THE POWER OF PREDICTION:

PREDICTIVE ANALYTICS, WORKPLACE COMPLEMENTS, AND BUSINESS PERFORMANCE

Erik Brynjolfsson, Wang Jin, and Kristina McElheran

NBER Summer Institute

July 23, 2021

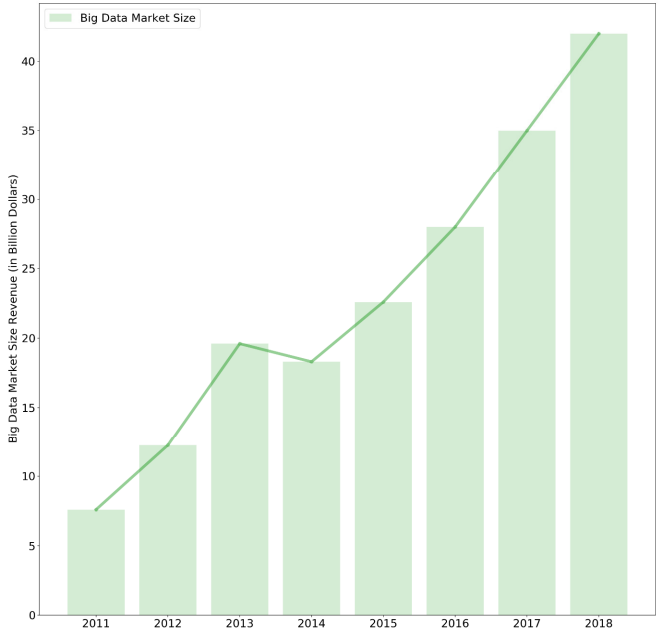
UNIVERSITY OF TORONTO

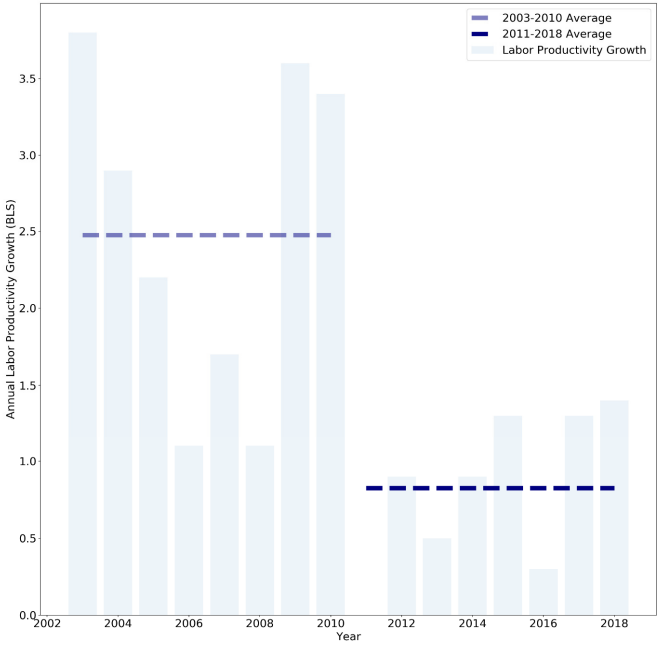
FEDERAL STATISTICAL RESEARCH DATA CENTERS DISCLAIMER

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

All results have been reviewed to ensure that no confidential information is disclosed.

MOTIVATION:
THE PREDICTIVE ANALYTICS “PAYOFF PARADOX”





UNIVERSITY OF TORONTO

RESEARCH QUESTION(S)

What can a rigorous empirical approach to studying predictive analytics tell us about:

1. **Prevalence**
2. **Performance Implications**
3. **Complements & Heterogeneity**

