network neighbors
    - Our contribution: prescription convergence depends on network (endogenous signal structure)
4. Papers measuring healthcare disparities: Dartmouth Atlas (Wennberg et al 1996, Munson et al 2013) and Cooper et al (2015)
    - Wide variation in healthcare quality, efficiency across regions (e.g. high vs. low generic prescription share)
    - Explained by unobserved differences in patient types or in knowledge?
    - Our contribution: our network-learning mechanism can explain observed convergence—*changes* in the extent of medical care disparities
        - Central doctors learn faster; peripheral doctors slower
        - This mechanism can rationalize treatment disparities

Data: Prescriptions from IMS Health (IQVIA, Xponent)

- We observe prescriptions by doctor, drug, and month
- Our data cover January 2000 through December 2010
  - This time period covers sequence of 12 drug innovations
  - Key entrants: three generics, three new 'molecular entities'
  - Today, within-spell analysis. In future, introduction of new drugs.
- Includes all U.S. doctors with at least 10 cholesterol drug prescriptions in *both* 2000 and 2010
  - 131,323 doctors $\times$ 132 months $\times$ up to 18 drugs
  - We observe location (five-digit zipcode) for all doctors
  - For $\sim$ 47K, medical school, cohort, specialty also observed
- Extraordinarily precise data on repeated technology adoption and prescription decisions by the universe of relevant individuals
- Unfortunately, no data on patients or insurance plans

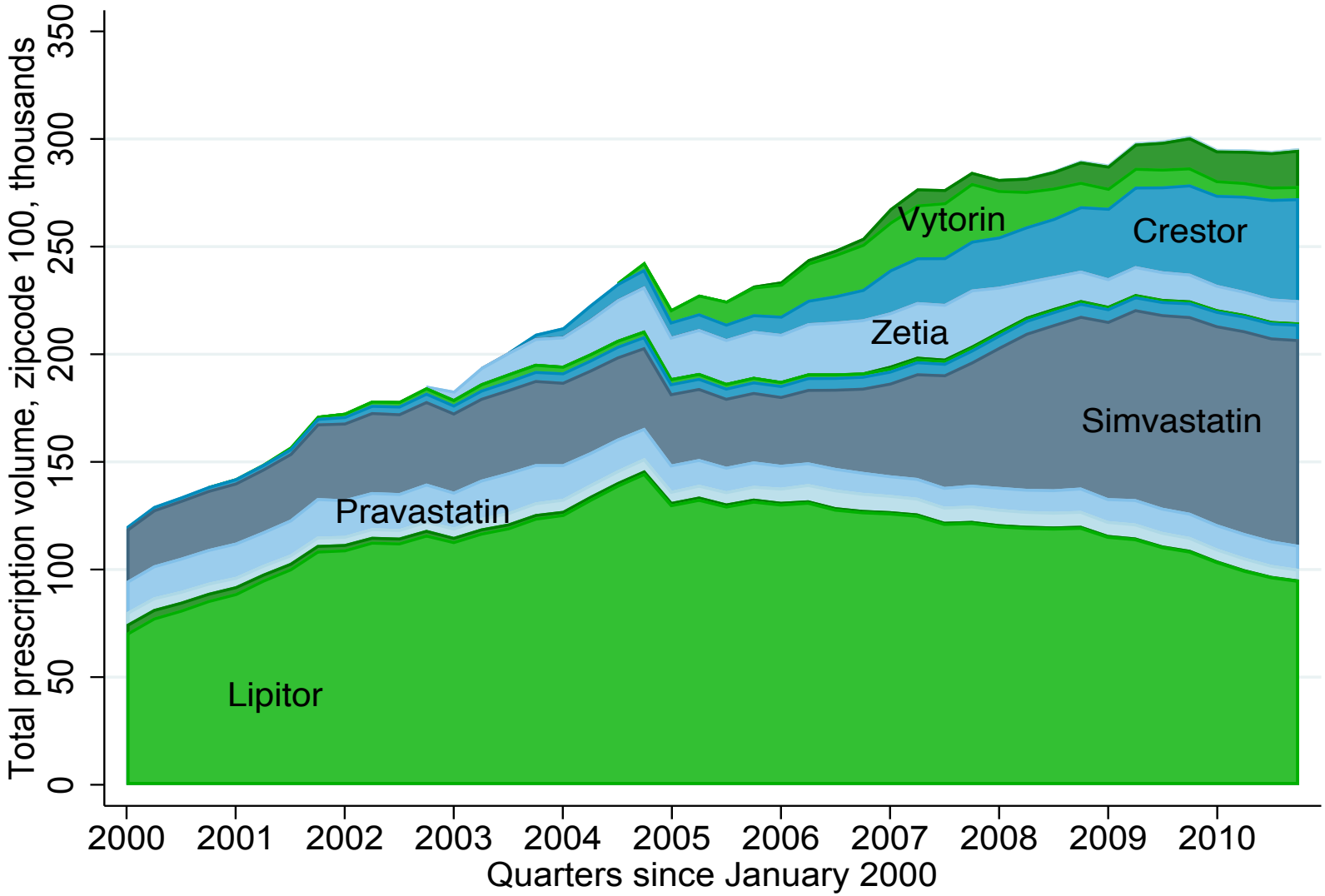
Aggregate Evolution in Prescribing, Jan. 2000–Dec. 2010



$\longrightarrow$ Suggests changes in perceived drug 'qualities'

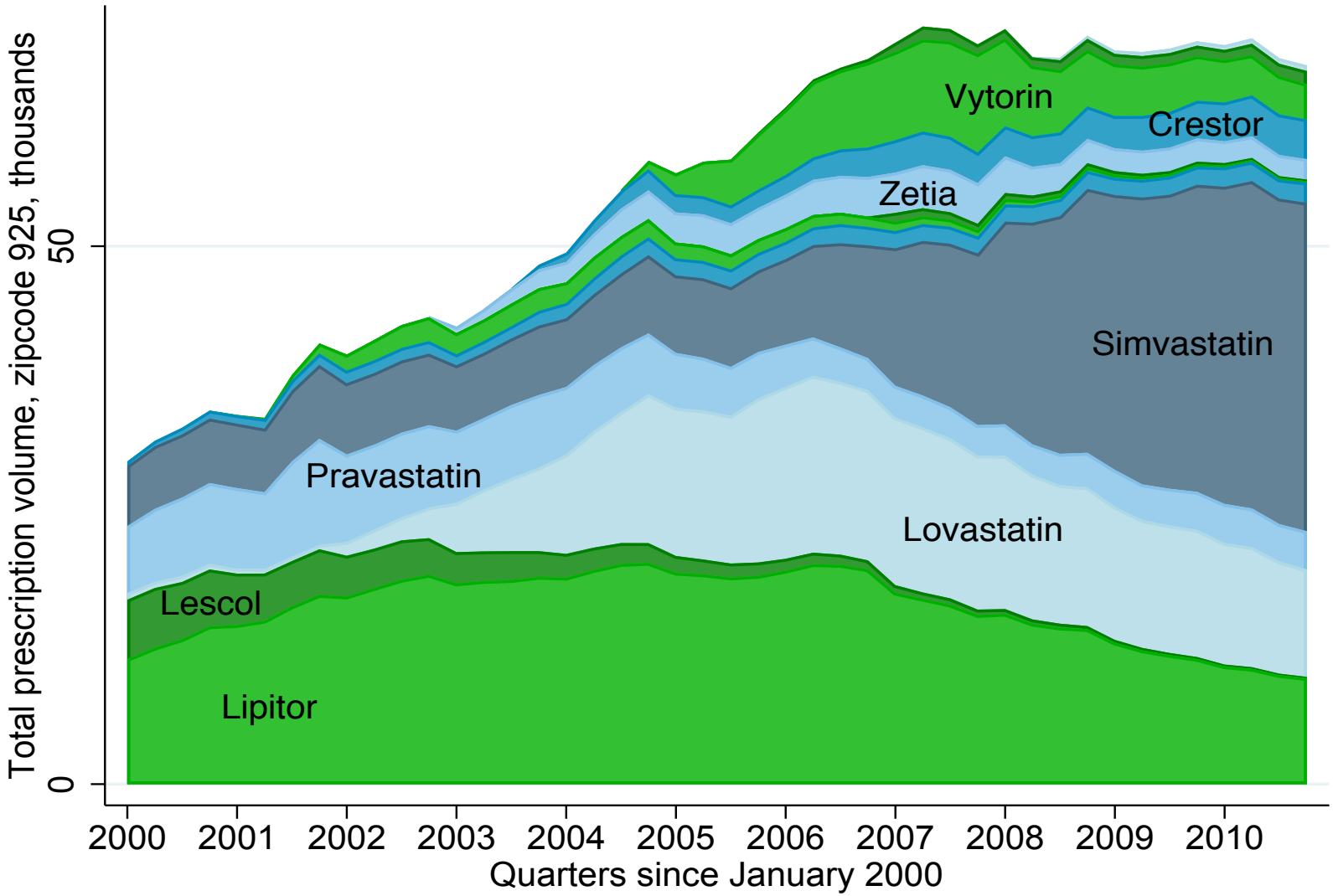
Evolution in Prescribing: NYC

$\longrightarrow$ Variation and evolution is different for areas like New York, NY...



Evolution in Prescribing: Hemet, CA

$\longrightarrow$ ...Relative to remote areas like Hemet, CA



Within-molecule Substitution to Generic also varies

- Gradual diffusion is observed even for generics
- Generic substitution is only partial six months after generic release, but is essentially complete by December 2010 (final month)

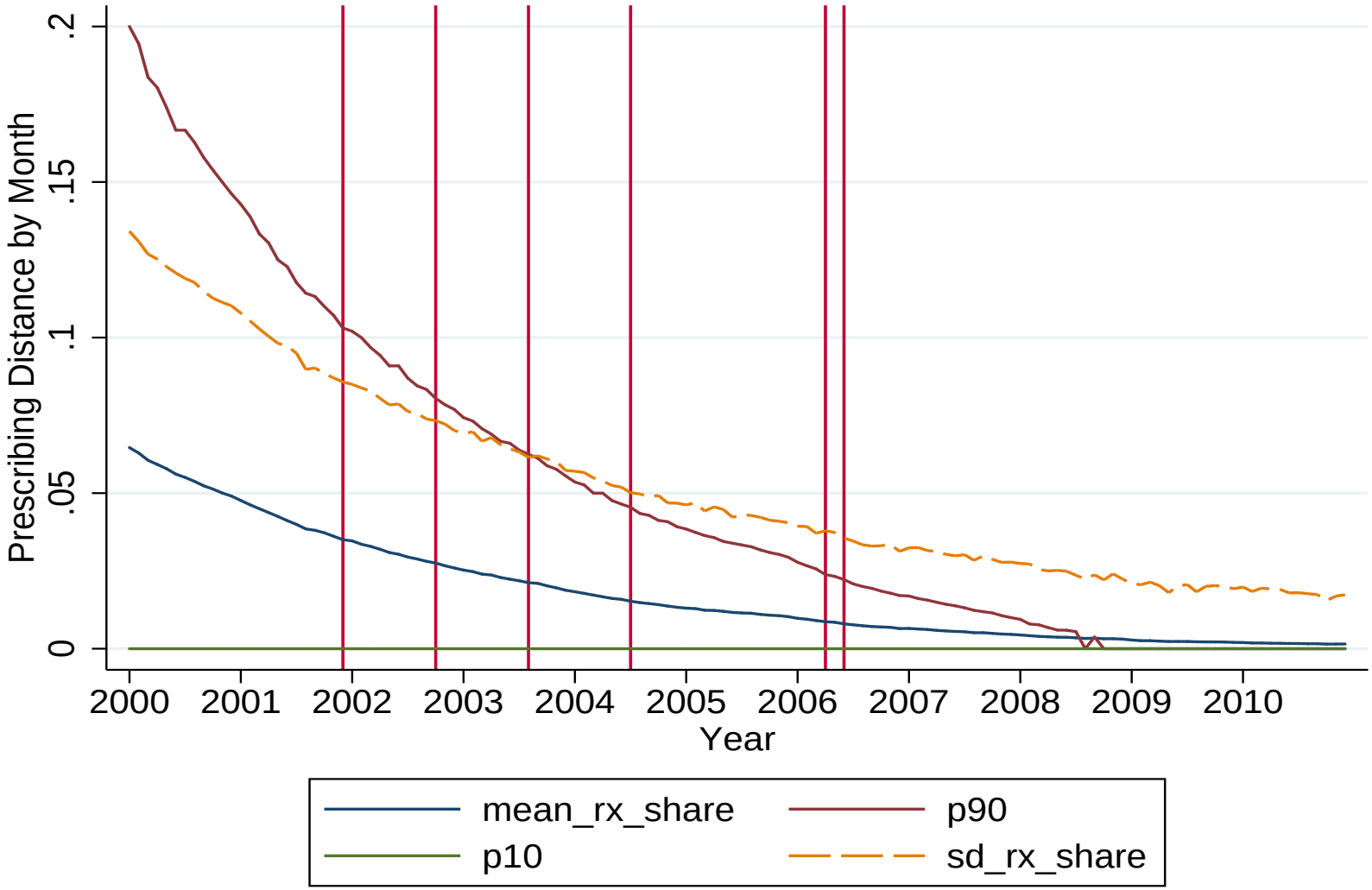
	Generic Share in Prescriptions		
	Lovastatin	Pravastatin	Simvastatin
After six months			
Mean	0.8280	0.8197	0.8616
St Dev	0.3382	0.2793	0.2079
5^{th} Percentile	0	0	0.448
95^{th} Percentile	1	1	1
December 2010			
Mean	0.9995	0.9930	0.9970
St Dev	0.0188	0.0588	0.0276
5^{th} Percentile	1	1	0.995
95^{th} Percentile	1	1	1

Different Types of Convergence

- Will summarize doctor scripting patterns by drug shares
- Could have all doctors moving steadily towards same final-period drug shares
  - Should see decreasing distance between average share and final share over time
- Could have doctors drug-shares becoming more similar to each other each period, but big fluctuations in those drug shares across periods
  - Should see decreasing variation in drug shares across doctors over time.
- Reminiscent of β and σ convergence in growth regression literature
- Will now show pictures that show within-spell both types of convergence are present

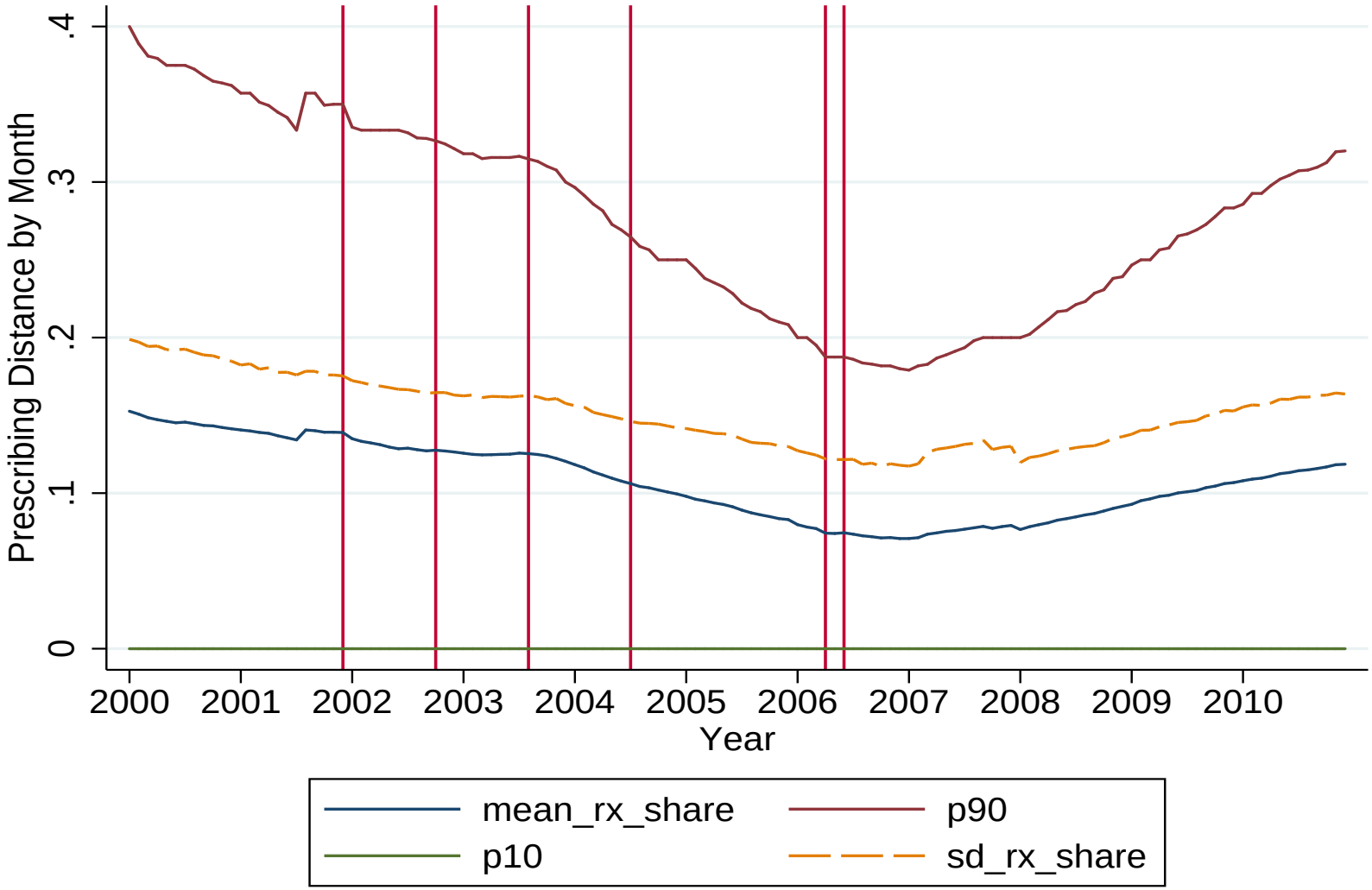
Fact #1: Prescriptions are converging across doctors

