Can Statistical Discrimination Explain Inequality?

Christian Hunkler

Frühjahrstagung der Sektion „Modellbildung und Simulation“
der Deutschen Gesellschaft für Soziologie
21.März 2013 in Konstanz

MAX PLANCK INSTITUTE
FOR SOCIAL LAW AND SOCIAL POLICY
mea Munich Center for the Economics of Aging

Research Questions, Motivation and Method

- **Research Questions**
  1. Can *Phelp's* model of statistical discrimination explain inequality in *hiring/allocation?*

     Labor market inequality is mainly caused by allocation to positions
     [e.g. Petersen & Morgan 1995]

  2. Is it the same outcome considering hiring for qualified positions?
  3. Does reduced mobility of (e.g. ethnic migrants) applicants change the outcome?

- **Motivation**
  - Debate in ZFS [Seibert & Solga 2005, 2006 vs. Kalter 2006a, 2006b]
  - Misleading "textbooks" [Cain 1986: 724 vs. England 1992; Kalter 2003]
  - „Rescue" a realistic, plausible, precise, simple, and formalized mechanism [Esser 1999, Hedström & Swedberg 1998,1996]

- **Method**
  - Dynamic micro simulations, implemented in Stata
  - → [Derive corollary hypotheses, to later on test with real data]

Phelp's Measurement Model of Statistical Discrimination: "The Statistical Theory of Racism and Sexism"

- **Hiring – Employers have incomplete information**
  - Investment Decision under uncertainty → interpretation of signals [Phelps 1972; Spence 1973: 355f.; Arrow 1973a, 1973b, 1998; …]

- **Measurement Model – „Phelps Model"** [Phelps 1972, Aigner & Cain 1977]
  - *y* : „Some indicator of skill", e.g. test scores, grades, all „signal" combinations
  - *q* : true skill level; productivity = ability = *job relevant* human capital)
  - α : group mean of q
  - γ : reliability of the test score *y*

- **Employer evaluates each applicant:**

Reliability [0;1] (how good is signal)

$$\hat{q} = E(q|y) = (1-\gamma)\cdot\alpha + \gamma\cdot y$$

Average Product. of Group

Signal (e.g. grade)

Textbooks: … cannot explain group inequality!

However

- **Cain 1986: 724**

Accordingly, for a given y-score [roughly corresponding to "holding X constant" in Model (I)], majority workers receive a higher wage than minority workers for y-scores above the mean, α, and lower wages for y-scores below the mean. Thus, group discrimination, defined by $E(w_1 - w_0) > 0$, is not present.

Allocation to positions

- **England & Lewin 1989: 243**

	Motivation	Creates Group Differences in Earnings
TASTE	Nonpecuniary individual gain in utility	Yes
STATISTICAL	Pecuniary individual gain through saving screening	No

- **Kalter 2006: 146**
  (quoting Cain 1986: 723)

Statistische Diskriminierung im engen Sinne dieser Modelle kann die Residualeffekte somit nicht erklären, solange die Leistungsunterschiede der Gruppen im Mittel nicht wirklich verschieden sind. Was so-

… cannot explain group inequality: True for wages, but not for hiring!

- **Wages**
  - Assume *correct* perception of same average productivity α (no pre-market differences)
  - Assume *non-distorted* perception of signal y (same grades…)
  - Then on average $y_W = y_B = \alpha_W = \alpha_B$
  - Then (some math omitted):

$$\hat{q} = E(q|y) = (1-\gamma)\cdot\alpha + \gamma\cdot y$$

$$wage_W - wage_B = (y-\alpha)(\gamma_W - \gamma_B)$$

- **Hiring – *Not everybody is hired!***
  - Based on the same assumptions
  - Hiring 1 applicant: worker 1 is hired → inequality
  - Hiring 2 applicants: workers 1, 3 hired → no inequality
  - Hire 3 applicants: workers 1, 3, 4 hired → inequality → **discrimination against W group!**

→ **It depends on relation of open positions to applicants**

Table 1: A simple four-worker example

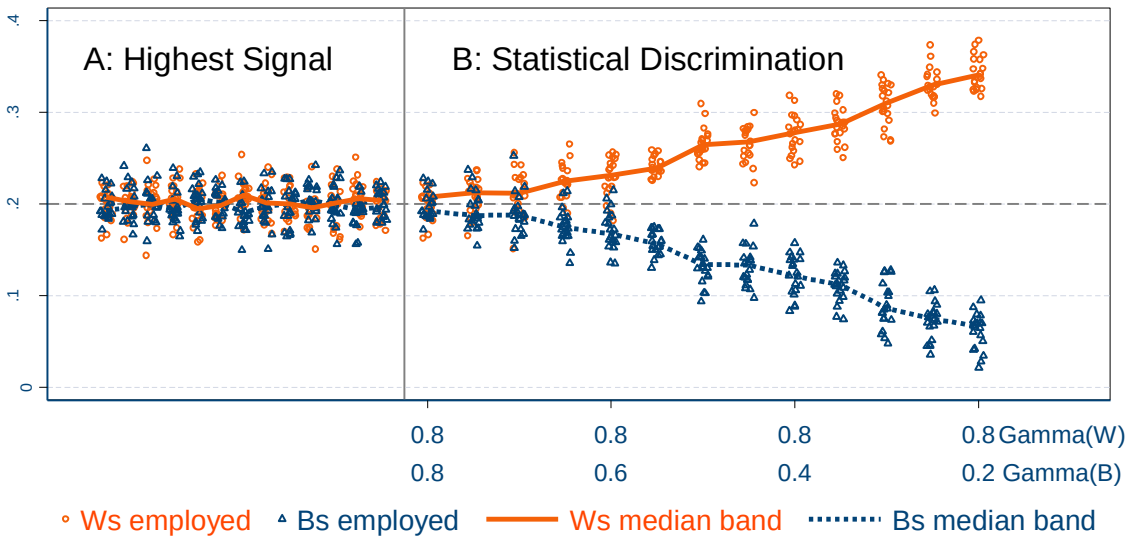
Worker	Group	„Raw“ Signal	**Statistical Disc.**	α	γ
1	W	60	**58**	50	0.8
2		40	**42**		
3	B	60	**54**	50	0.4
4		40	**46**		

