Discussion of “High Wage Workers Work for High Wage Firms” by Borovičková and Shimer

John M. Abowd
Associate Director for Research and Methodology and Chief Scientist,
U.S. Census Bureau

CHCS Annual Employment Conference:
Labor Supply Sorting and Wages
September 21-22, 2017

Acknowledgments and Disclaimer

- Parts of this discussion are based on forthcoming papers estimated using confidential data from the U.S. Census Bureau. All results have been reviewed to insure that no confidential data are disclosed.
- Parts of this discussion were supported by the National Science Foundation, the Sloan Foundation, and the U.S. Census Bureau (before and after my appointment started).

- The opinions expressed in this discussion are my own, not necessarily those of the U.S. Census Bureau or other research sponsors.

What I Like About This Paper

- Takes seriously the difference between the AKM statistical decomposition and the use of AKM to generate target quantities for structural estimation.
- Does a careful construction of a framework that is an alternative to AKM for generating summary models.
- Show a defensible sorting effect (similar to the one documented in Abowd, Kramarz, Lengermann, McKinney and Roux 2012 [IZA-JOLE open access]) by using an alternative form of smoothing (over random graph realizations).

But the AKM Bias Issue Is Overstated

- This statement is problematic:

  "It follows that the empirical average log wage is a noisy measure of a worker's or firm's type even with 36 years of data." (page 2)

- It is only true in the hypothetical world of infinite *N* asymptotics for finite population estimators.
- In the world we actually inhabit (finite *N* populations, near-universe samples), the statistical bias, even evaluated under worst-case hypotheses using Borovičková and Shimer's formulas, is much closer to zero when fit from the universe U.S. data.
- The difference between the correlation of estimated person and firm effects and the hypothetical correlation between true effects in their analysis is a model-based bias that depends upon the model, not the data.

The Reason Is Subtle

- The standard econometric asymptotic arguments do not distinguish between the realized population and the super-population from which it was selected.
- The super-population formulation reconciles the quantities produced in standard (read: statistical offices) design-based estimation with the view that any particular parameter could be sampled from a space for that parameter when the population under study is instantiated.
- They are Bayesian in origin, but deliver frequentist formulas by conditioning on the realized population (See Imbens and Rubin, 2015, for extensive use in the causal modeling framework).
- When you assume that the AKM estimand is the population conditional mean in the actual labor market under study, the finite-population correction to the formulas on page 35 (eq. 10) substantially reduces the statistical bias from assuming an infinite population.
- How one interprets the model-based bias depends on characteristics of the model, not the data.

AKM From AMZ (2018)

Observations (Person Year Job)	2,014,000,000	M	3	
Years	2004 to 2013	N	12	
Persons	200,700,000	N*(M-1)	24	
Firms	14,650,000	M*(N-1)	33	
Sampling rate (largest connected group)	99%	MN-M-N	21	
		1-Samping rate	1%	
Variances/Correlations				
	ln y	person	firm	AKM residual
ln_real_ann_earn	3.5970	0.4323	0.6225	0.5446
person effect	0.3836	0.3530	0.0155	-0.0004
firm effect	0.5206	0.0413	0.3975	-0.0015
AKM residual	0.3879	-0.0010	-0.0033	0.5479

Source: Abowd, McKinney and Zhao (JOLE 2018) [preprint]

Borovičková Shimer Bias Correction

Observations (Person Year Job)	2,014,000,000	M	3	
Years	2004 to 2013	N	12	
Persons	200,700,000	N*(M-1)	24	
Firms	14,650,000	M*(N-1)	33	
Sampling rate (largest connected group)	0%	MN-M-N	21	
		1-Samping rate	100%	
Variances/Correlations				
	ln y	person	firm	AKM residual
ln_real_ann_earn	3.5970	0.4323	0.6225	0.5446
person effect	0.3836	0.3530	0.0155	-0.0004
firm effect	0.5206	0.0413	0.3975	-0.0015
AKM residual	0.3879	-0.0010	-0.0033	0.5479
	ln y	person	firm	AKM residual
ln_real_ann_earn				
person effect		0.2087	-0.0261	
firm effect		0.1917	0.0717	
AKM residual				