CONTRIBUTION

- **High firm investment yet low returns** (Syverson 2017; Brynjolfsson, Rock, & Syverson 2021)
- **Rising concentration, "superstar firms," and inequality potentially linked to IT** use (Autor et al. 2020; Lashkari et al. 2020, Barth et al. 2020)
- **Complementarities in the Digital Age**
  - General-Purpose IT (Black and Lynch 1996 & 2001; BBH 2002; FGG 2012; Tambe et al. 2012)
  - Applications (e.g., Dranove et al. 2014)
  - Data and Data-Centered Management Practices (Aral et al. 2012; Saunders &Tambe 2013; Brynjolfsson and McElheran 2019)
  - Role of Intangibles (Brynjolfsson et al. 2002; CHS 2005, 2009; Haskel & Westlake 2018)
- **This paper:**
  - **Novel measures (tech and "intangibles")**
  - **Causality**
  - **Importance of "Humans in the Loop"**

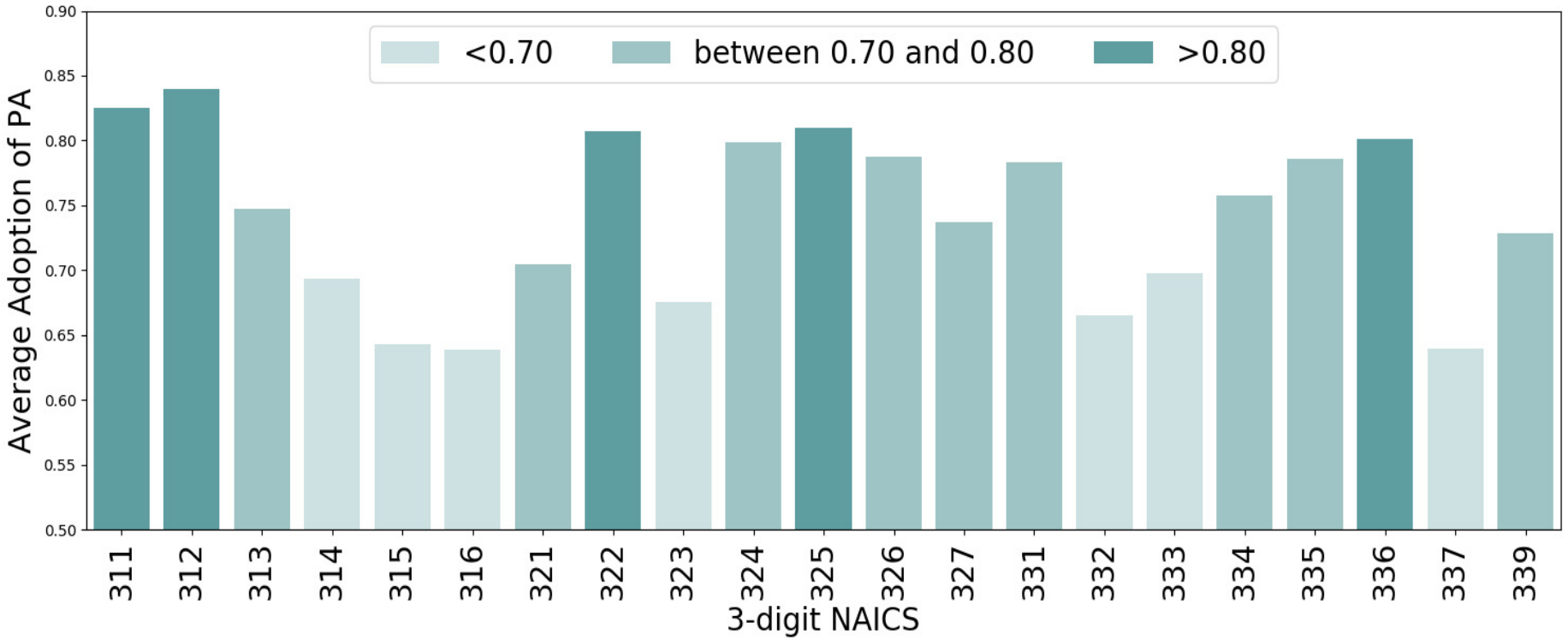
WHAT IS “PREDICTIVE ANALYTICS”?