Simulation Model

- **Hiring**
  - 100 employers try to hire each one applicant
  - Each of 500 applicants (50% Ws and 50% Bs) randomly apply to ten employers
  - Signals of productivity constant across groups (mean=50; Std. Dev.=10)
  - Sequence of hiring is randomly assigned
  - Once hired, all other applications of the respective applicant are deleted

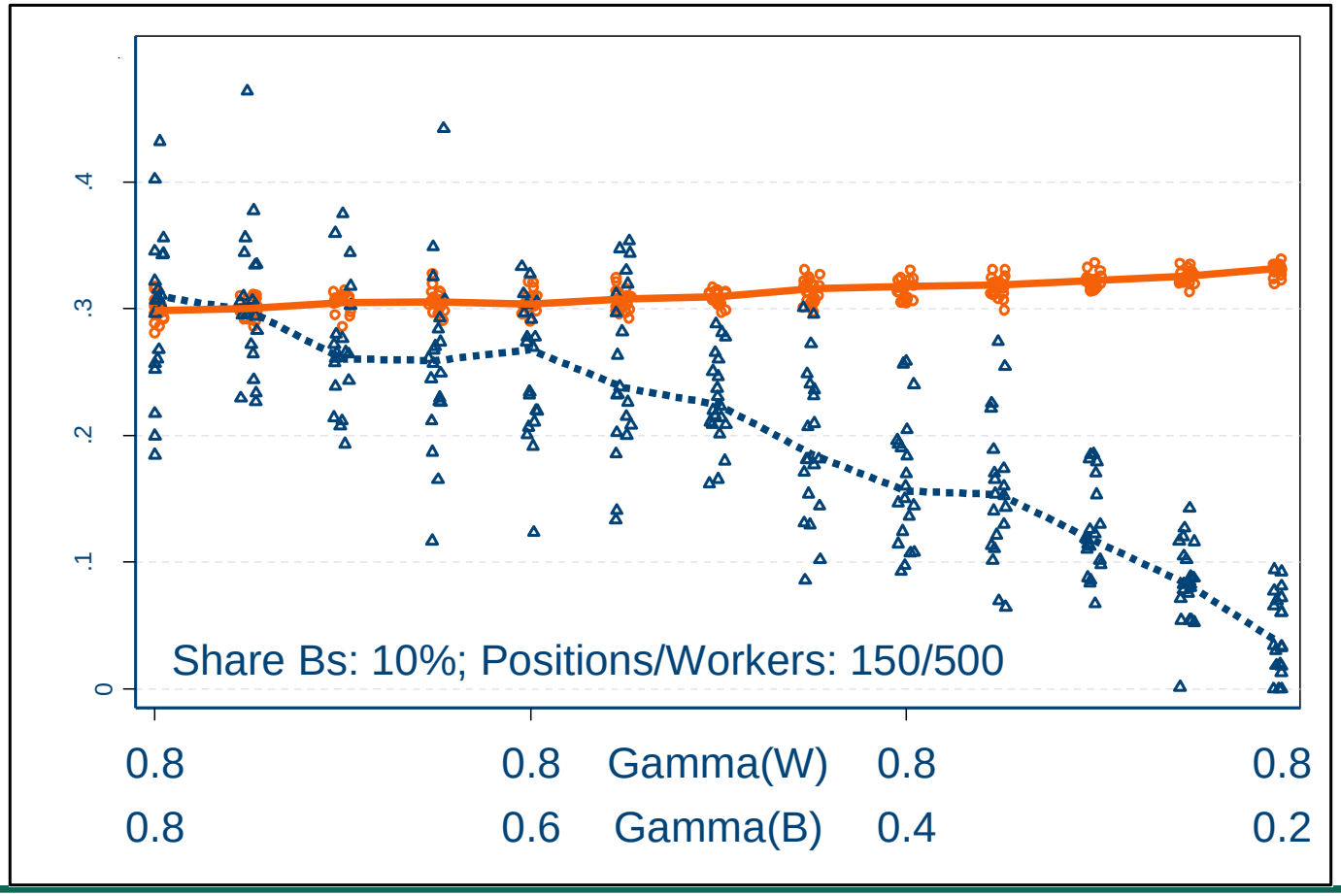
- **Apply two Selection routines**
  - Hiring according to highest signal
  - Statistical discrimination (as in formula above)