Lescol



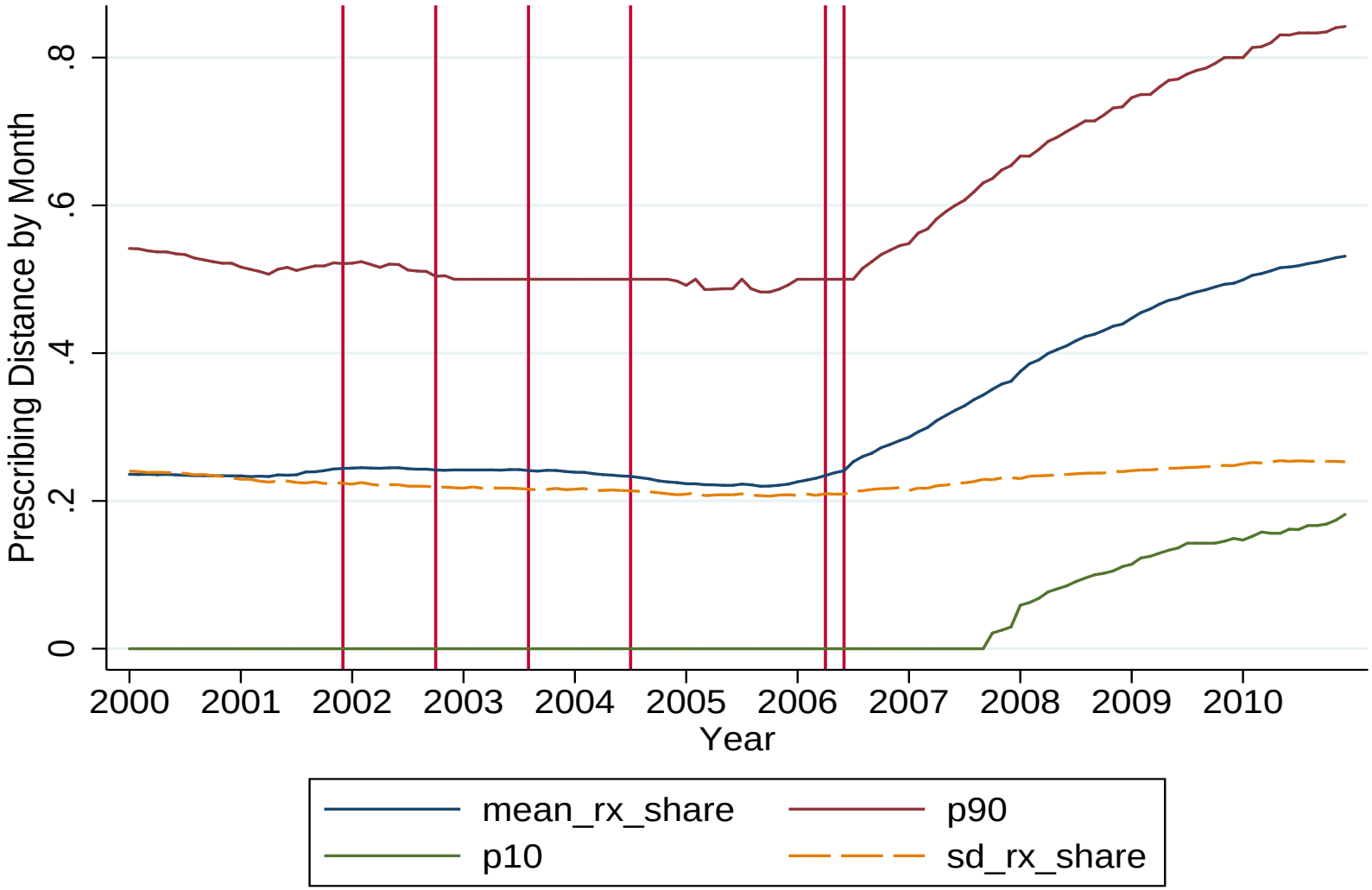
Fact #1: Prescriptions are converging across doctors

Pravastatin



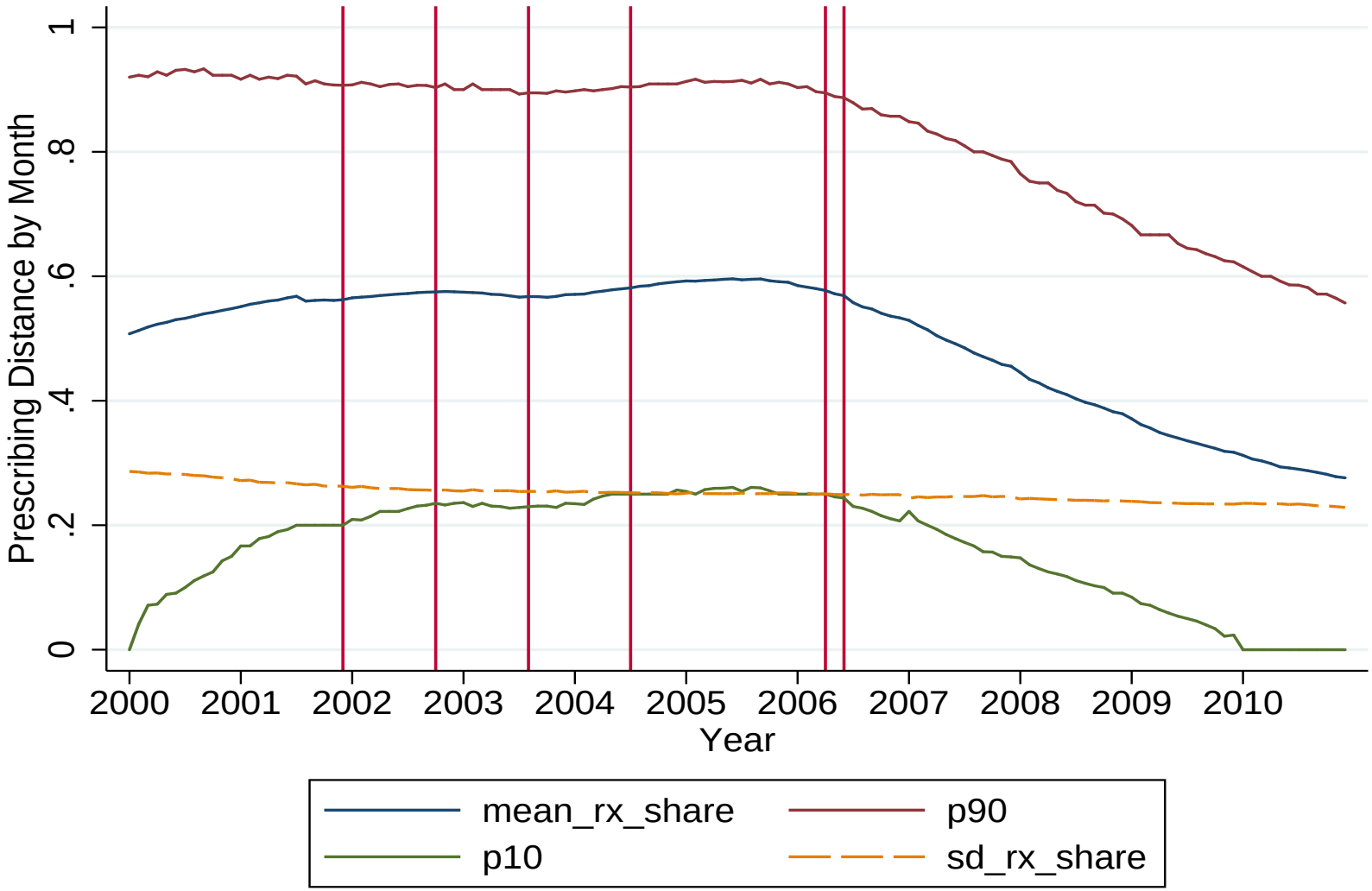
Fact #1: Prescriptions are converging across doctors

Simvastatin



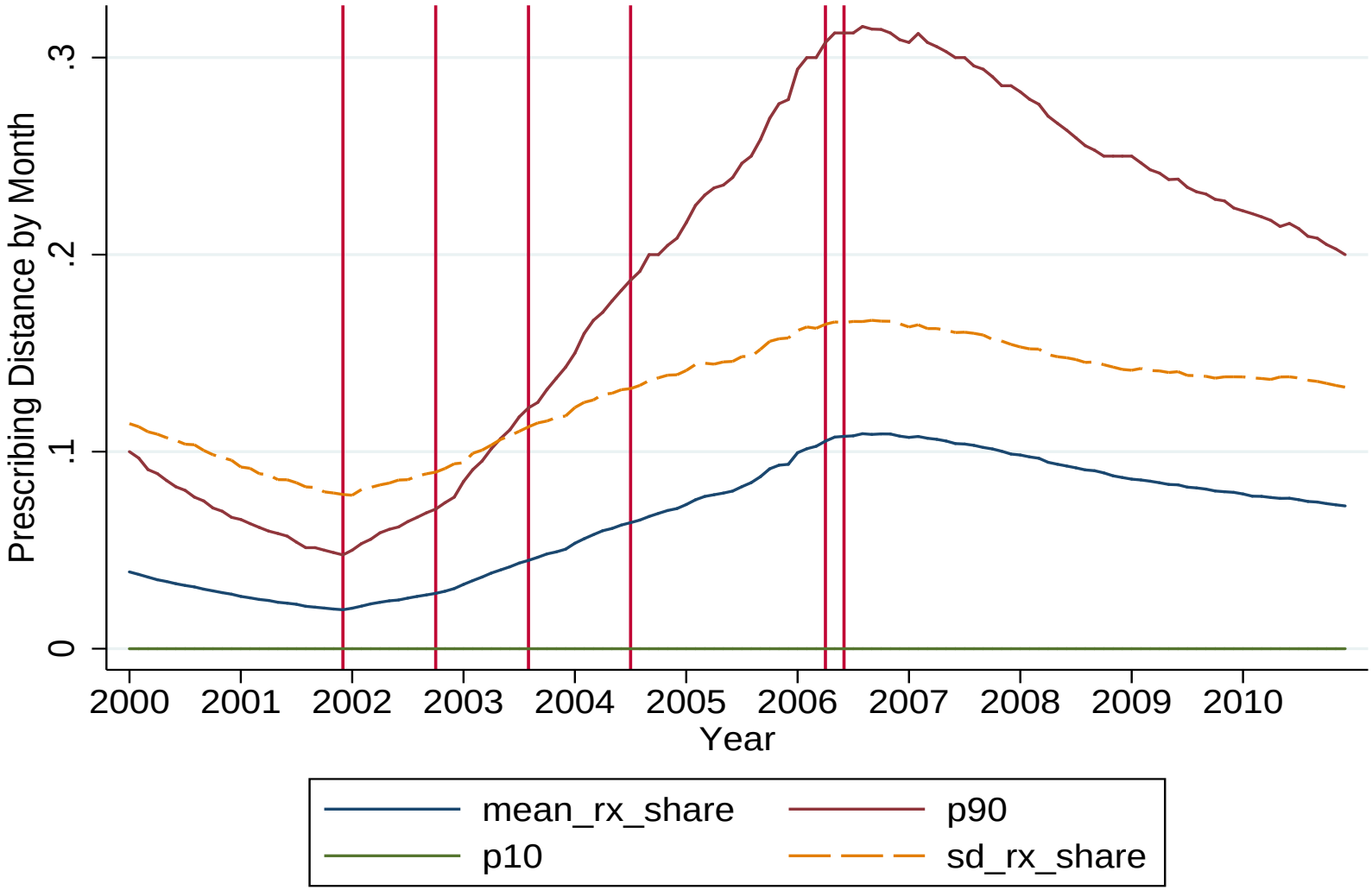
Fact #1: Prescriptions are converging across doctors

Lipitor



Fact #1: Prescriptions are converging across doctors

Lovastatin



Fact #1: Prescriptions are converging across doctors

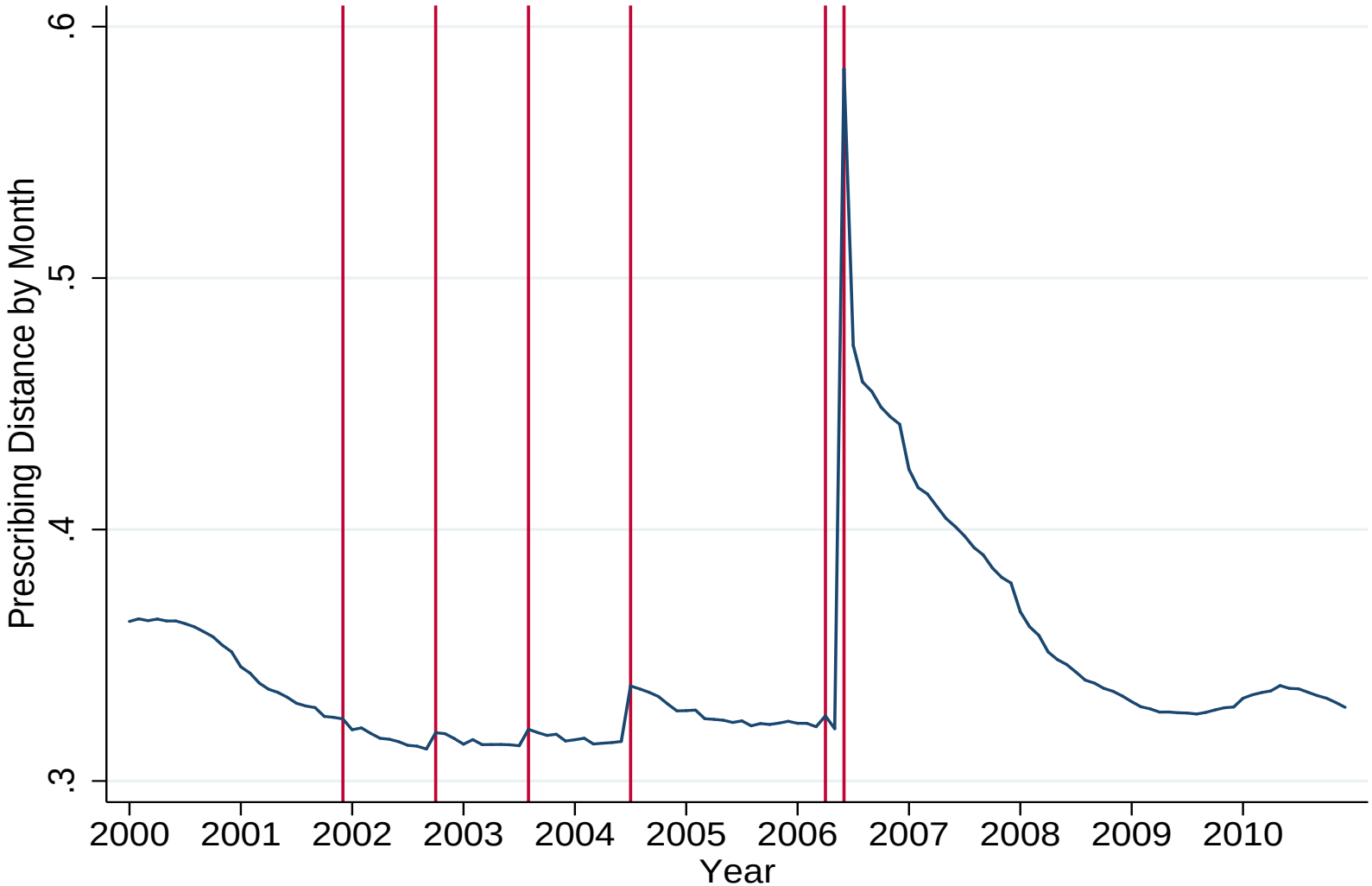
- Consider drugs $\mathcal{D}_t = \{1, 2, ..., D_t\}$ available in month t
- Group months into spells defined by major drug introductions
- π_{idt} is drug-d share of doctor i's month-t prescriptions