Finite Population Bias Correction

Observations (Person Year Job)	2,014,000,000	M	3	
Years	2004 to 2013	N	12	
Persons	200,700,000	N*(M-1)	24	
Firms	14,650,000	M*(N-1)	33	
Sampling rate (largest connected group)	99%	MN-M-N	21	
		1-Samping rate	1%	
Variances/Correlations				
	ln y	person	firm	AKM residual
ln_real_ann_earn	3.5970	0.4323	0.6225	0.5446
person effect	0.3836	0.3530	0.0155	-0.0004
firm effect	0.5206	0.0413	0.3975	-0.0015
AKM residual	0.3879	-0.0010	-0.0033	0.5479
	ln y	person	firm	AKM residual
ln_real_ann_earn				
person effect		0.0021	-0.0003	
firm effect		0.0421	0.0007	
AKM residual				

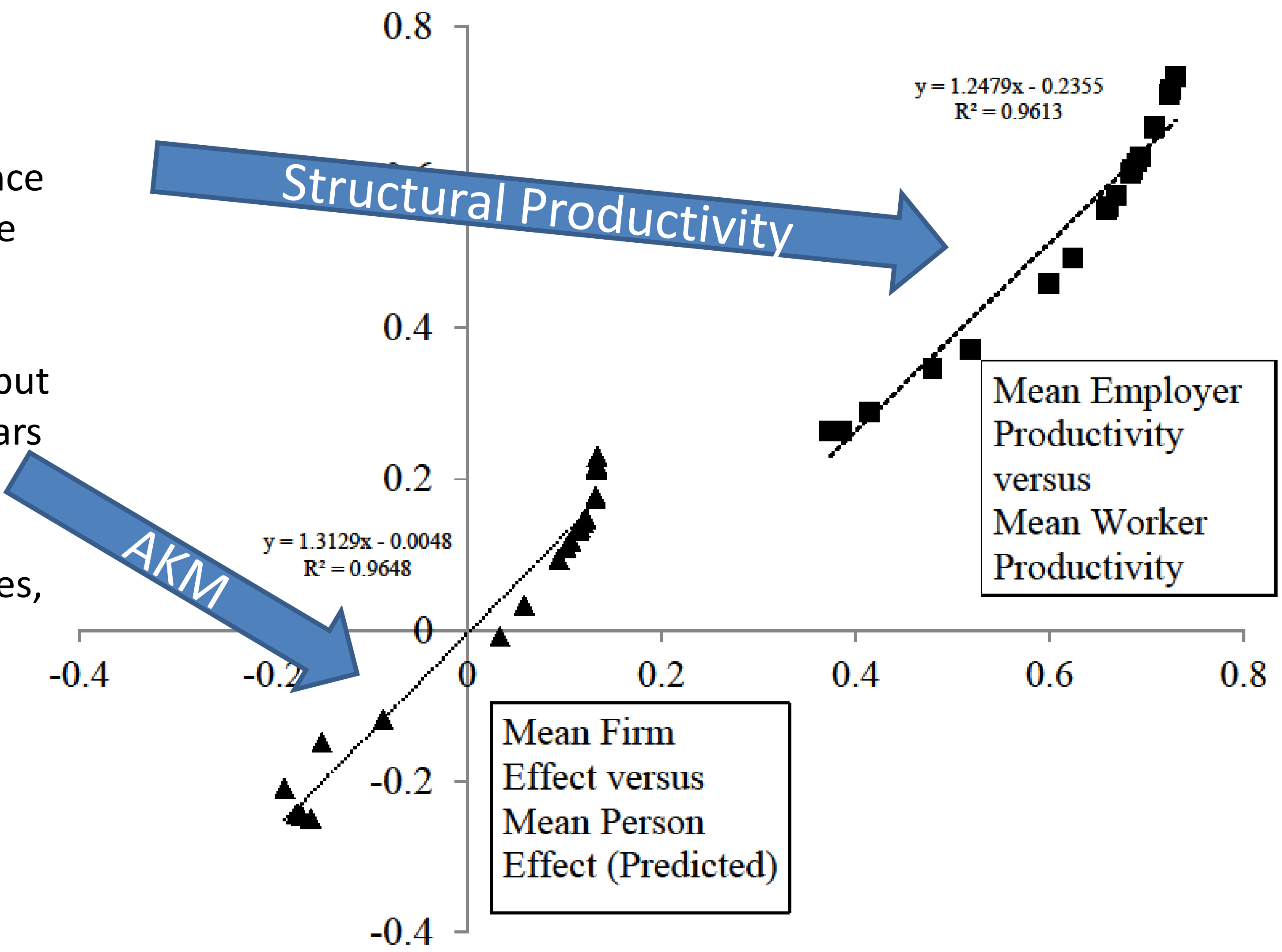
But, …

- Francis Kramarz and I actually agree with the authors' point that this covariance (or the associated correlation) is not a useful indicator of the assortative matching in the labor market without further structural assumptions
  - This point was actually in the original AKM working paper (http://www.nber.org/papers/w4917.pdf, pages 5-11)
  - The final version excluded it at the editor's recommendation (we've all been there)
- We made the point again in our *Handbook of Labor Economics* paper (Abowd and Kramarz, 1999, [preprint, pages 29-36])
- And again in our *Labour Economics* paper, 1999, [preprint, page 18]

Then, We Also Modeled It

- The Abowd, Kramarz, Lengermann, Perez-Duarte (2004) paper cited in the text is the first version of a paper that has been in wide circulation for a long time
- It became Abowd, Kramarz, Perez-Duarte, and Schmutte (forthcoming *Annals of Economics and Statistics*) [2009 preprint; 2014 preprint; 2017 final preprint]
  - Uses the AKM decomposition to generate empirical moments that are then fit to structural productivity heterogeneity with the Shimer (2005) model