- **In all simulations**
  - Repeat each combination of parameters 20 times
  - Count who is hired on group level and calculate share
  - Plot shares from single runs and include median bands



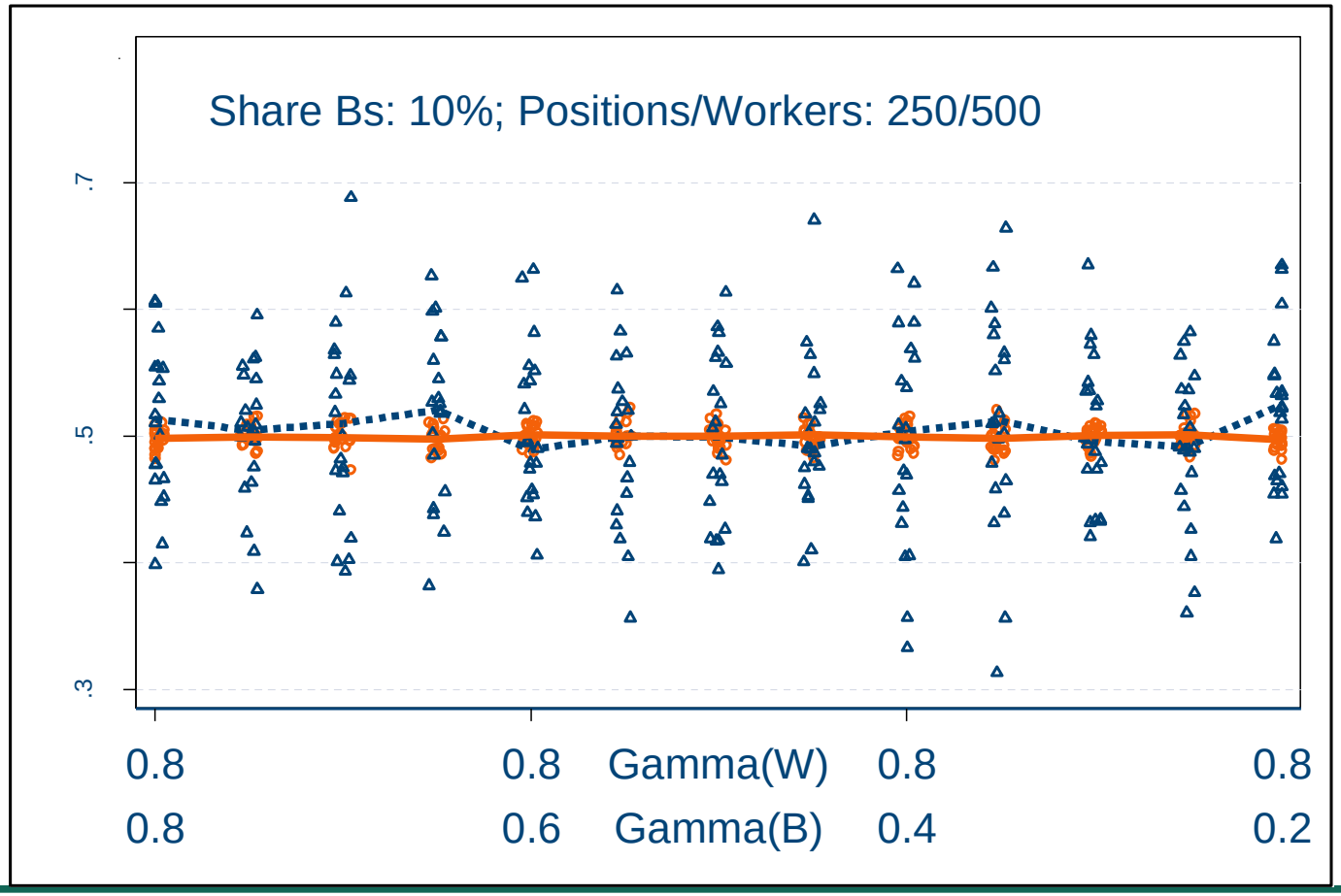
Does it depend on ratio of open positions to applicants?

- **Figure 2a:** **ratio** open positions to applicants **< 1:2** (150:500)
  - → group who is assigned lower reliability is discriminated



Does it depend on ratio of open positions to applicants?