Distance measure y_{it}: Euclidean prescription distance between i at t and average doctor at end of spell $T(t)$,

$$y_{it} \equiv \left(\sum_{d=1}^{D_t} \left(\pi_{idt} - \overline{\pi}_{dT(t)} \right)^2 \right)^{\frac{1}{2}}$$

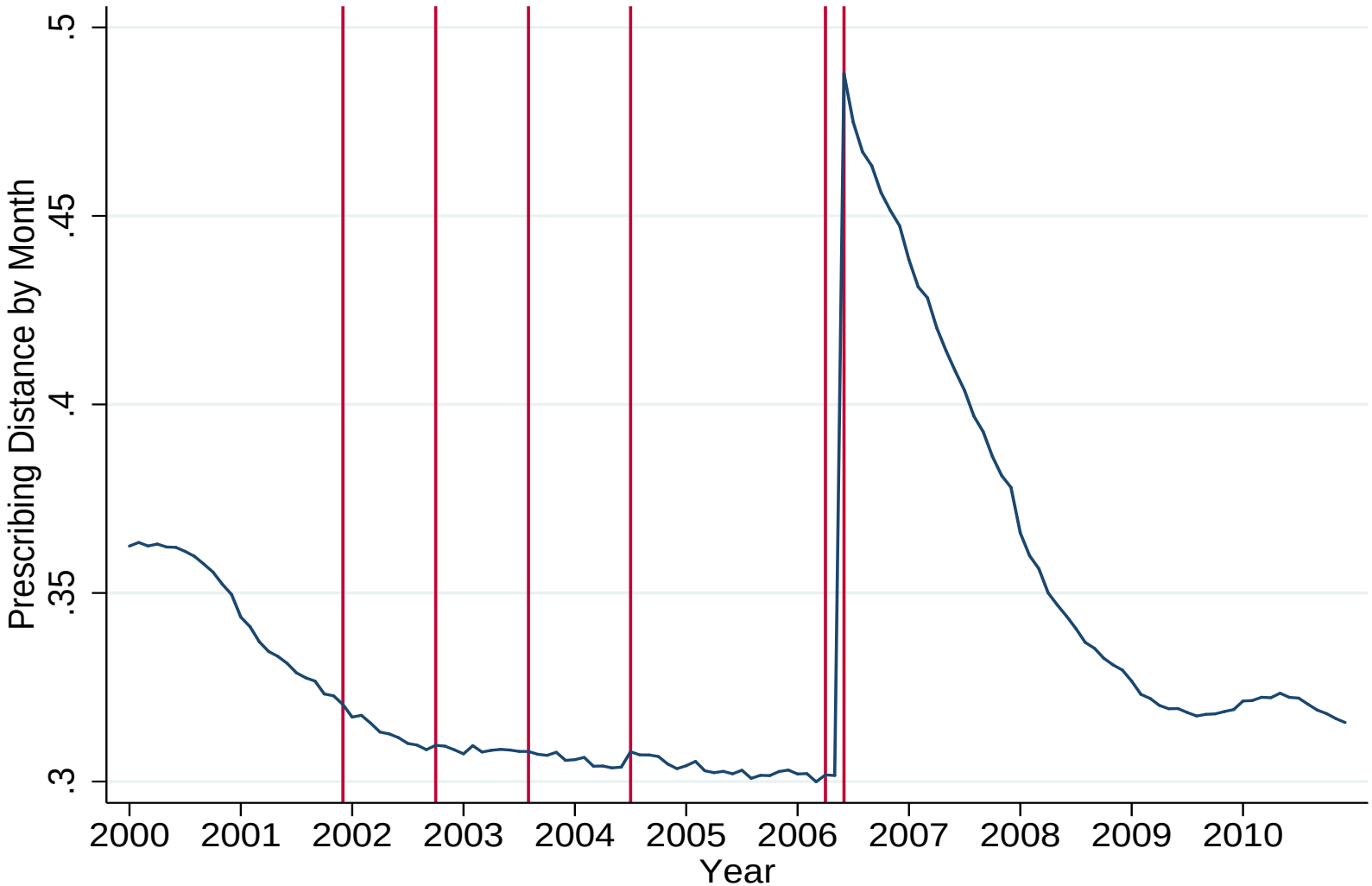
- $\overline{\pi}_{dT(t)}$ is average prescription vector at end of current spell
- Changes in $\mathcal{D}_t$ can affect convergence mechanically; perform analysis using a) all drugs, b) six molecules available at $t = 0$
- Start with simple plots of $\overline{y}_t = \frac{1}{I} \sum_i y_{it}$ over time

Scripting Patterns Converge across Doctors



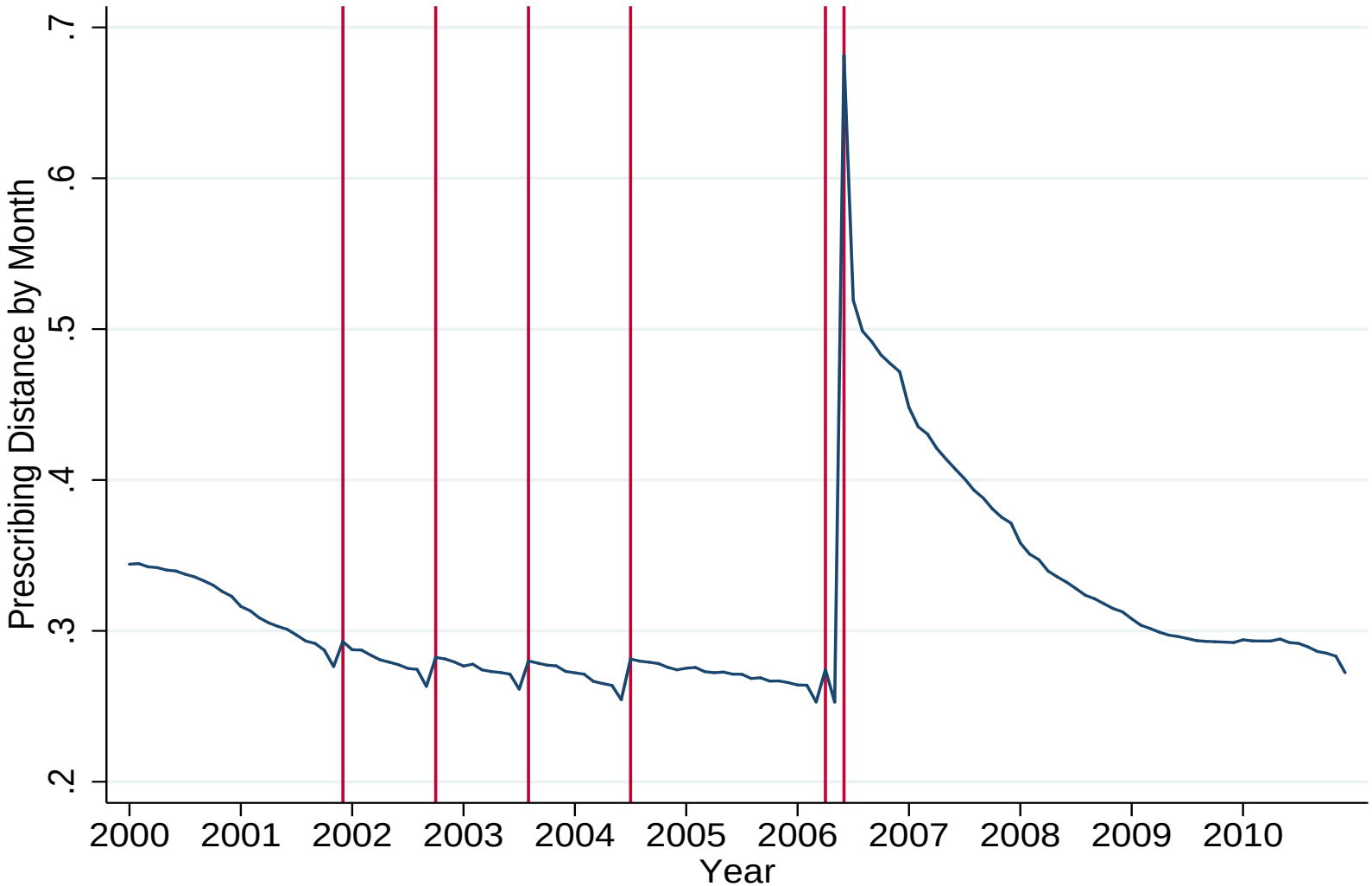
Scripting Patterns Converge across Doctors

Same convergence pattern for initial 6 drugs



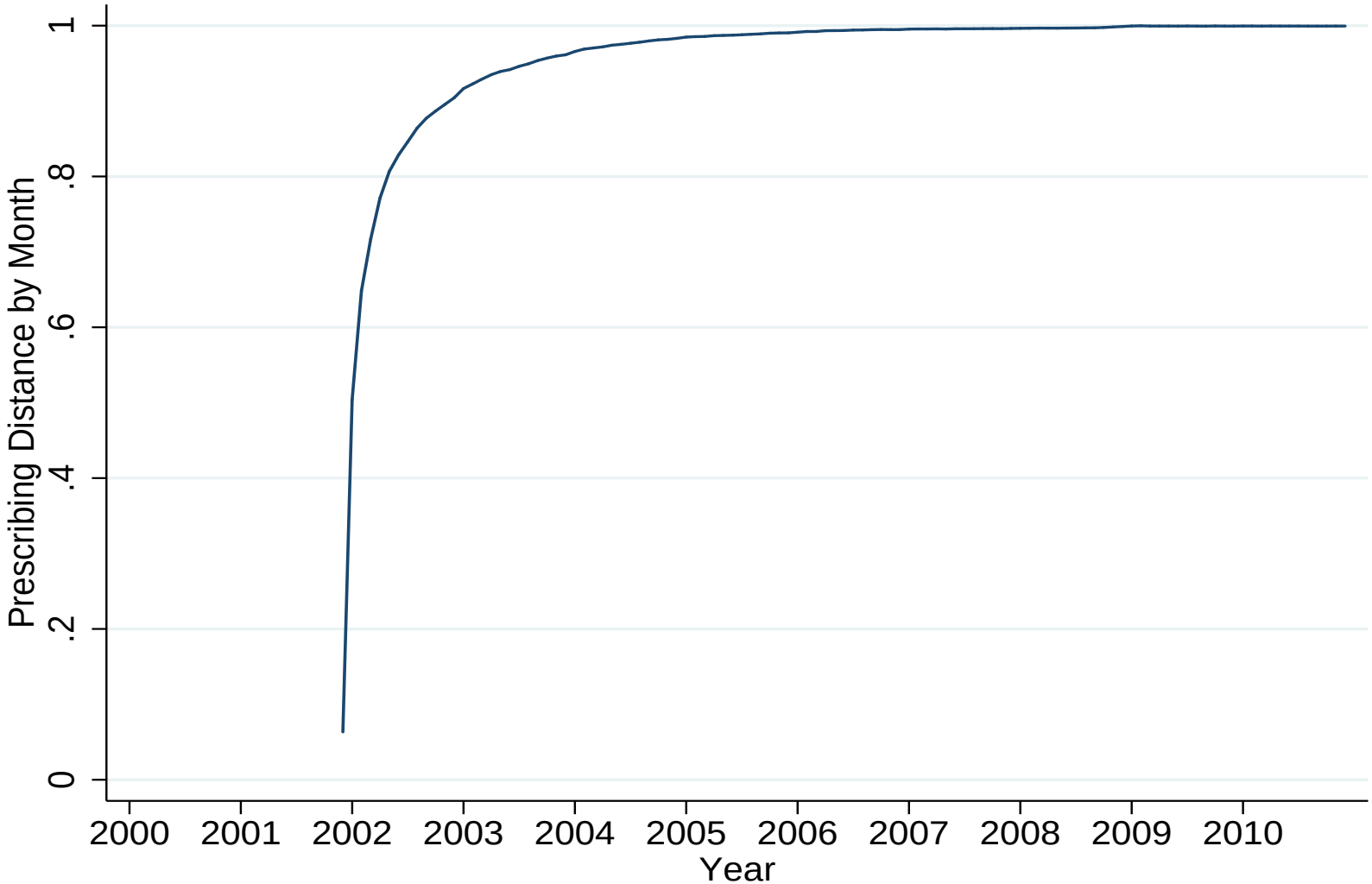
Scripting Patterns Converge across Doctors within Zipcode

Same convergence pattern for initial 6 drugs



Generic Prescription Share Converges

$\longrightarrow$ Within-molecule generic substitution, Lovastatin



Fact #2: Central doctors converge faster

1. Measure simple doctor-i specific convergence rate β_i

$$y_{it} = \beta_i \times t + \delta_i + \delta_t + \epsilon_{it}$$

2. Regress β_i on measure of the centrality of each doctor
   2.1 'Gravity' regression to construct centrality measure:

$$\widetilde{X}_{ij} = \alpha_1 \widetilde{dist}_{ij} + \alpha_2 school_{ij} + \alpha_3 cohort_{ij} + \alpha_4 specialty_{ij} + \delta_i + \delta_j + \epsilon_{ij}$$

where $X_{ij} \equiv \left(\sum_d (\overline{\pi}_{id} - \overline{\pi}_{jd})^2\right)^{1/2}$, $\widetilde{Z}_{ij} = 1/(1 + Z_{ij})$

- Estimate using a random sample of ∼3,200 doctors
- Medical specialty is dominated by three groups: internal medicine (36%), family practice (28%), and cardiology (10%)
- ∼ 250 medical schools, 8.5K zipcodes
- Median doctor graduates in 1981

Fact #2: Central doctors converge faster

2.2 Adjacency matrix A $\equiv [A_{ij}]$:

$$A_{ij} \equiv \exp\{\hat{\alpha}_1 \widetilde{dist}_{ij} + \hat{\alpha}_2 school_{ij} + \hat{\alpha}_3 cohort_{ij} + \hat{\alpha}_4 specialty_{ij}\}$$

2.3 Centrality: the eigenvector c $\equiv [c_i]$ associated with the largest eigenvalue of A contains the centrality index for each doctor i

2.4 Regress doctor-specific convergence rate β_i on centrality index c_i

$$\hat{\beta}_i = \lambda c_i + x_i + \eta_{z(i)} + \epsilon_i,$$

where x_i is prescription volume, advertising, $\eta_{z(i)}$ are zipcode FE

'Gravity' Estimates— Bilateral prescription 'proximity'

Proximity ($\widetilde{dist}$)	.1228^a	.1175^a
	(.0013)	(.0017)
Same medical school	.0036^a	.0038^a
	(.0002)	(.0002)
Same medical school cohort	.0004^a	.0003^a
	(.00003)	(.00003)
Same medical specialty	.0026^a	.0025^a
	(.00004)	(.00004)
Same school $\times \widetilde{dist}$		-.0281^a
		(0073)
Same cohort $\times \widetilde{dist}$		.0184^a
		(.0028)
Same specialty $\times \widetilde{dist}$		.0054^c
		(.0029)
Doctor i FE	Y	Y
Doctor j FE	Y	Y
R^2	0.6338	0.6338
Observations	10220788	10220788

Convergence rate is increasing in centrality index

- Equation: $\hat{\beta}_i = \lambda c_i + x_i + \eta_{z(i)} + \epsilon_i$

	Rx Distance		Generic Share	
Centrality	-.3957[a]	-.3748[a]	.0324[a]	.0322[a]
	(.0436)	(.0436)	(.0090)	(.0090)
Rx volume	-.0519[a]	-.0073	-.0141[a]	-.0145[a]
	(.0190)	(.0154)	(.0032)	(.0033)
Advertising		-.1290[a]		.0012
		(.0110)		(.0022)
Zipcode FE	Y	Y	Y	Y
R^2	.2462	.2497	.2039	.2039
Observations	37417	37417	35907	35907

Robustness: a) degree centrality, b) estimate gravity coefficients using excluded physician sample or observations from excluded time period

Summary of descriptive results

- Fact #1: Prescription shares are converging across doctors
- Fact #2: Convergence rates are increasing in network centrality
- These facts are also evident in generic substitution context
- Both facts are robust to alternative measures of convergence and centrality

Next steps motivated by descriptive results:

1. Develop structural model of scripting and learning
2. Use structural equations to estimate doctor-specific learning rates, network structure, and initial knowledge
3. Counterfactuals in simulated environment

Why a Structural Model?