Basically the same structural result as Borovičková and Shimer, which is not surprising since the theoretical models are very similar. This figure is estimated using Shimer 2005 as the base model, but the correlation also appears in the AKM estimates because they have been smoothed across industries, which is one way to implement structural mobility assumptions.



Source: Figure 9, Abowd, Kramarz, Perez-Duarte, and Schmutte (2018) "Sorting Between and Within Industries: A Testable Model of Assortative Matching, Annals of Economics and Statistics (forthcoming), [final preprint]

And Endogenous Mobility Matters Too

- Different random graph models have much different implications
- Originally modeled in two separate papers
  - Abowd, McKinney and Schmutte [2010 conference version preprint]
  - Abowd and Schmutte [2013 conference version preprint]
- Final paper Abowd, McKinney and Schmutte (forthcoming *Journal of Business and Economic Statistics*) [2015 preprint; advance final]
- Estimates presented in that paper use the full super-population framework to fit underlying worker, firm and match types with endogenous mobility
- Borovičková and Shimer is an alternative random-graph model

Table OA8: Correlation Matrix of Earnings Equation Parameters: Extended Work Histories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	$\ln w$	$X\beta_{AKM}$	θ_{AKM}	ψ_{AKM}	μ_{AKM}	ε_{AKM}	$X\beta_{Gibbs}$	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
$\ln w$	1.00										
$X\beta_{AKM}$	0.44	1.00									
θ_{AKM}	0.39	−0.49	1.00								
ψ_{AKM}	0.50	0.07	0.17	1.00							
μ_{AKM}	0.34	0.03	0.00	−0.00	1.00						
ε_{AKM}	0.20	−0.02	0.00	0.00	−0.00	1.00					
$X\beta_{Gibbs}$	0.78	0.56	0.25	0.24	0.04	−0.02	1.00				
θ_{Gibbs}	0.50	0.14	0.38	0.27	0.00	0.00	0.25	1.00			
ψ_{Gibbs}	0.27	0.02	0.12	0.42	0.11	0.00	0.10	0.04	1.00		
μ_{Gibbs}	0.06	0.05	−0.05	−0.10	0.28	0.00	−0.00	−0.23	−0.74	1.00	
ε_{Gibbs}	0.27	0.00	0.02	0.08	0.17	0.78	0.00	0.00	0.00	0.00	1.00

Thank you.

john.maron.abowd@census.gov