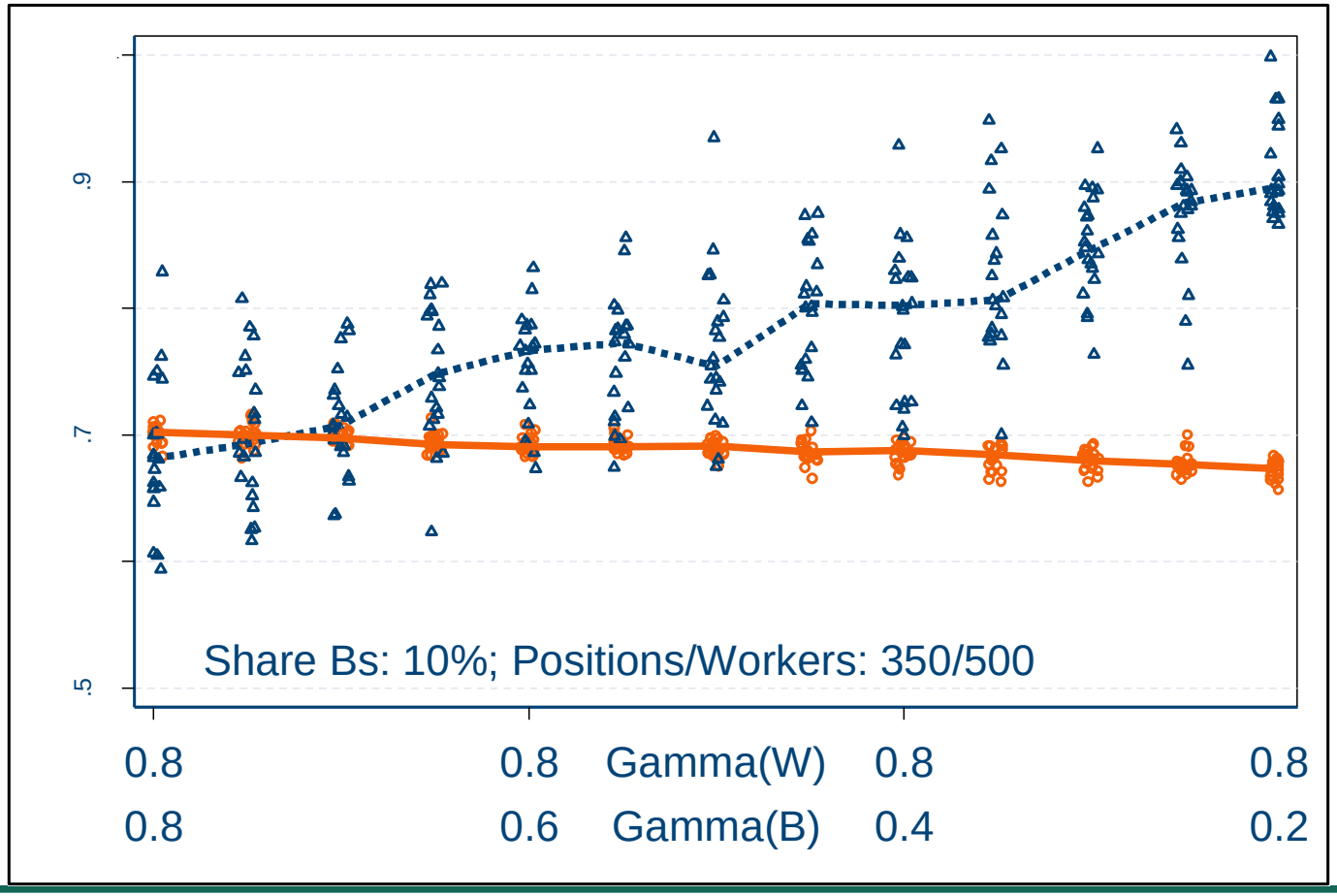
- **Figure 2b: ratio** open positions to applicants **= 1:2** (250:500)
  - → No discrimination



Does it depend on ratio of open positions to applicants?

MAX PLANCK INSTITUTE
FOR SOCIAL LAW AND SOCIAL POLICY
mea Munich Center for the Economics of Aging

- **Figure 2c:** **ratio** open positions to applicants **> 1:2** (350:500)
  - → Discrimination against majority!