- With structural model, can see how lowering barriers to knowledge diffusion will affect learning, accounting for endogeneity of effort
- Quantify how much barriers to knowledge affect drug mis-prescribing
  - Could connect to the value of improving doctors' knowledge (continuing ed., electronic medical records with or without decision support tools, etc.)
- Quantify which variables most affect knowledge diffusion: geography, age, common med school, etc. (analogy to distance, common border, common language, etc. in trade)
  - Which doctors would it be most useful to educate? Those who know least? Those most connected? Probably both important, but not one in the same, so meaningful tradeoff.
- Use detailed data as model selection guide: different models of networks and learning

Roadmap: Summary of Structural Model

- Drug discrete choice problem relates drug-shares π_{idt} to beliefs about drug quality and risk aversion (static problem given beliefs)
- Beliefs are distributed Normal and update according to Bayesian learning
  - LoM for beliefs depends on number of signals and signal distribution
- Social learning: Signals received depend on network
  - Number of signals depends on who doctor is connected to and how much those doctors know
  - Signal distribution depends on who doctor is connected to, how much those doctors know, and *what those doctors believe* (not today)
- Doctors choose investment in learning (number of signals) to maximize PDV of their expected utility (not today)
- No doctor-specific patient population and no advertising (as of now)

Patient Utility and Doctor Beliefs

- Consider a spell with a fixed set $\mathcal{D}$ of available drugs
- Each doctor $i \in \{1, ..., N\}$ treats a unit measure of patients $\nu_i \in [0, 1]$ in period $t = 1, 2, ..., T$
- Patient ν_i reward from drug $d \in \mathcal{D}$ is

$$u_{dt}(\varepsilon_{dt}(\nu_i)) = \beta_d + \varepsilon_{dt}(\nu_i)$$

  - The true unconditional efficacy of drug d, β_d
  - An idiosyncratic observed shock, $\varepsilon_{dt}(\nu) \sim$ Gumbel $F(\cdot)$

- However, doctors are only imperfectly informed about $\{\beta_d\}_{d \in \mathcal{D}}$
  - Doctor i beliefs at t about β_d are summarized by a Normal distribution $G_{idt}(\cdot)$ with mean $\tilde{\beta}_{idt}$ and variance σ^2_{idt}

Doctor Expected Utility from Treating Patient ν_i

- Let doctors have CARA preferences represented by $\mathcal{U}_{idt}(\varepsilon_{dt}(\nu_i), x)$
- Expected utility of doctor i treating patient ν_i with drug d given beliefs $x \sim G_{idt}(\cdot)$ is

$$\begin{aligned}
\mathcal{U}_{idt}(\varepsilon_{dt}(\nu_i), \tilde{\beta}_{idt}, \sigma_{idt}) &:= \int -\exp(-\alpha(x + \varepsilon_{dt}(\nu_i)))dG_{idt} \\
&= -\exp(-\alpha\varepsilon_{dt}(\nu_i)) \int \exp(-\alpha x)dG_{idt} \\
&= -\exp(-\alpha\varepsilon_{dt}(\nu_i)) \exp\left(-\alpha\tilde{\beta}_{idt} + \frac{1}{2}\alpha^2\sigma^2_{idt}\right) \\
&= -\exp\left(-\alpha\tilde{\beta}_{idt} + \frac{1}{2}\alpha^2\sigma^2_{idt} - \alpha\varepsilon_{dt}(\nu_i)\right)
\end{aligned}$$

Expected Utility, Beliefs, and Prescription Patterns

Assume learning is independent of prescription choice (verified later).

- Doctor i chooses a drug at t for each patient ν_i to maximize expected doctor utility given current beliefs, yielding expected utility

$$U_{it}(\boldsymbol{\varepsilon_{dt}}(\nu_i), \tilde{\boldsymbol{\beta}}_{it}, \boldsymbol{\sigma}_{it}) := \max_{d \in D}\{\mathcal{U}_{idt}(\varepsilon_{dt}(\nu_i), \tilde{\beta}_{idt}, \sigma_{idt})\}.$$

- Given $\varepsilon_{dt}(\nu_i)$ is distrubted Gumbel, the doctor faces a standard multinomial choice problem, such that doctor i chooses drug d for patient ν_i with probability π_{idt}:

$$\pi_{idt}(\boldsymbol{\varepsilon_{dt}}(\nu_i), \tilde{\boldsymbol{\beta}}_{it}, \boldsymbol{\sigma}_{it}) :=$$
$$\Pr\left\{\tilde{\beta}_{idt} - \alpha\sigma^2_{idt}/2 + \epsilon_{idt}(\nu_i) > \tilde{\beta}_{id't} - \alpha\sigma^2_{idt}/2 + \epsilon_{id't}(\nu_i),\ \forall d' \neq d\right\}$$
$$= \frac{\exp(\tilde{\beta}_{idt} - \alpha\sigma^2_{idt})}{\sum_{d' \in \mathcal{D}} \exp(\tilde{\beta}_{id't} - \alpha\sigma^2_{id't})}$$

- Empirical analog: π_{idt} is drug-d share in doctor i's portfolio at time t

Doctor Expected Utility from Treating all Patients

- Finally, doctor-i's period payoff considering all the patients she treats at t is

$$W_{it}(\tilde{\boldsymbol{\beta}}_{it}, \sigma_{it}) \equiv \int U_{it}(\varepsilon_{dt}(\nu_i), \tilde{\boldsymbol{\beta}}_{it}, \sigma_{it}) dF(\varepsilon_{dt}(\nu_i)).$$

- And the dynamic problem of a doctor is to choose $l_{it} \in \mathcal{R}^+$:

$$V_{it}(\tilde{\boldsymbol{\beta}}_{it}, \sigma_{it}) \equiv \max_{\{l_{i\tau}\}_{\tau=t}^{\infty}} \sum_{\tau \geq t} \delta^{\tau - t}(W_{i\tau}(\tilde{\boldsymbol{\beta}}_{i\tau}, \sigma_{i\tau}) - c(l_{i\tau}))$$

L.O.M. for $\boldsymbol{\beta_{it}}$ and σ_{it} (which depends on l_{it})

Normal Bayesian learning: LoM for β_{it} and σ_{it} given f_{it}

- Prior beliefs are distributed around true values: $\tilde{\beta}_{id0} \sim N(\beta_d, \sigma_d^2)$
- Each doctor receives f_{idt} signals at t, values $\{x_{dn}\}, n = 1, ..., f_{idt}$
  - Signals accumulate into knowledge $S_{idt} = S_{id0} + \sum_{u=1}^{t} f_{idu}$
  - For today, signal values are centered around truth: $x_{dn} \sim N(\beta_d, \sigma_d^2)$
  - Let $\bar{x}_{idt} = \sum_{n=1}^{f_{idt}} x_{dn} / f_{idt}$

    $\implies \bar{x}_{idt} \sim \beta_d + \epsilon_{idt}$, where $\epsilon_{idt} \sim N(0, \sigma_d^2 / f_{idt})$
- Standard Bayesian learning then implies,

$$\tilde{\beta}_{idt+1} = \frac{\tilde{\beta}_{idt} S_{idt} + \bar{x}_{idt+1} f_{idt+1}}{S_{idt} + f_{idt+1}}$$

$$\sigma^2_{idt} = \frac{\sigma^2_{id0}}{S_{idt}}$$

- Assumptions for today:
  - Assume signals are general information: $f_{idt} = f_{it} \ \forall \ d, t$
  - Assume common prior knowledge: $S_{id0} = S_{i0} \ \forall \ d$
  - Then $S_{idt} = S_{it}$ and $\sigma^2_{idt} = \frac{\sigma_d^2}{S_{it}} \ \forall \ d, t$

Knowledge diffusion through networks: f_{it} and LoM for S_{it}

- Our model of i's knowledge flow at t (f_{it}) depends on (1) investment in learning I_{it}, (2) network connections, and (3) other doctors' stocks of knowledge as follows

$$f_{it+1} = I_{it} \times \sum_j \tau_{ij} S_{jt}$$

  where $\tau_{ij} \geq 0$ reflects the strength of network connections
- Then the law of motion for the stock of knowledge is

$$\begin{aligned} S_{it+1} &= Q(I_{it}, \mathcal{S}_t, \mathcal{T}) \\ &= S_{it} + f_{it+1} \\ &= S_{it} + I_{it} \times \sum_j \tau_{ij} S_{jt} \end{aligned}$$

- where $\mathcal{S}_t$ is the vector of S_{it} and $\mathcal{T}$ is matrix of τ_{ij}
- We've also worked with $Q\left(S_{i,t}, \left\{I_{ijt}\right\}, \mathcal{T}\right) = \delta S_{i,t} + I_{it} \sum_j \tau_{ij}^{-\varepsilon} I_{jt}$

Summary of Model

- Doctor discrete choice relates drug-shares π_{idt} to beliefs $\tilde{\boldsymbol{\beta}}_{it}, \sigma_{it}$ and risk aversion

$$\pi_{idt} = \frac{\exp(\tilde{\beta}_{idt} - \alpha\sigma^2_{idt})}{\sum_{d' \in \mathcal{D}} \exp(\tilde{\beta}_{id't} - \alpha\sigma^2_{id't})}.$$

- Beliefs update according to Bayesian Learning

$$\tilde{\beta}_{idt} = \tilde{\beta}_{idt-1}\frac{S_{it-1}}{S_{it}} + \overline{x}_{idt}\frac{f_{it}}{S_{it}} \quad \text{(Mean)}$$

$$\text{where } \overline{x}_{idt} \sim \beta_d + \epsilon_{idt} \text{ and } \epsilon_{idt} \sim N(0, \sigma_d^2 / f_{idt})$$

$$\sigma^2_{idt} = \frac{\sigma_d^2}{S_{it}} \quad \text{(Variance)}$$

- Amount of learning for doctor i given by network structure and knowledge of all doctors

$$\begin{aligned} S_{it} &= S_{it-1} + f_{it} \\ &= S_{it-1} + I_{it} \times \sum_j \tau_{ij} S_{jt-1} \end{aligned}$$

- Doctors choose I_{it} to maximize PDV of expected utility

Next Steps

- Assume $I_{it} = 1$ for today. No endogenous investment in learning
- Only use Bayesian learning + discrete choice equations to estimate f_{it}
- Then use these estimates of f_{it} with model of network learning to estimate τ_{ij} and S_{i0} (and thus S_{it} $\forall$ i, t)
  - Normalize $S_{N0} = 1$
- Perform counterfactuals in simulated model about speed of learning w.r.t. different networks and initial knowledge
- Note: we have some theoretical results on endogenous learning and relaxing some homogeneity assumptions, but not with empirical results ready for today

Bayesian Learning Estimating Equation

- Use $S_{it+1} = S_{it} + f_{it+1}$ and define $g_{it} := f_{it}/S_{it}$ to find

$$\begin{aligned}\tilde{\beta}_{idt+1} &= \tilde{\beta}_{idt}\, S_{it}/S_{it+1} + (\beta_d + \epsilon_{idt+1}) f_{it+1}/S_{it+1} \\ &= \tilde{\beta}_{idt}\,(1 - g_{it+1}) + \beta_d g_{it+1} + \epsilon_{it+1} g_{it+1}\end{aligned}$$

- Substituting $\tilde{\beta}_{idt} = \ln \pi_{idt} + \alpha\sigma_{it} + \eta_{it}$, we find

$$\begin{aligned}\ln \pi_{idt+1} = \ln \pi_{idt} \cdot \underbrace{(1 - g_{it+1})}_{=\rho_{it+1}} + \underbrace{\eta_{it}(1 - g_{it+1}) - \eta_{it+1}}_{=\kappa_{it}} \\ + \underbrace{\beta_d}_{=\delta_d} \cdot \underbrace{g_{it+1}}_{=\gamma_{it}} + \underbrace{\epsilon_{idt+1} g_{it+1}}_{=u_{idt+1}} \\ = \ln \pi_{idt} \cdot \rho_{it+1} + \kappa_{it} + \delta_d \cdot \gamma_{it+1} + u_{idt+1}\end{aligned}$$

- Introduce reference drug d' and spells τ to get main equation

$$\ln(\pi_{idt+1}/\pi_{id't+1}) = \ln(\pi_{idt}/\pi_{id't}) \cdot \rho_{it+1} + \delta_{d\tau} \cdot \gamma_{it+1} + u_{idt+1}$$

Estimation results for persistence ρ_{it} :

- Estimate by quarter; aggregate to molecule level
- Replace $\delta_d \cdot \gamma_{it}$ with $\delta_d + \gamma_i + \gamma_t$

Spell	1	2	3	4	5	6	7
Persistence—							
Mean	0.81	0.80	0.80	0.81	0.81	0.85	0.87
Std Dev	0.39	0.43	0.40	0.33	0.25	0.42	0.22
25^{th} pctile	0.69	0.68	0.68	0.70	0.72	0.71	0.78
75^{th} pctile	0.96	0.97	0.95	0.96	0.93	1.00	0.97
Doctor FE	Y	Y	Y	Y	Y	Y	Y
Drug FE	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y
Observations	787K	362K	558K	516K	1.4M	231K	4.3M

Estimating τ_{ij} and S_{i0}

Combining network signal structure and Bayesian learning yields:

$$\begin{aligned}\hat{\rho}_{it} &= 1 - f_{it}/S_{it} \\ &= 1 - \frac{\kappa \times \sum_j \tau_{ij} S_{jt-1}}{S_{it}} \\ &= 1 - \frac{\kappa \times \sum_j \tau_{ij} \frac{S_{j0}}{\prod_{u=1}^{t-1} \hat{\rho}_{ju}}}{\frac{S_{i0}}{\prod_{u=1}^{t-1} \hat{\rho}_{iu}}},\end{aligned}$$

Parameterize $\tau_{ij}(\boldsymbol{\alpha}) = \tau\left(\widetilde{dist}_{ij}, school_{ij}, cohort_{ij}, specialty_{ij}; \boldsymbol{\alpha}\right)$ as follows,

$$\tau_{ij}(\boldsymbol{\alpha}) = \alpha_d \widetilde{dist}_{ij} + \alpha_s school_{ij} + \alpha_c cohort_{ij} + \alpha_{sp} specialty_{ij}$$

to get

$$\hat{\rho}_{it} = 1 - \frac{\sum_j \tau_{ij}(\boldsymbol{\alpha}) \frac{S_{j0}}{\prod_{u=1}^{t-1} \hat{\rho}_{ju}}}{\frac{S_{i0}}{\prod_{u=1}^{t-1} \hat{\rho}_{iu}}}$$

Estimation results for initial knowledge S_{i0} - preliminary

- Distribution of standardized S_{i0} highly skewed

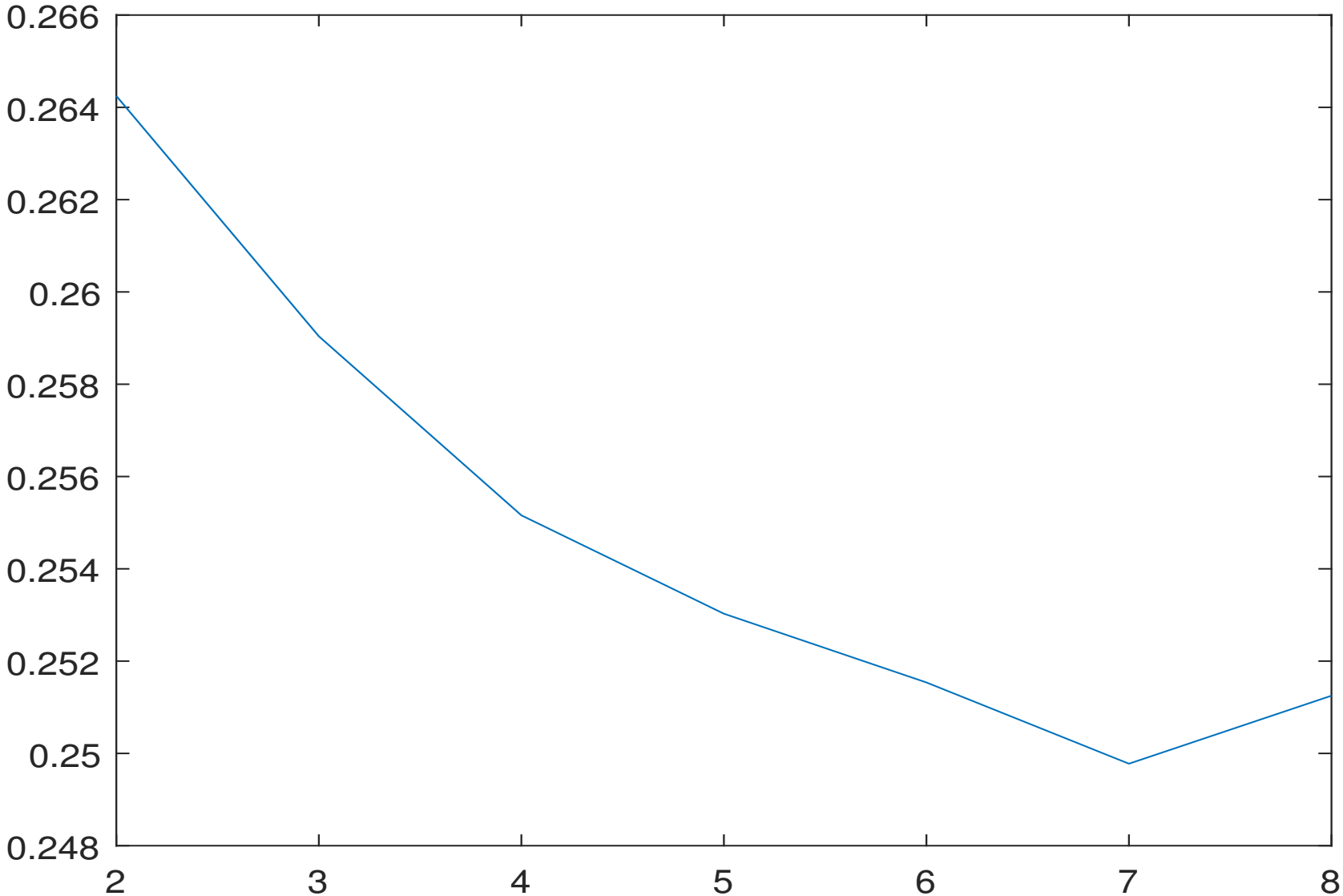
Distribution of $\frac{S_{i0}-\bar{S}_0}{\sigma_{S_0}}$	
p1	-.110
p50	-.096
p75	-.067
p90	.025
p95	.155
p99	1.25
Skewness	24.5
Kurtosis	673