1. A set of analytic techniques (e.g., data mining, statistical modeling, machine learning and AI)
2. That analyze historical and current data to make predictions

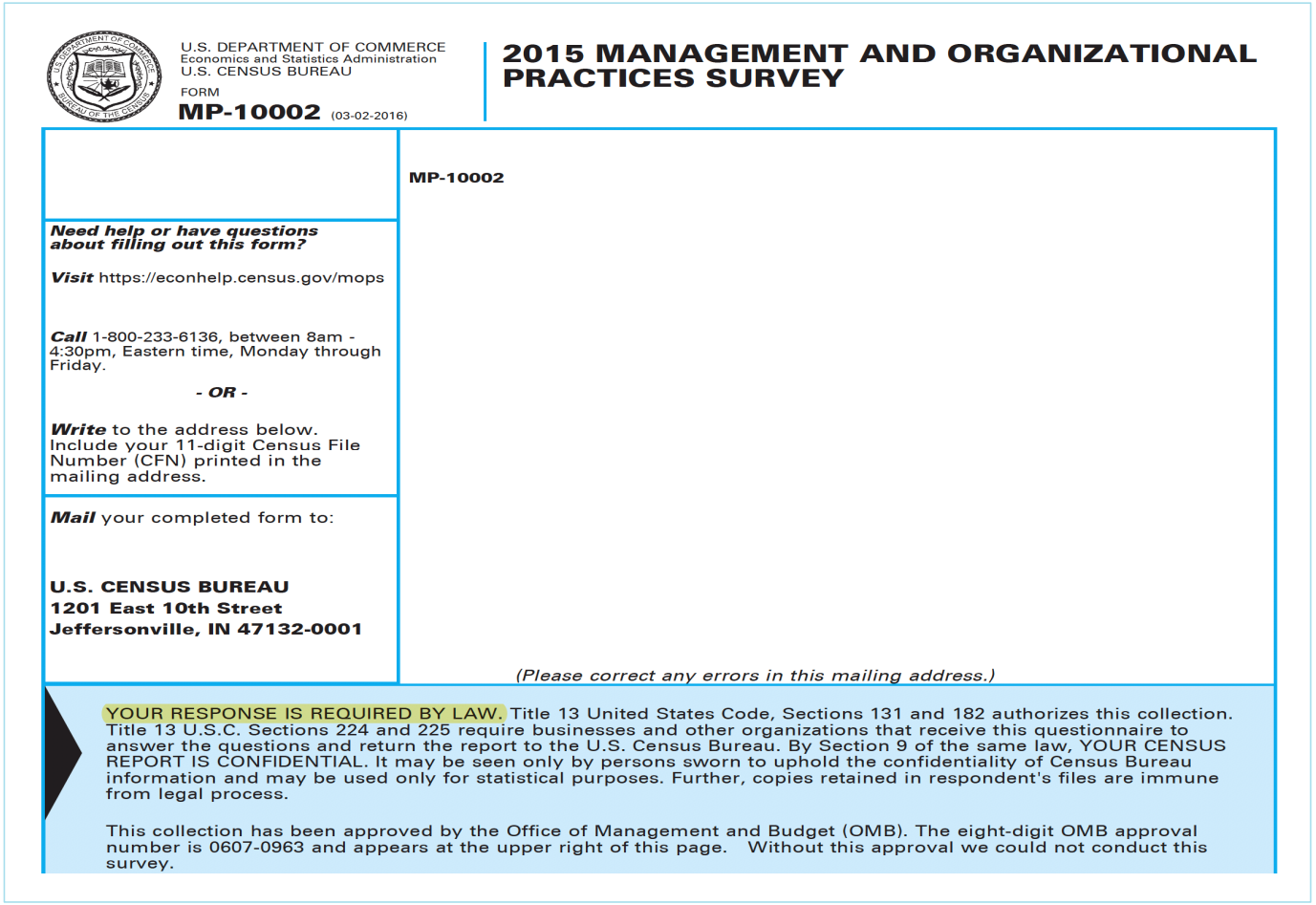
Conceptually: *automation / augmentation of cognitive work*