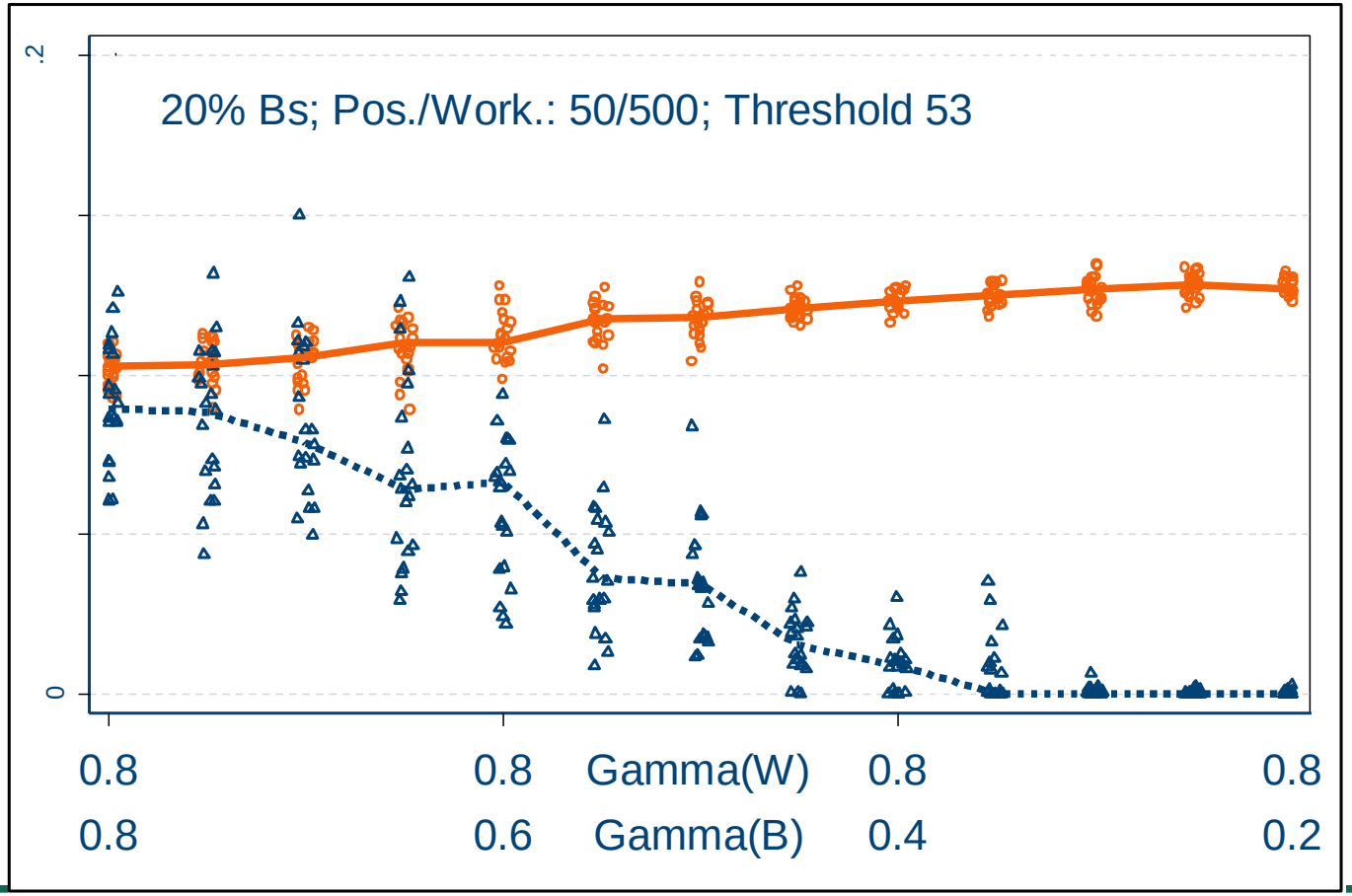


Access to qualified positions

- **Assumption:**
  - A qualified position, requires a productivity above the average (of all applicants)
  - Simulation using two different thresholds:
    - minimum of perceived productivity of 53 vs. 56

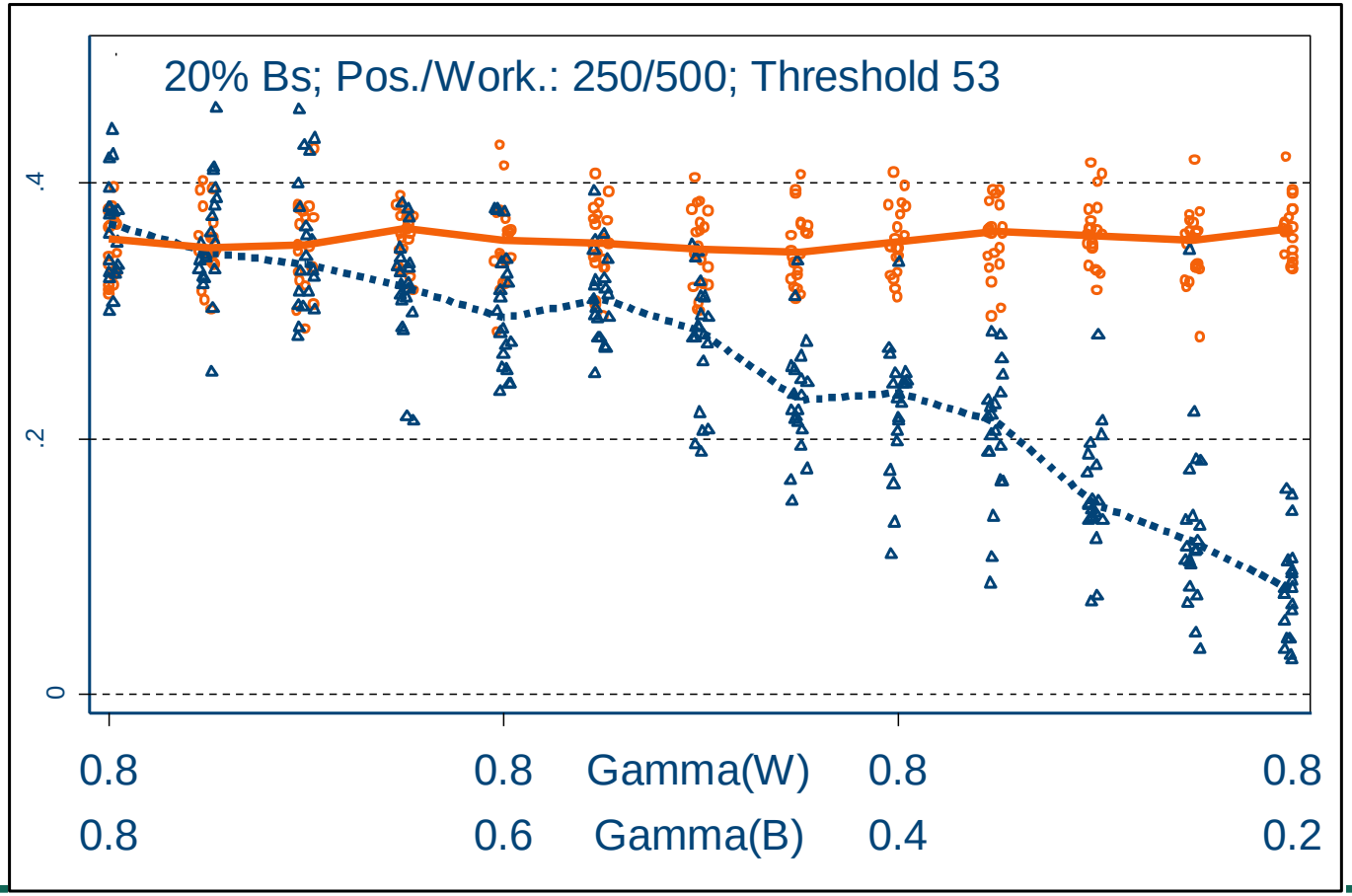
Access to qualified positions

- **Figure 3a: ratio** open positions to applicants **< 1:2** (50:500)
  - Discrimination against minority / lower reliability of signals
  - Even more pronounced, when threshold is higher (not shown)