Simulate the Economy

- Restrict analysis to first spell ($D = 6$, $T = 8$ quarters)
- Set $\sigma^2 = \#$ (chosen to eyeball-match rate of convergence)
- Set initial period beliefs $\tilde{\beta}_{i,d,1}$ consistent with initial period drug shares $\pi_{i,d,1}$
- Set true β_d equal to that consistent with average of final period drug shares across doctors.
- Use estimated α_d and S_0 to construct f_{it} $\forall$ i, t
- Use Bayesian updating to construct $\tilde{\beta}_{i,d,t}$ $\forall$ $i, d, (t > 1)$
  - Signals centered around truth, but finite signals means noisy updating
- Use $\tilde{\beta}_{i,d,t}$ to construct π_{idt} $\forall$ $i, d, (t > 1)$

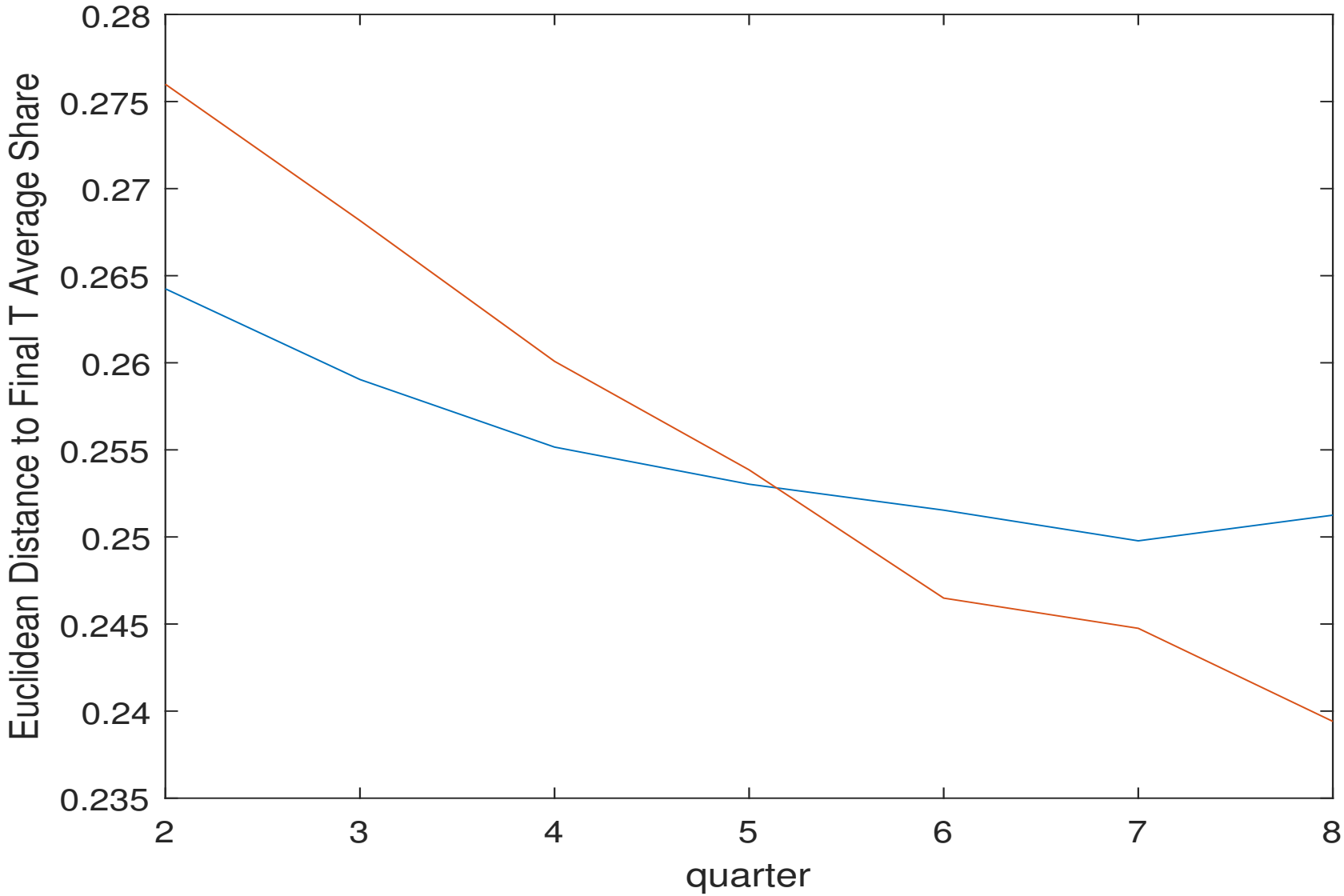
Data: Drug Shares Converge Across Doctors

Euclidean Distance of Prescription Shares to Average T-shares



Simulation: Drug Shares Converge Across Doctors

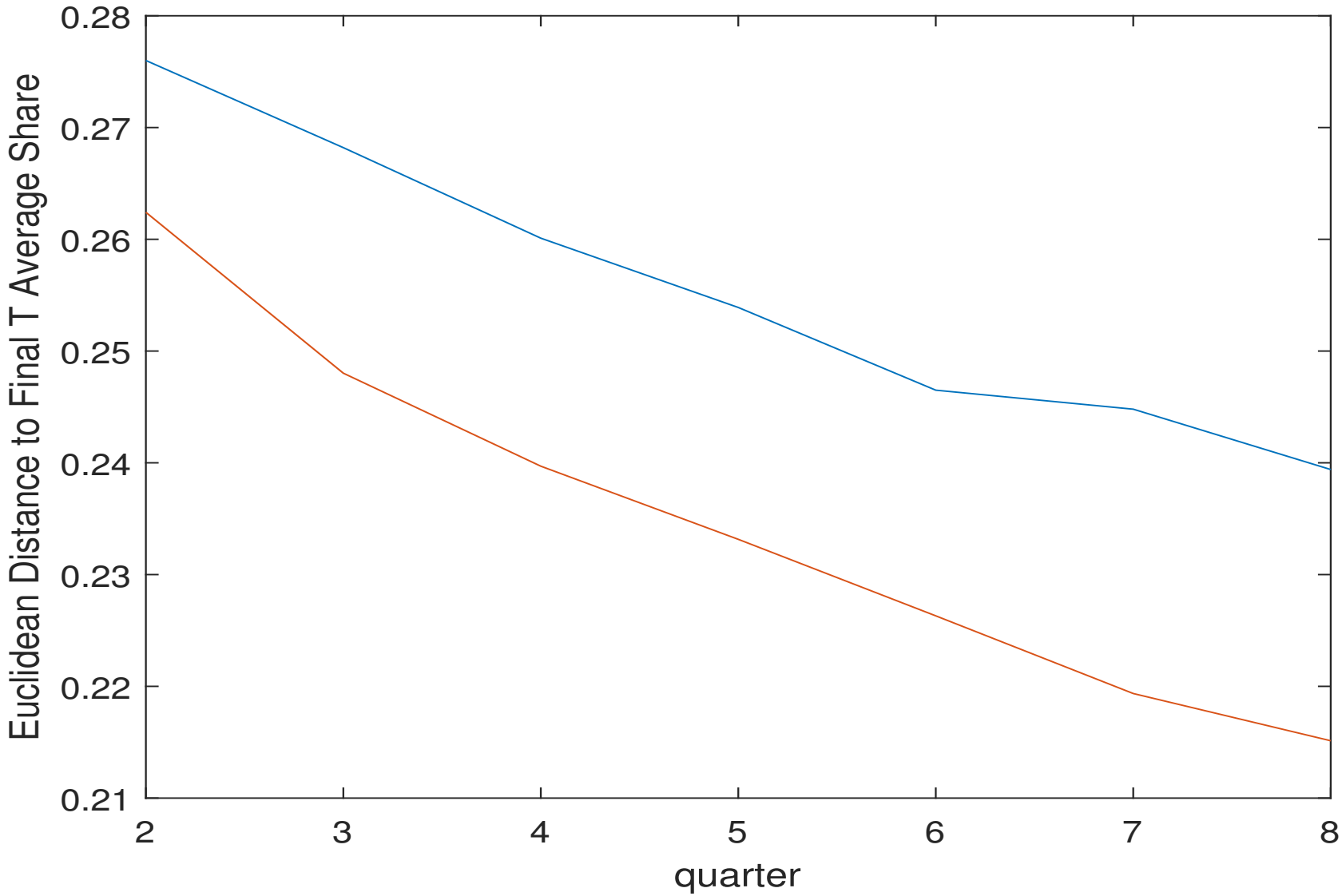
Euclidean Distance of Prescription Shares to Average T-shares



Sim: Double Initial Knowledge Uniformly

Set $S_{i0} = 2 \times S_{i0}^{\text{baseline}}$

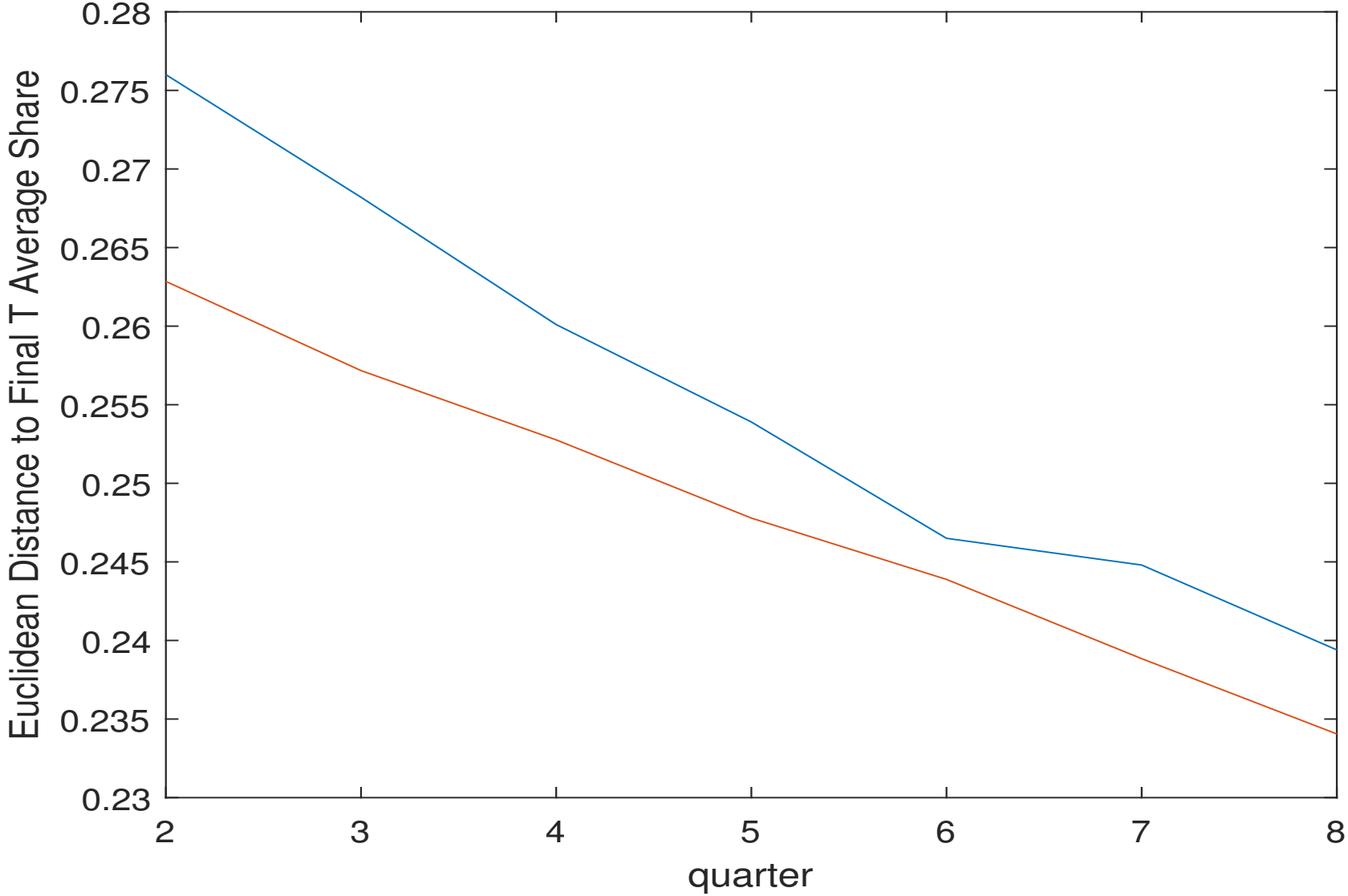
A bit less distance and more convergence. Meaningful only because not uniform network.



Sim: Increase Initial Knowledge of Least Knowledgeable

Set all $S_{i0} < p25(S_{i0}) = p25(S_{i0})$

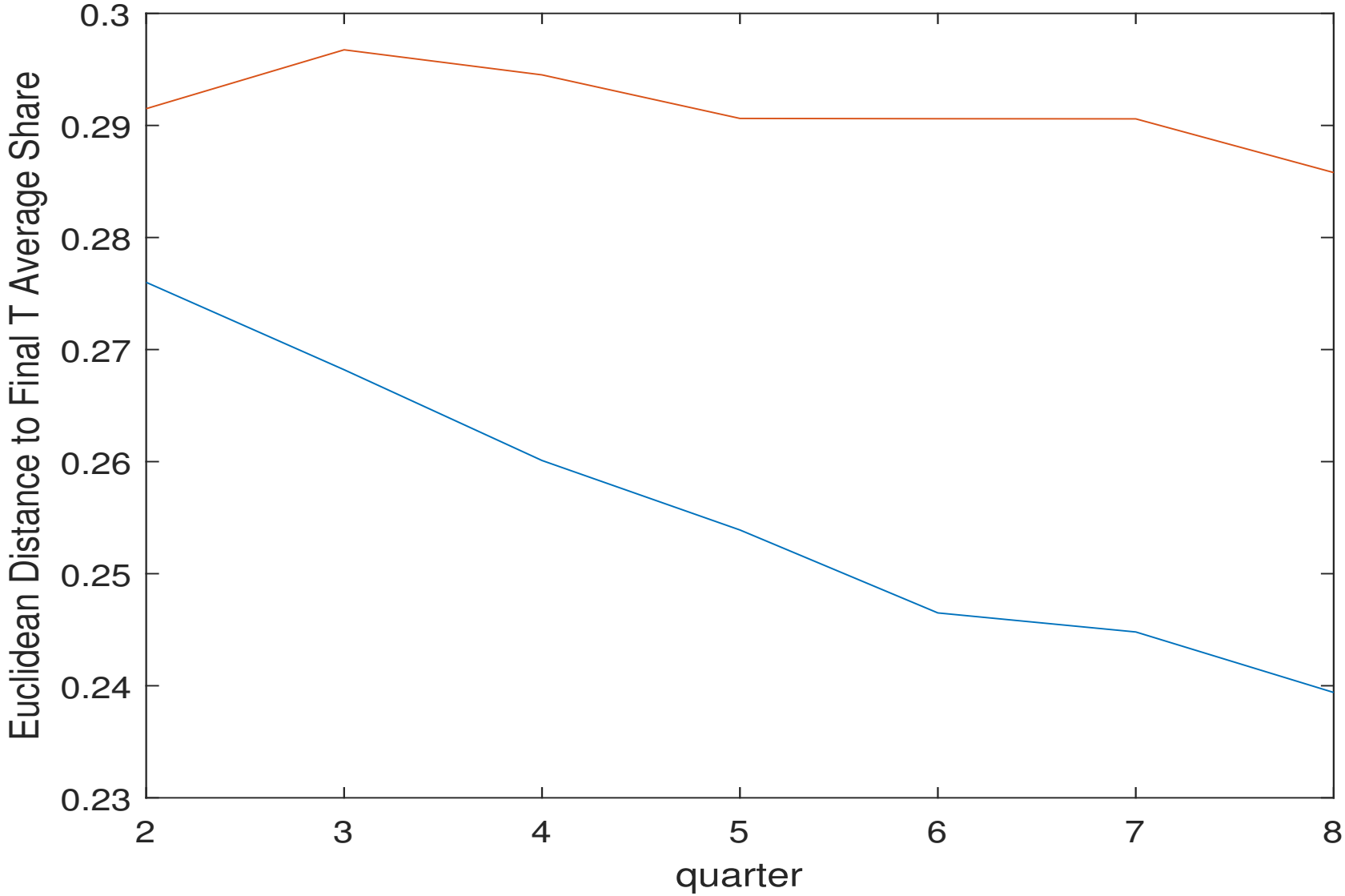
Less initial distance, but also less convergence.