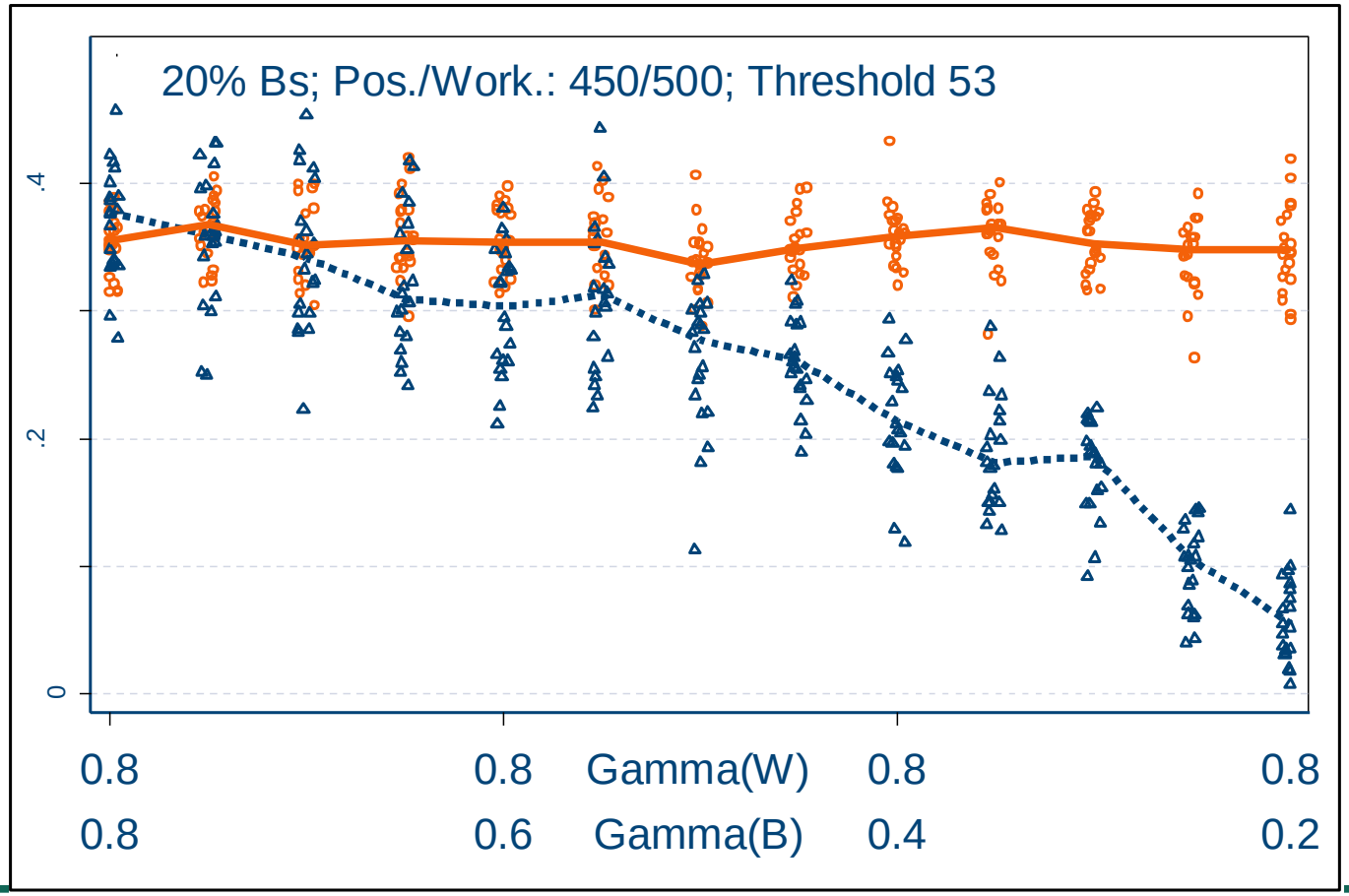
Access to qualified positions

- **Figure 3b: ratio** open positions to applicants **= 1:2** (250:500)
  - Again! Discrimination against Bs
  - Even more pronounced, when threshold is higher (not shown)



Access to qualified positions

- **Figure 3c: ratio** open positions to applicants **> 1:2** (450:500)
  - Again!! Discrimination against Bs
  - Even more pronounced, when threshold is higher (not shown)



Introducing Reduced Mobility of Applicants

- **What happens, if minority group is less informed about open positions, or less mobile?**
  - Or: What happens in a *perfect* vs. *less perfect* labor market, where all/some applicants apply to all/some open positions…

- **Translated into the Simulation Model**
  - Minority group B applies to less employers
  - [Before all workers applied to 10 random employers]
  - Now variation in number of applications: 10, 60, 110, 160 (maximum)

  - Actually, due to the randomization, in reality the numbers are lower:
    - 10 9.7
    - 60 50.2
    - 110 79.7
    - 160 101.3
  - Number of total open positions fixed at 160, 500 applicants, 20% Bs

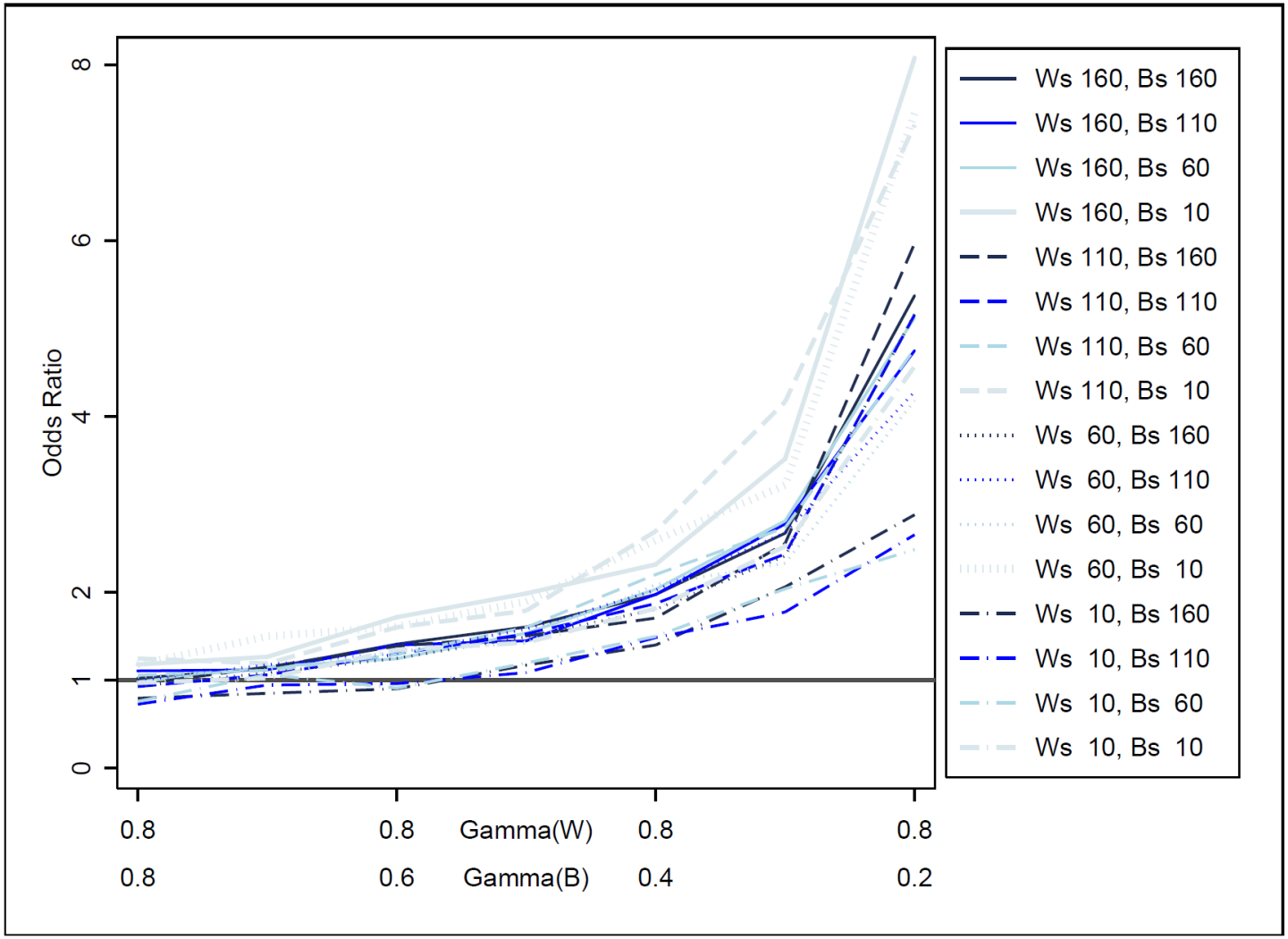
Introducing Reduced Mobility of Applicants

MAX PLANCK INSTITUTE
FOR SOCIAL LAW AND SOCIAL POLICY
mea Munich Center for the Economics of Aging

Odds Ration > 1
advantage for majority group/ group with higher gamma

Odds Ration < 1
advantage for minority group/ group with lower gamma

Figure 4: "Perfect" Labor Market Conditions and "Reduced Mobility"



Summary & Conclusions

1. Phelp‘s measurement model of statistical discrimination explains inequality in hiring.
   - The direction of discrimination mainly depends on the ratio of open positions to applicants,
   - and partly on the relative size of the minority group (not shown).

2. For access to qualified positions, the mechanism will always result in discrimination against the minority group.
   - More pronounced when minority group is smaller (not shown)

3. Reduced mobility substantially pronounces these effects.

→ As inequality is mainly attributed to allocation, not to within-job wage differentials, inequality research should take statistical discrimination into account.

→ But is it reasonable to stick with Phelp’s formula ?