Sim: Decrease Initial Knowledge of Most Knowledgeable

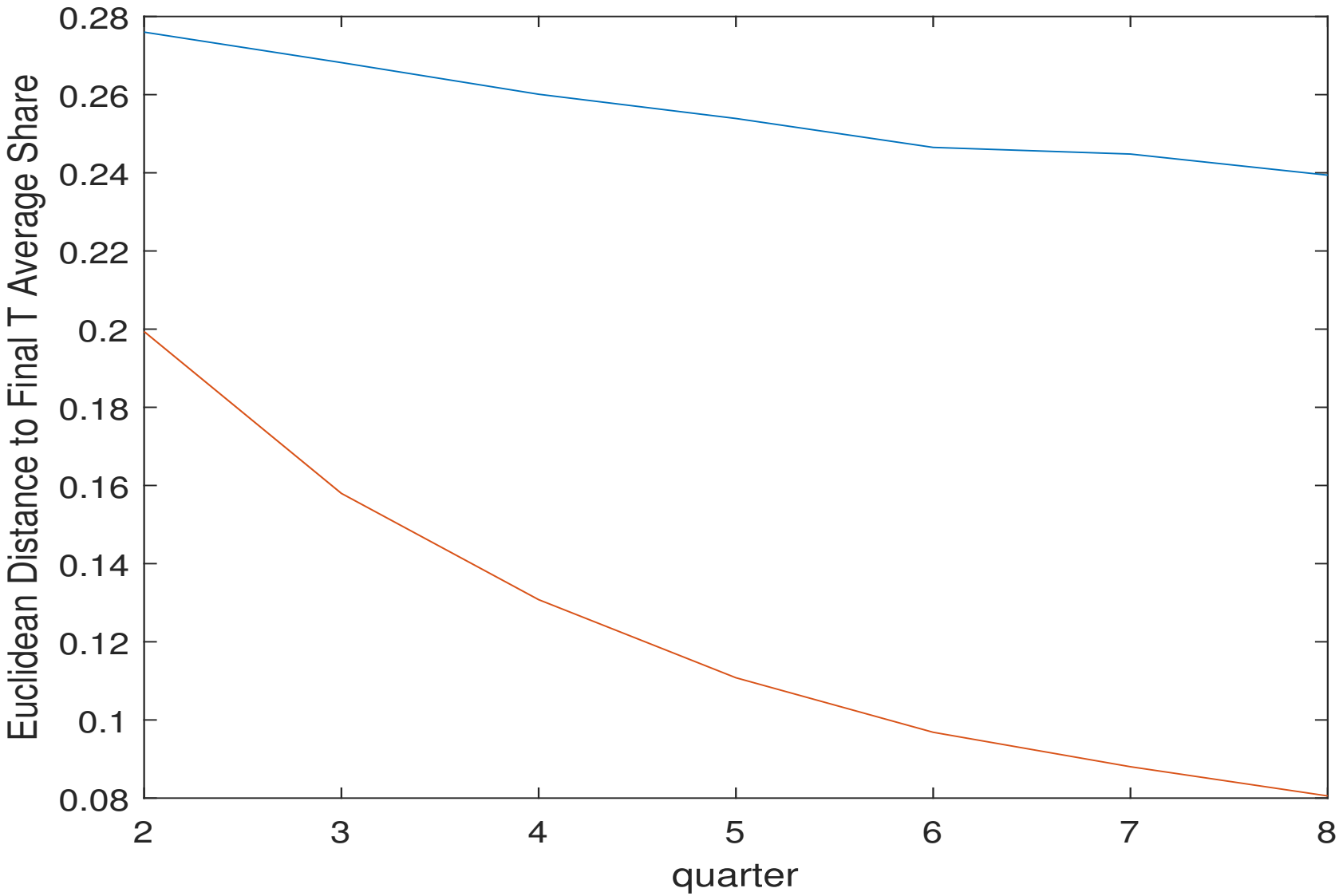
Set all $S_{i0} > p75(S_{i0}) = p75(S_{i0})$

Much bigger distances, much slower convergence. Skewness.



Sim: Drug Shares Converge Faster when Distance is Shorter

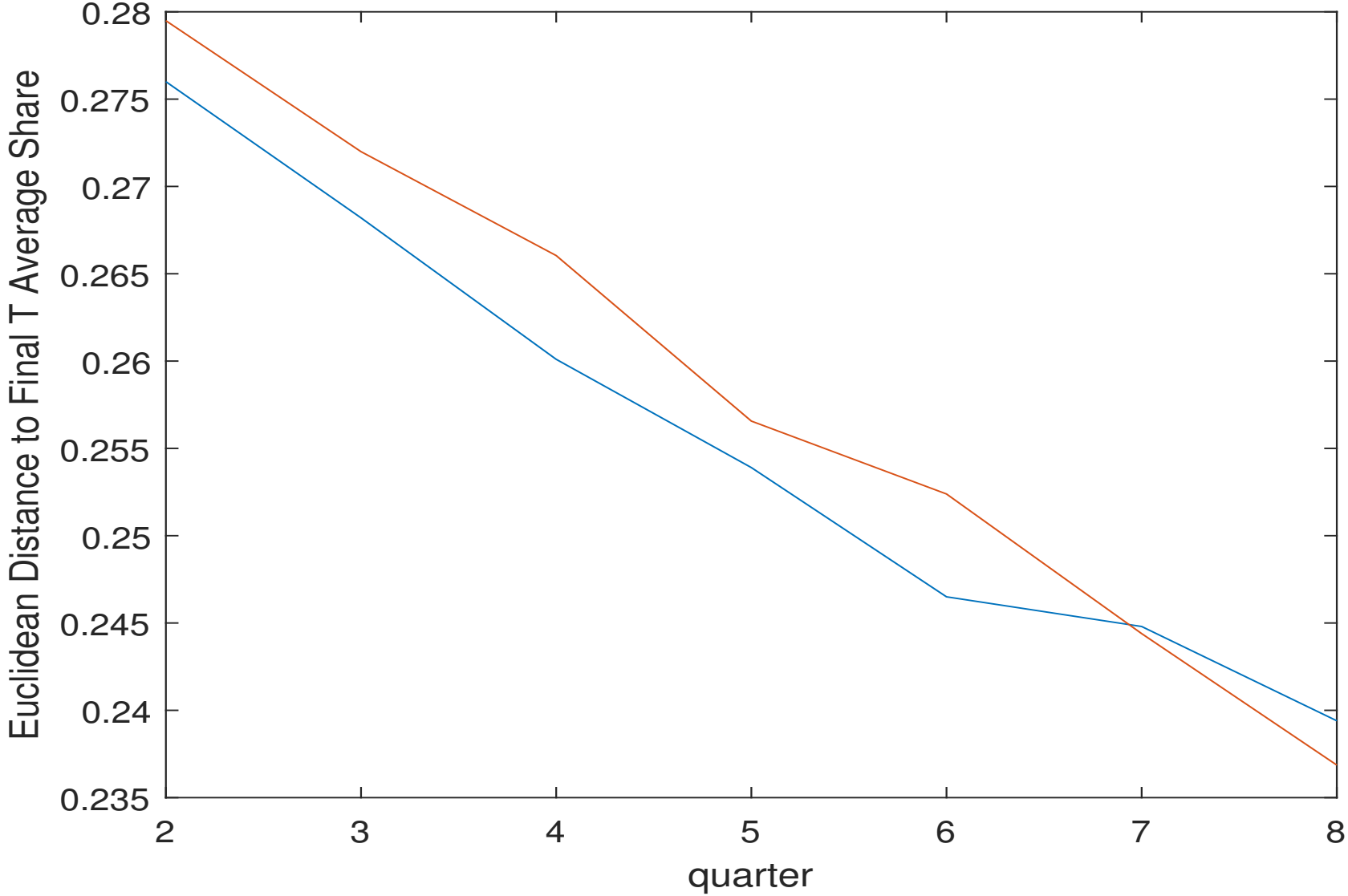
Set $\tau_{ij} = 10 \times \tau_{ij}^{\text{baseline}}$



Sim: Make Most Distant Doctors More Connected

Set all $\tau_{ij} < p25(\tau_{ij}) = p25(\tau_{ij})$

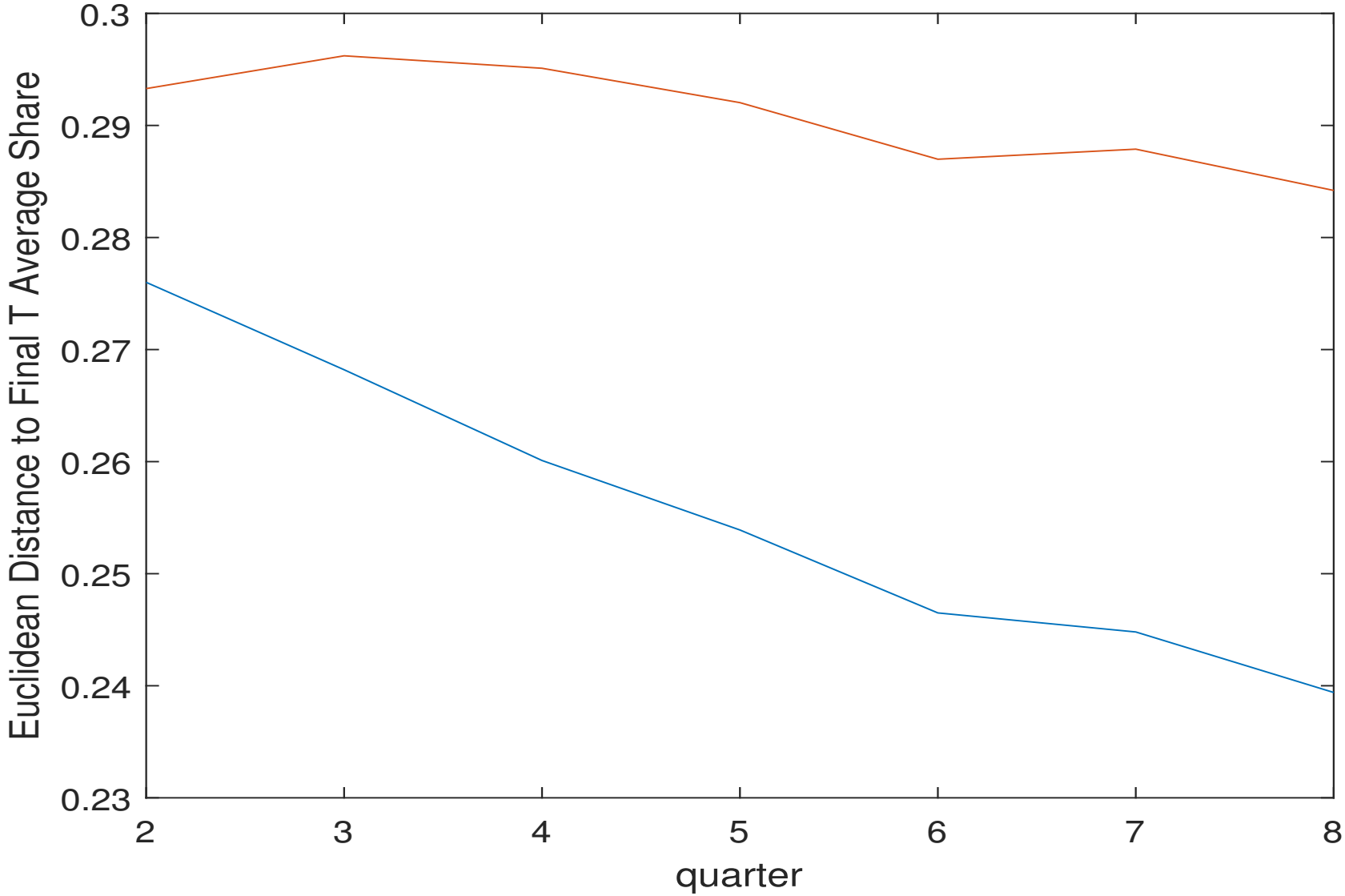
Not much changes



Sim: Make Most Connected Doctors Less Connected

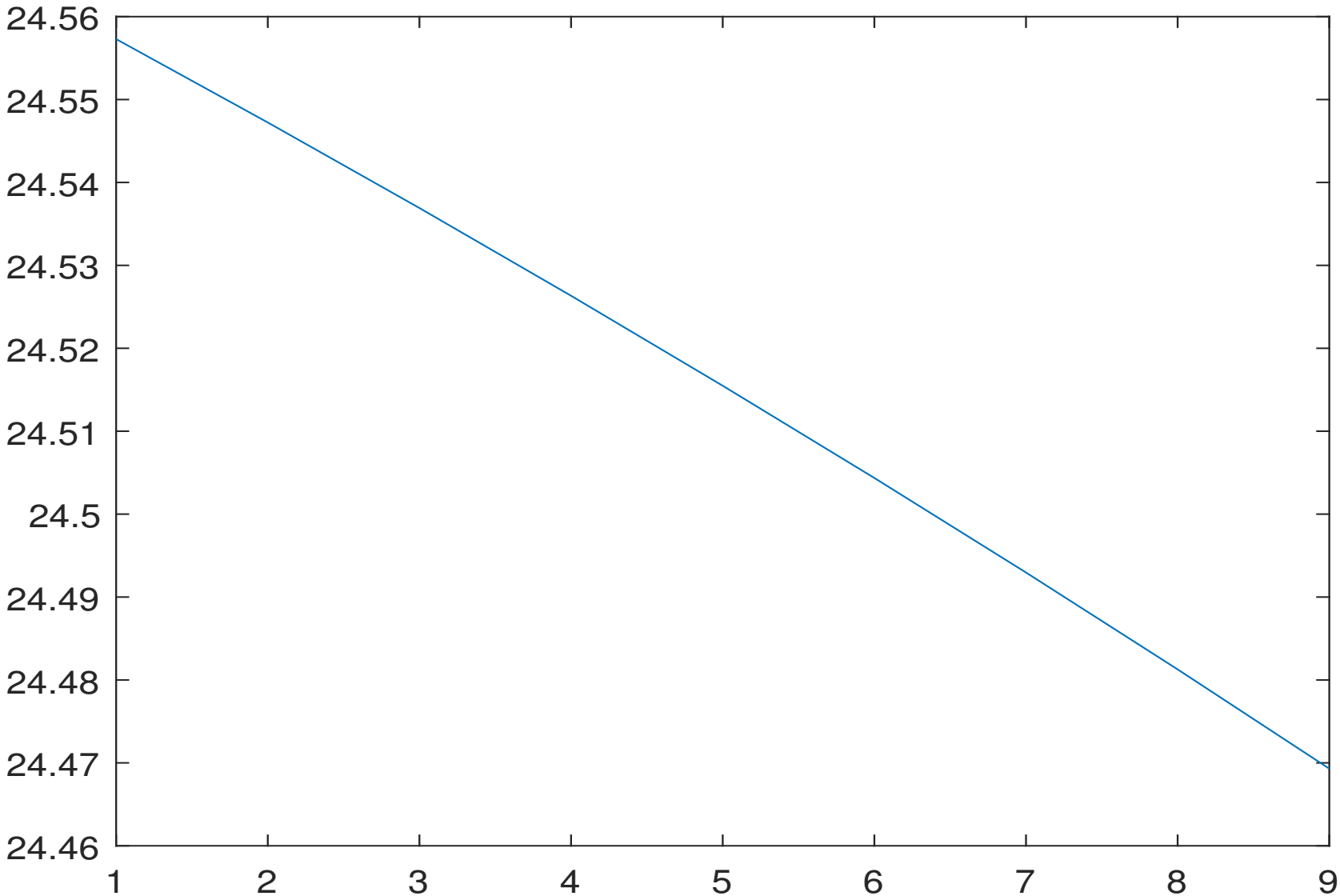
Set all $\tau_{ij} > p75(\tau_{ij}) = p75(\tau_{ij})$