Thank you!

Method: Simulation using Stata

- **Use Stata data matrix to store …**
  - employers and their parameters in columns;
  - applicants (signals and group W_{hites}, B_{lacks}) in lines
- **Use some random function to** (standard normal Gaussian)
  - assign productivity signals (mean=50, std=10) to applicants (L_1 … L_n)
  - assign them group membership afterwards
  - assign who does apply with which employer ("x")
- **Set for employers**
  - same average productivity
  - some meaningful variation in reliabilities γ

Table 2: Example Matrix of Data Setup

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				E_1	E_2	…	E_j	…	E_m
			α_W	50	50	…	50	…	50
			α_B	50	50	…	50	…	50
			γ_W	0.8	0.8	…	0.8	…	0.8
			γ_B	0.6	0.6	…	0.6	…	0.6
	Signal	Group							
L_1	50	W		x			x		
L_2	45	B			x				
…	…	…							
L_i	55	B		x					
…	…	…							x
L_n	50	W					x		

Simulation Model: Descriptives of the 260 runs in Figure 1 [Backup]

MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
mea Munich Center for the Economics of Aging

	Mean	Std. Dev.	Min.	Max.
A: Parameter Settings				
E - Number of Employers	100.0	0	100.0	100.0
γ_W	0.8	0.000	0.8	0.8
γ_B	0.5	0.187	0.2	0.8
L - Number of Workers	500.0	0	500.0	500.0
Share of B Workers	50.0	0	50.0	50.0
Number of Applications Ws	10.0	0	10.0	10.0
Number of Applications Bs	10.0	0	10.0	10.0
B: Simulated Values				
All Workers: Mean Signal	49.991	0.430	48.704	51.478
W Workers: Mean Signal	50.014	0.641	48.445	51.994
B Workers: Mean Signal	49.967	0.641	48.212	51.774
All Workers: Std. Dev. Signal	9.986	0.318	8.963	10.771
W Workers: Std. Dev. Signal	9.984	0.441	8.923	11.026
B Workers: Std. Dev. Signal	9.976	0.421	8.504	11.201

Higher threshold causes more inequality [Backup]

- In plots it is hard to see that higher threshold, causes more discrimination against Bs
- Here Odds-Ratios (show effects, regardless of absolute shares of hired workers)

Table 4: Simulation of Access to Qualified Positions

		Number of Employers and Threshold														
		—50—			—150—			—250—			—350—			—450—		
γ_W	γ_B	53		56	53		56	53		56	53		56	53		56
.8	.5	5.17	$<$	6.58	1.51	$<$	2.41	1.44	$<$	2.56	1.82	$<$	2.20	1.60	$<$	2.31
.8	.55	4.64	$>$	3.09	1.48	$<$	2.08	1.22	$<$	2.00	1.43	$<$	2.15	1.29	$<$	1.89
.8	.6	2.60	$<$	3.05	1.43	$<$	1.75	1.32	$<$	1.56	1.18	$<$	1.93	1.34	$<$	1.63
.8	.65	2.05	$>$	1.66	1.25	$<$	1.67	1.26	$<$	1.47	1.25	$<$	1.39	1.23	$<$	1.51
.8	.7	1.39	$<$	1.78	1.30	$>$	1.23	1.03	$<$	1.31	1.14	$<$	1.30	1.09	$<$	1.27
.8	.75	1.37	$=$	1.37	1.06	$<$	1.18	1.02	$<$	1.11	1.03	$<$	1.07	1.06	$<$	1.23
.8	.8	1.20	$>$	1.16	1.12	$>$	1.10	0.97	$<$	1.06	1.03	$<$	1.09	0.95	$<$	1.21

BACKUP

OLS-Regression of Share of B-Workers employed

VARIABLES	(1) Model_1	(2) Model_2
γ_W - γ_B	-0.233***	-0.230***
	(0.007)	(0.004)
W Workers: Mean Signal		-0.014***
		(0.001)
B Workers: Mean Signal		0.016***
		(0.001)
W Workers: Std. Dev. Signal		-0.011***
		(0.002)
B Workers: Std. Dev. Signal		0.009***
		(0.002)
Constant	0.209***	0.115
	(0.002)	(0.094)
Observations	260	260
R^2	0.826	0.932

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[BACKUP]

OLS-Regression of Share of W-Workers employed

VARIABLES	(1) Model_1	(2) Model_2
γ_W - γ_B	0.238***	0.235***
	(0.007)	(0.004)
W Workers: Mean Signal		0.014***
		(0.001)
B Workers: Mean Signal		-0.016***
		(0.001)
W Workers: Std. Dev. Signal		0.012***
		(0.002)
B Workers: Std. Dev. Signal		-0.008***
		(0.002)
Constant	0.190***	0.243**
	(0.002)	(0.096)
Observations	260	260
R^2	0.833	0.931

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

MAS-Simulations [Backup]

- **Multi-Agent-System (MAS)**

Reality *Makro 1*
- Population
- Labor Market
- Migration
- …

Simulation Model
- Agents
- Environment
- …

Theory
- Human Capital Th.
- Statistical Discr.
- …

?

Reality *Makro 2*
- Population
- Labor Market
- Inequality (?)

Reality vs. Simulation

How good is theory?
(How good is programming)

"Final State"
- E.g. estimated level of inequality