Much slower convergence



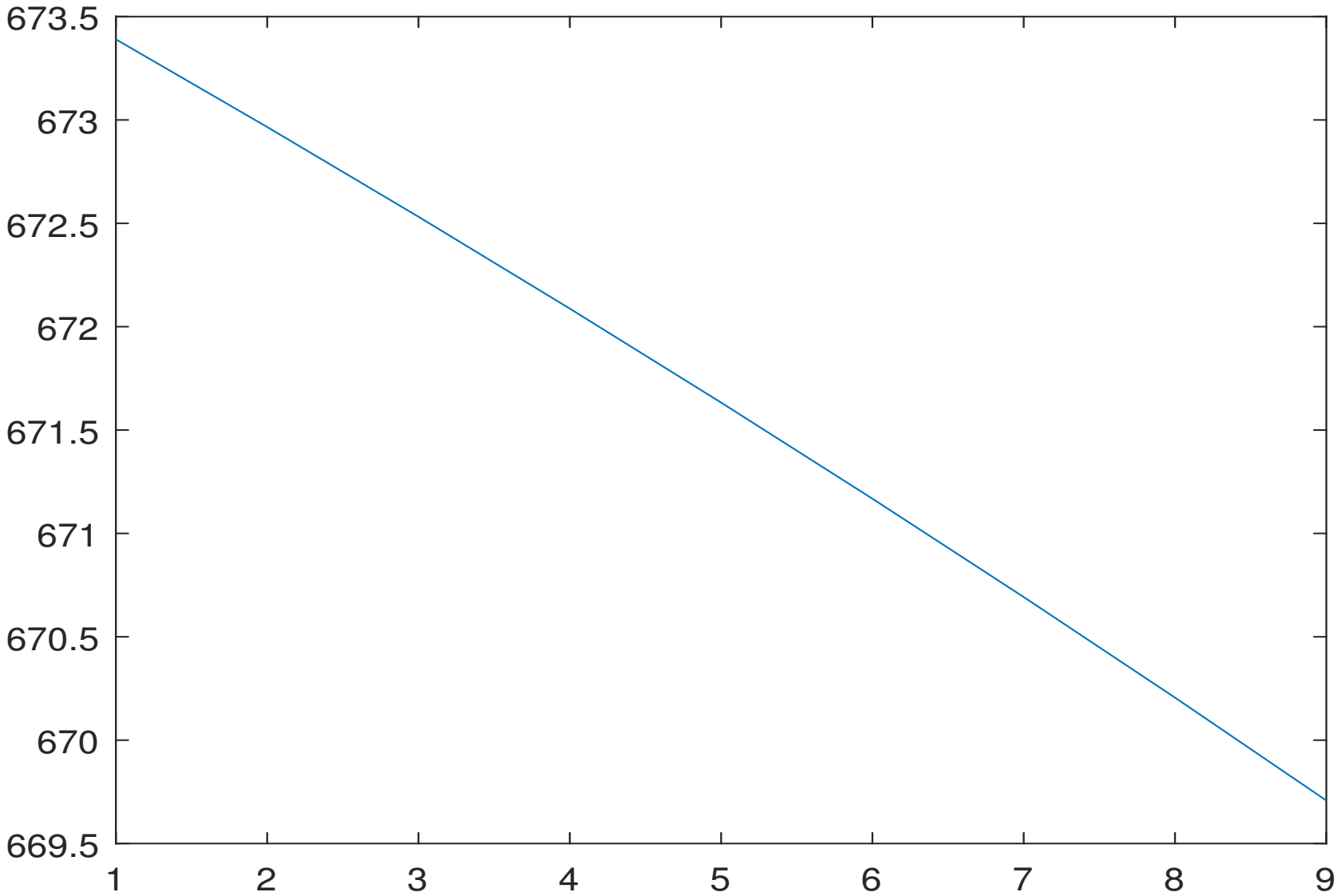
Simulation: S_{it} Skewness Decreasing Over Time

Diffusion of information is an equalizing force

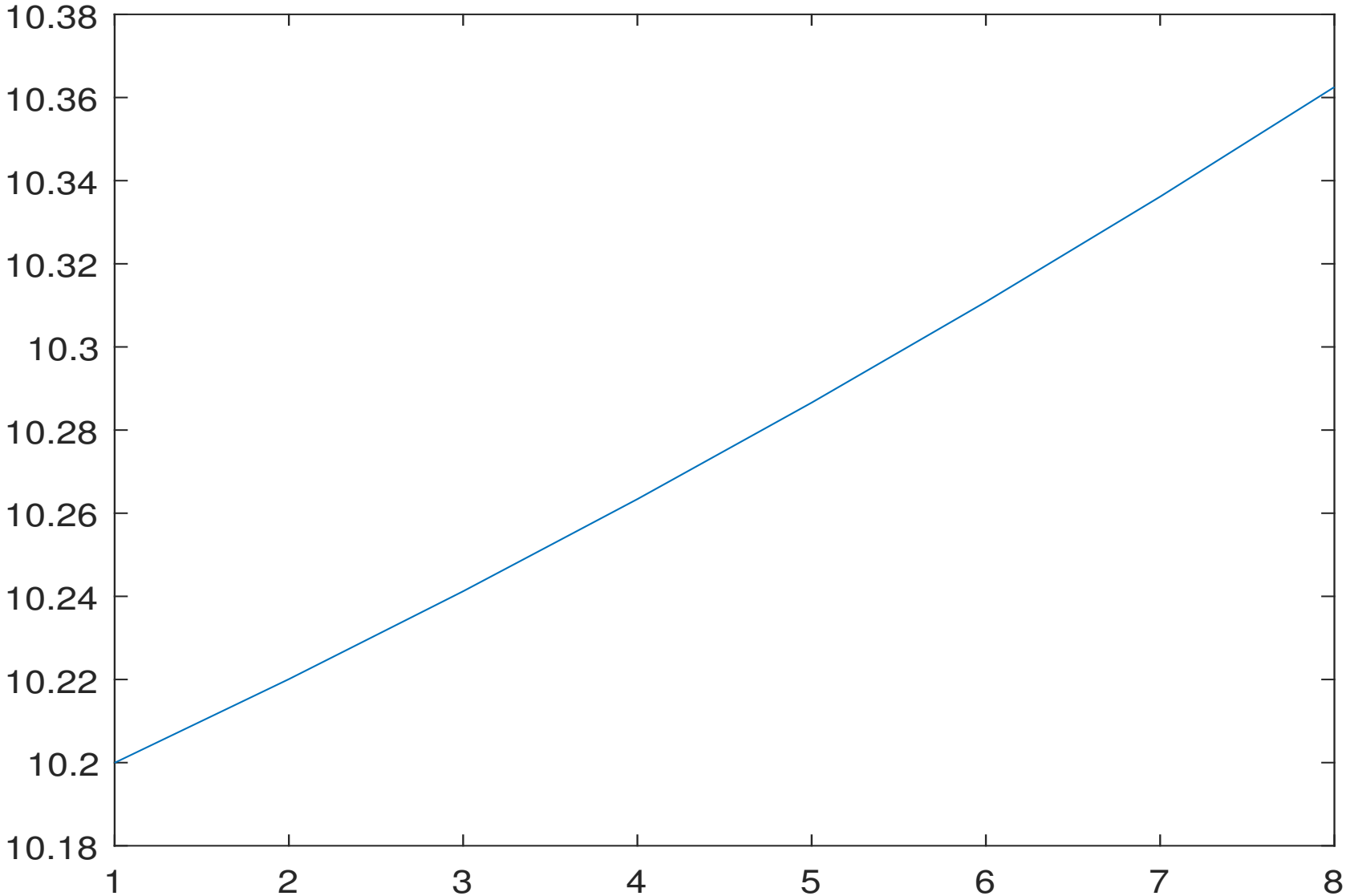


Simulation: S_{it} Kurtosis Decreasing Over Time

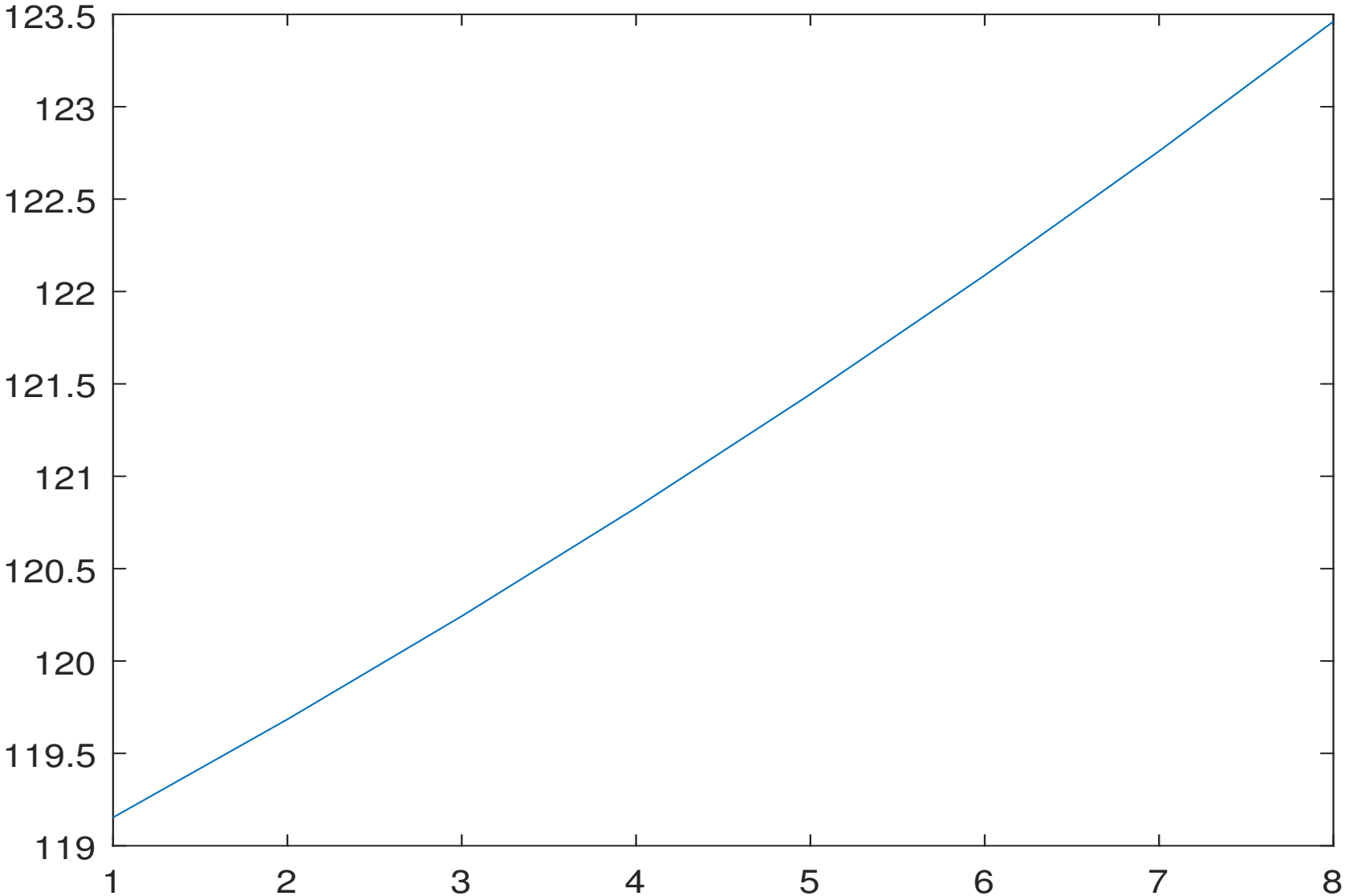
Diffusion of information is an equalizing force



Simulation: f_{it} Skewness Less Than S_{it}

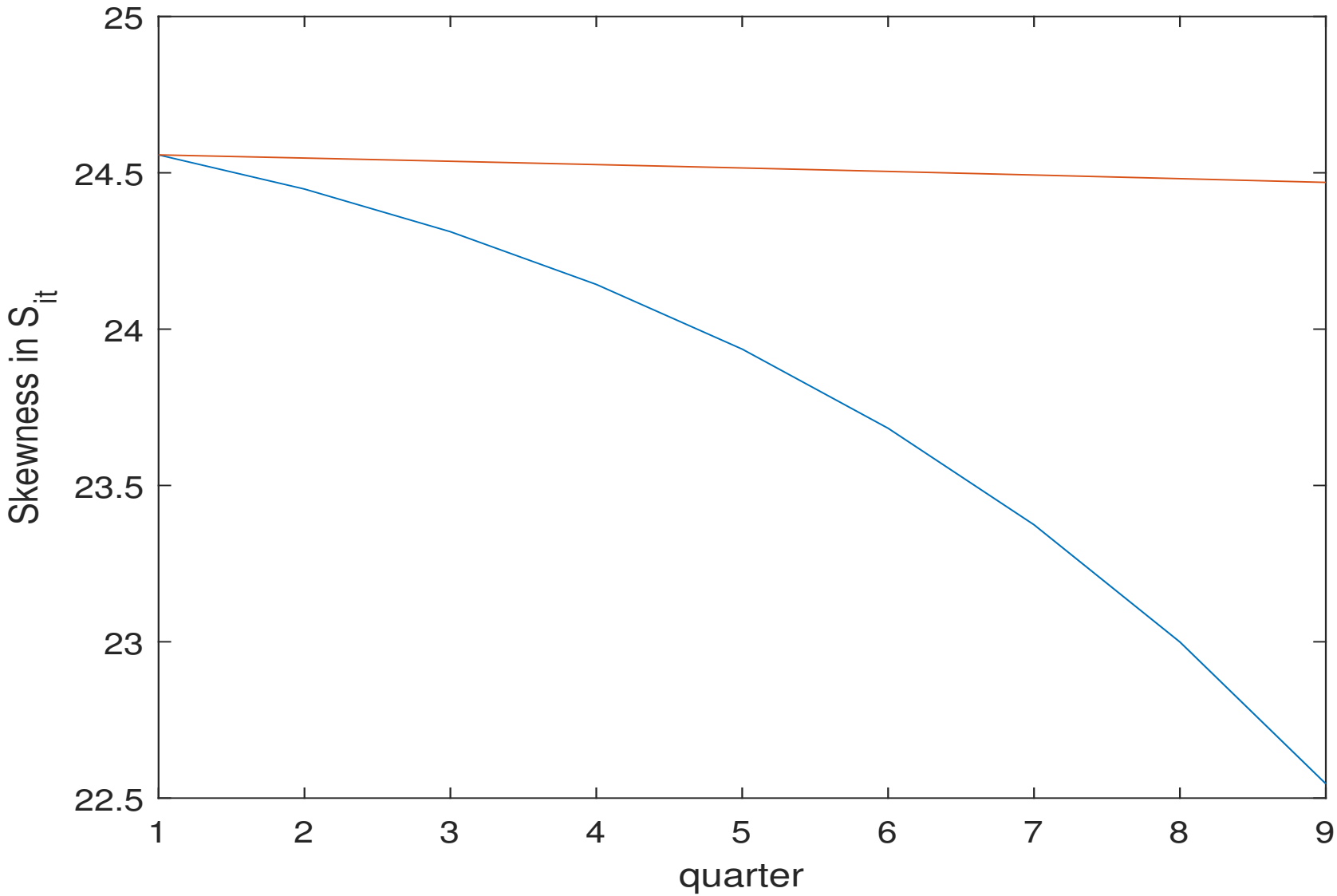


Simulation: f_{it} Kurtosis Less Than S_{it}



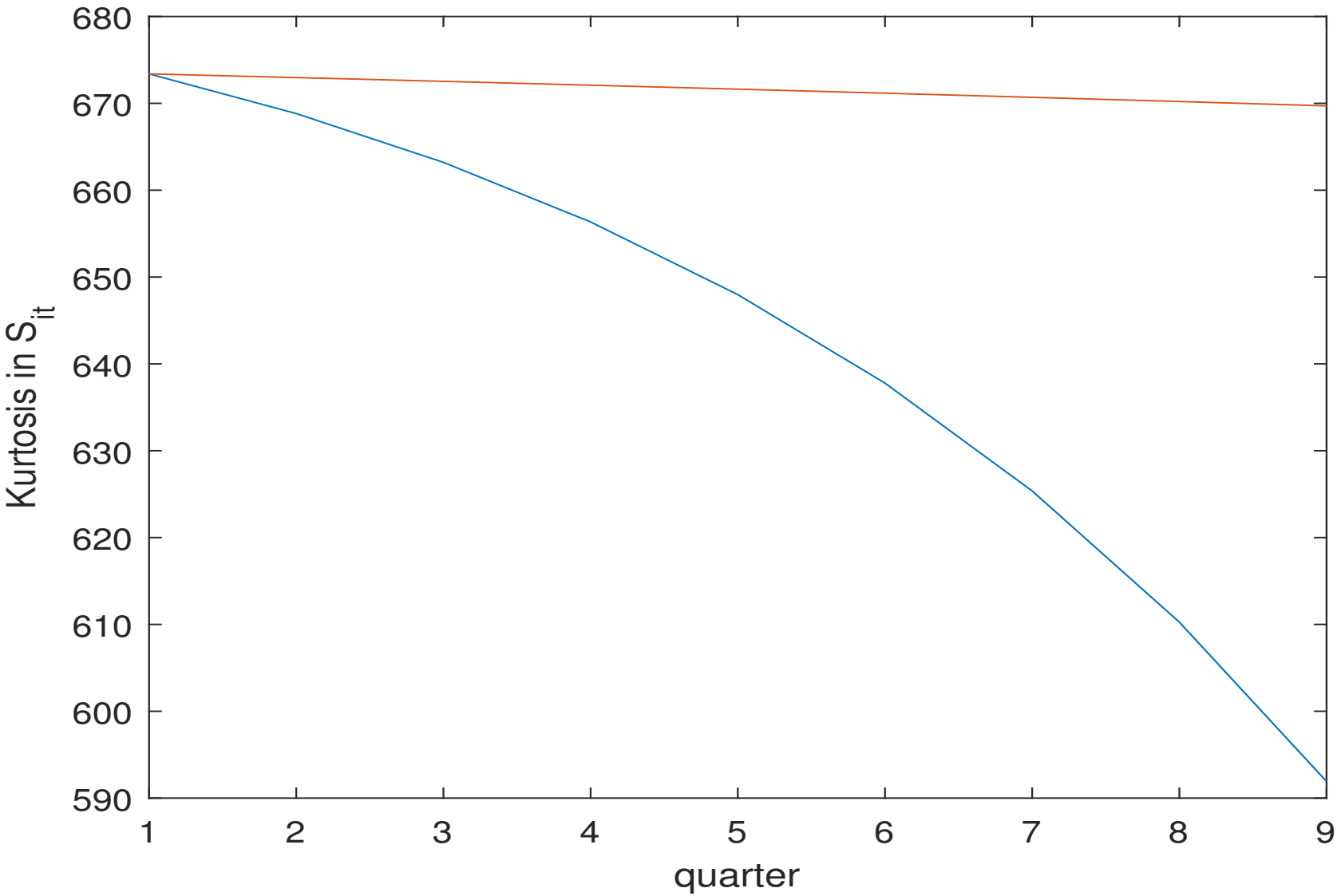
Simulation: S_{it} Skewness Decreases Faster When Distances are Shorter

Set $\tau_{ij} = 10 \times \tau_{ij}^{\text{baseline}}$



Simulation: S_{it} Kurtosis Decreases Faster When Distances are Shorter

Set $\tau_{ij} = 10 \times \tau_{ij}^{\text{baseline}}$



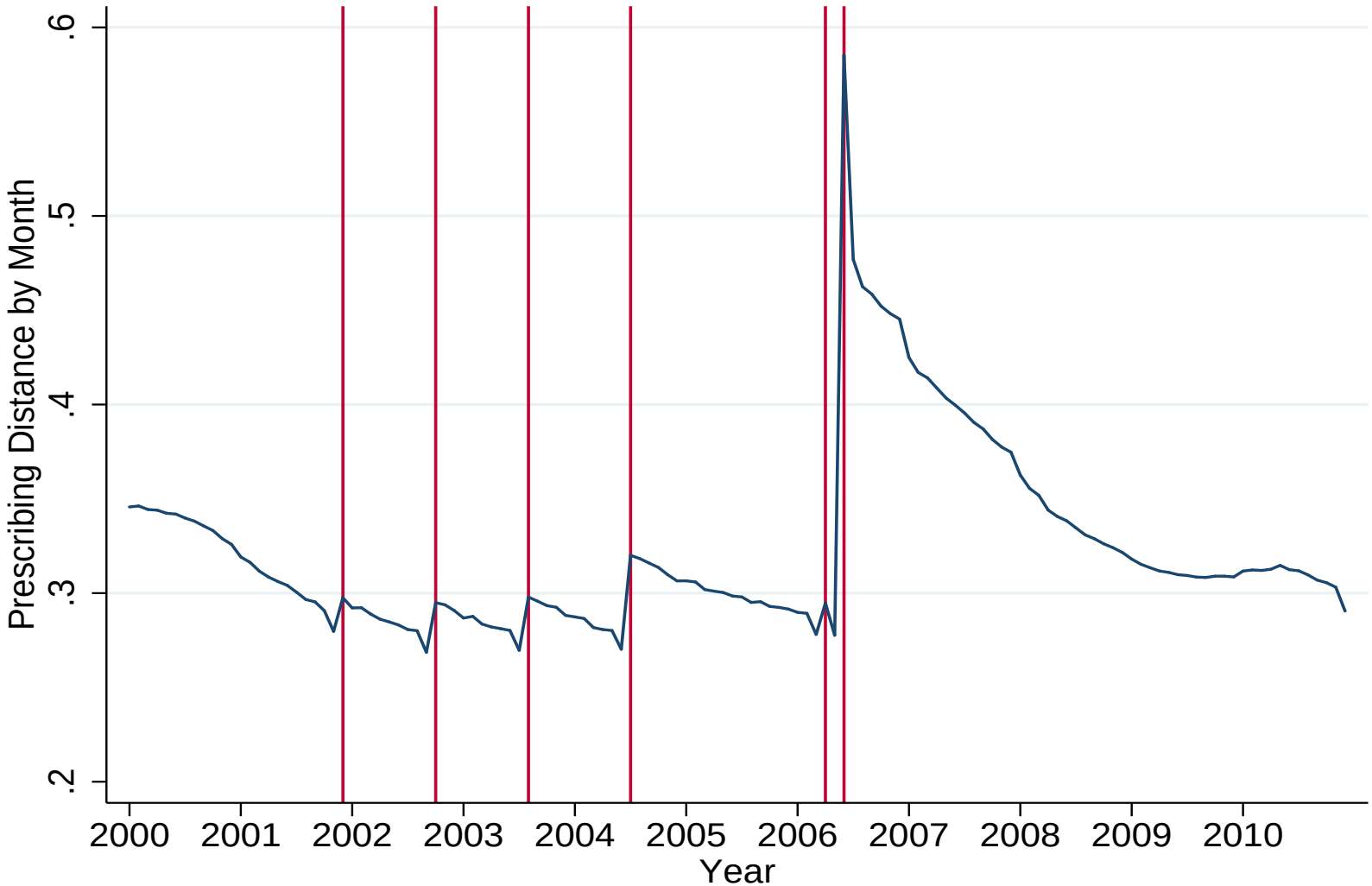
Conclusion

- Prescription shares converge over time, faster for central doctors
- Model-generated and estimated learning paths both indicate central doctors learn faster and that learning declines over time
- Many next steps:
  - Richer distance variables (cohort, medschool, specialty, facebook social connectedness of zipcode)
  - Learning from other doctors' beliefs
  - Risk aversion ($S_{idt} \neq S_{id't}$)
  - Endogenous investment in learning
  - Joint estimation of full model (not two step Bayesian learning and then network module), with more structural parameters (β_d, σ_d^2, CARA, etc.)
  - Alternative models of the network
  - Introduction of new drugs—across spell analysis

Appendix

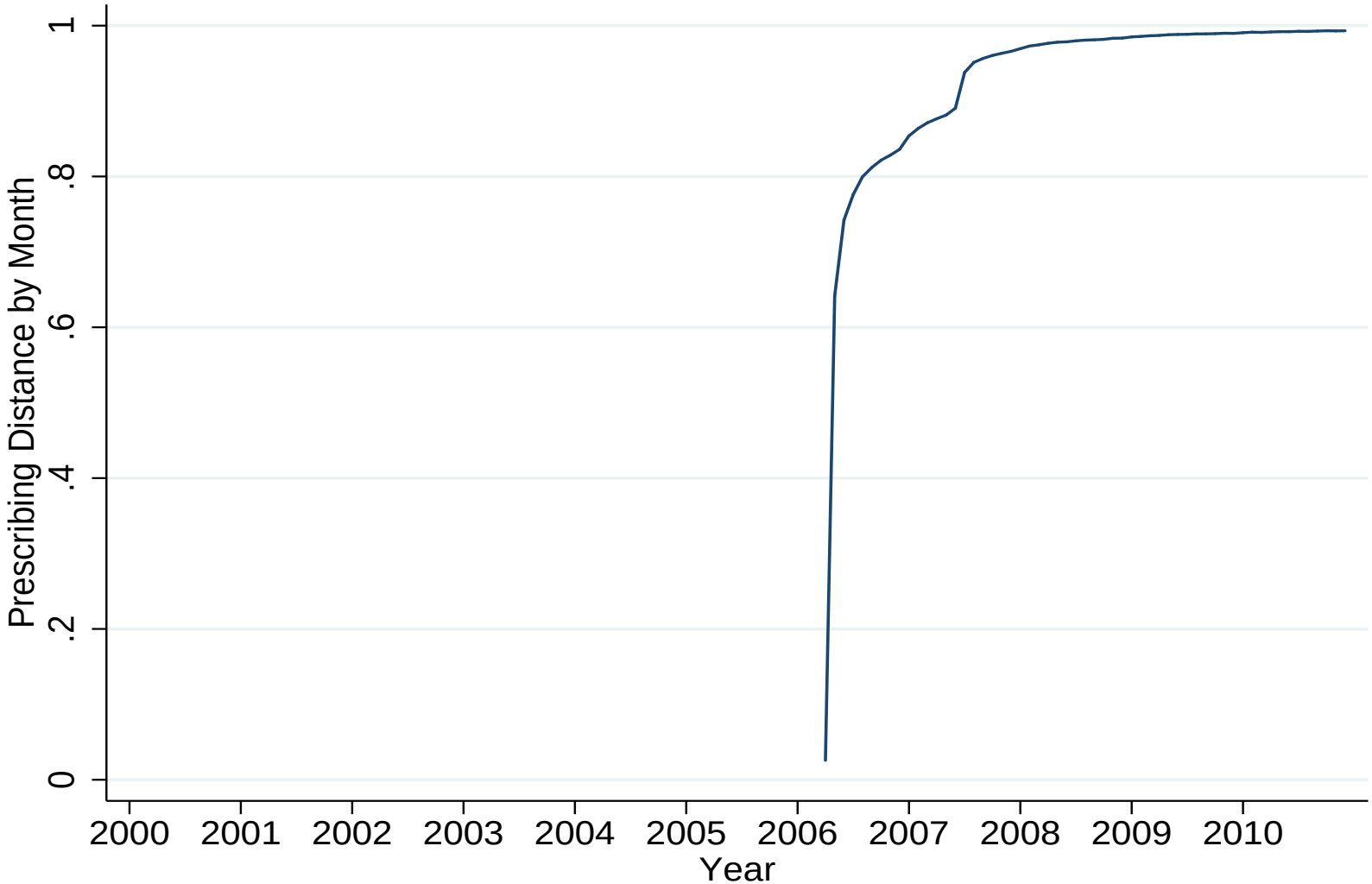
Scripting Patterns Converge across Doctors within Zipcode

Convergence within zipcode suggests not just composition effect



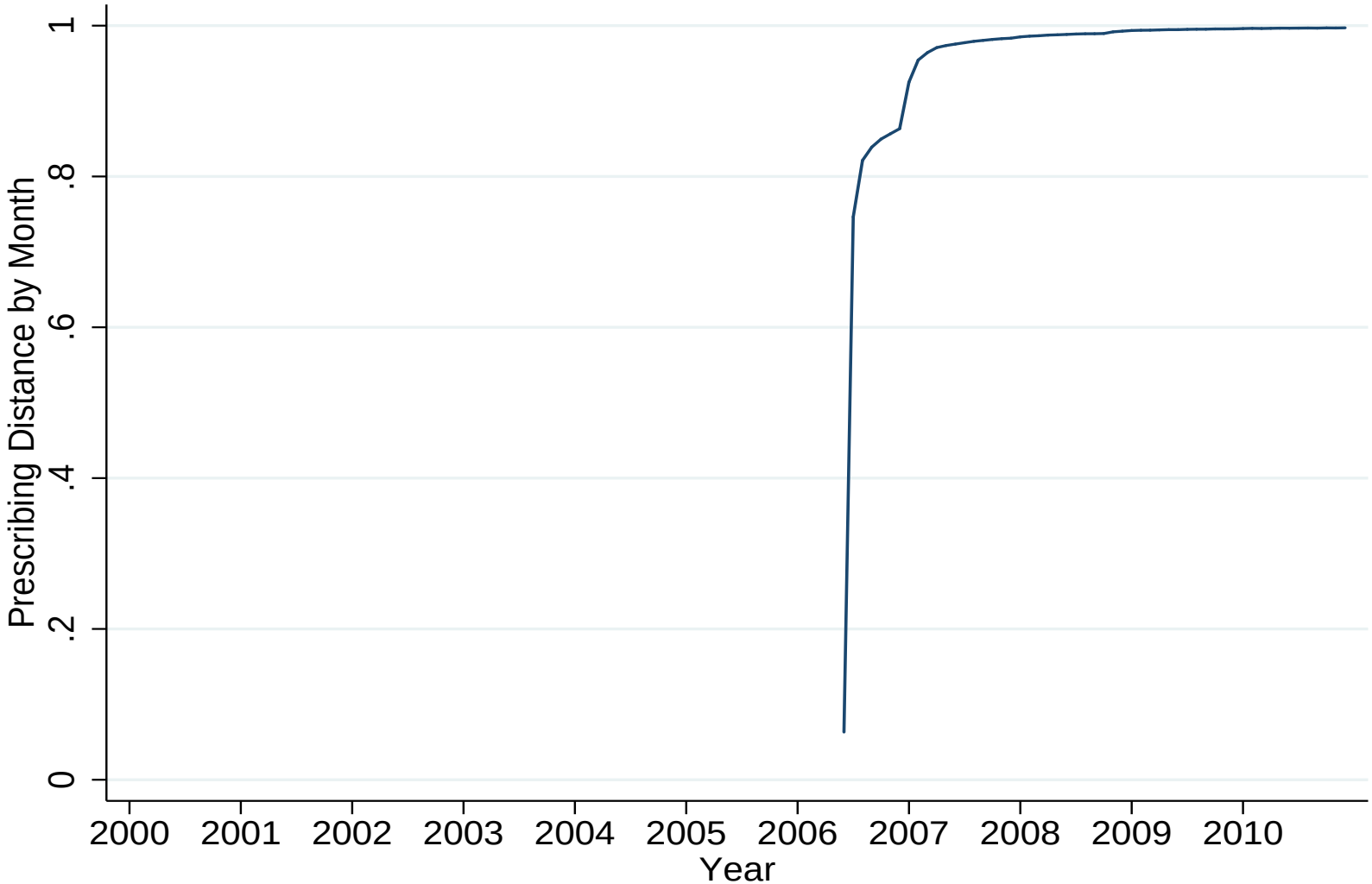
Generic Prescription Share Converges

$\longrightarrow$ Within-molecule generic substitution, Pravastatin



Generic Prescription Share Converges

$\longrightarrow$ Within-molecule generic substitution, Simvastatin



Estimates of Average Convergence Rates

- Convergence target $\overline{\pi}_{dT(t)}$ differs by spell, $\overline{\pi}_{dzT}$ by zipcode-spell
- Control for prescribing intensity and doctor FE

Dep variable:	Average Prescription Distance			
Drugs	All	Six	Six	Six
Target	U.S. $\overline{\pi}_{dT}$	U.S. $\overline{\pi}_{dT}$	U.S. $\overline{\pi}_{dT}$	Zipcode $\overline{\pi}_{dzT}$
Time (months)	-.0005^a	-.0005^a	-.0003^a	-.0006^a
	(3.64e-06)	(3.85e-06)	(2.73e-06)	(2.92e-06)
Rx Volume			-93.99^a	-92.09^a
			(.0628)	(.0670)
Doctor FE	N	N	Y	Y
Observations	17334636	17334636	16241174	16241174

Notes: *a* denotes 1% significance, *b* denotes 5% significance, *c* denotes 10% significance.

Estimates of Convergence Rates: Generic Substitution

Dep variable:	Average Generic Share		
Molecule	Lovastatin	Pravastatin	Simvastatin
Target	1	1	1
Time (months)	.0013[a]	.0051[a]	.0038[a]
	(1.79e-06)	(5.12e-06)	(3.68e-06)
Rx Volume	.5644[a]	15.05[a]	9.206[a]
	(.1252)	(.2340)	(.1443)
Doctor FE	Y	Y	Y
Observations	7300484	4683682	6171908

Notes: *a* denotes 1% significance, *b* denotes 5% significance, *c* denotes 10% significance.

- Control for prescribing intensity and doctor FE