BIG INVESTMENT IN PREDICTIVE ANALYTICS

- Rise of data-driven management practices in the early 2000s (Brynjolfsson and McElheran 2016)
- $189 Billion in 2015 ⬆ to $274 Billion in 2022 worldwide (IDC)
- Manufacturing industries (among top adopters) – *see also Zolas et al. (2020)*



OUR APPROACH:
PARTNER WITH CENSUS FOR DATA COLLECTION



U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

FORM
MP-10002 (03-02-2016)

2015 MANAGEMENT AND ORGANIZATIONAL PRACTICES SURVEY

MP-10002

Need help or have questions about filling out this form?

Visit https://econhelp.census.gov/mops

Call 1-800-233-6136, between 8am - 4:30pm, Eastern time, Monday through Friday.

- OR -

Write to the address below. Include your 11-digit Census File Number (CFN) printed in the mailing address.

Mail your completed form to:

U.S. CENSUS BUREAU
1201 East 10th Street
Jeffersonville, IN 47132-0001

(Please correct any errors in this mailing address.)

YOUR RESPONSE IS REQUIRED BY LAW. Title 13 United States Code, Sections 131 and 182 authorizes this collection. Title 13 U.S.C. Sections 224 and 225 require businesses and other organizations that receive this questionnaire to answer the questions and return the report to the U.S. Census Bureau. By Section 9 of the same law, YOUR CENSUS REPORT IS CONFIDENTIAL. It may be seen only by persons sworn to uphold the confidentiality of Census Bureau information and may be used only for statistical purposes. Further, copies retained in respondent's files are immune from legal process.

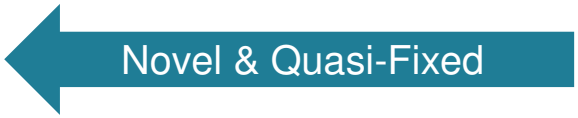
This collection has been approved by the Office of Management and Budget (OMB). The eight-digit OMB approval number is 0607-0963 and appears at the upper right of this page. Without this approval we could not conduct this survey.

UNIVERSITY OF TORONTO

PREVIEW OF FINDINGS

1. **Predictive analytics associated with much better performance**
   - Productivity gains of 1-3% ($464,000 – $918,000 per plant controlling for inputs)
     - Statistically and economically significant
     - Results support a **causal** interpretation
2. **Four complements determine gains from predictive analytics**
   - IT capital
   - College-educated labor

   - Flow-efficient production process



   - High Managerial Capacity
     & managerial headcount (*pending*)

EMPIRICAL APPROACH

1. **Devise new measures (cognitive testing and validation with Census)**
2. **Link to large-scale administrative data**
3. **Estimate firm performance (multi-factor productivity)**
4. **Explore causality**
   - Timing
   - Instrumental Variables (**government-mandated data collection**)

EMPIRICAL APPROACH, CON'T

5. **Explore Workplace Complements** (Milgrom and Roberts 1990 & 1995; Kandel and Lazear 1992; Holmstrom and Milgrom 1994; Athey and Stern 1998; Brynjolfsson and Milgrom 2013, etc.)
   - **Correlation Test**
   - **Performance Test**

EMPIRICAL STRATEGY (PERFORMANCE)

- **Revenue-Based Production Function:**
- $logsales_{ijt} = \alpha + \beta_{pa} PA_{ijt} + \beta_k K_{ijt} + \beta_l L_{ijt} + \beta_x X_{ijt} + \varepsilon_{ijt}$ (1)
  - Establishment *i* in industry *j* at time *t*
  - *Log sales*: total value of shipments
  - *PA*: adoption (frequency) of predictive analytics
  - *K*: non-IT and IT capital stock
  - *L*: total number of employees
  - Other inputs *X*: cost of materials and energy, known productivity shifters (e.g., MU status)
  - Control for data-driven decision making (Brynjolfsson and McElheran 2019) and "structured management" (Bloom et al. 2019)

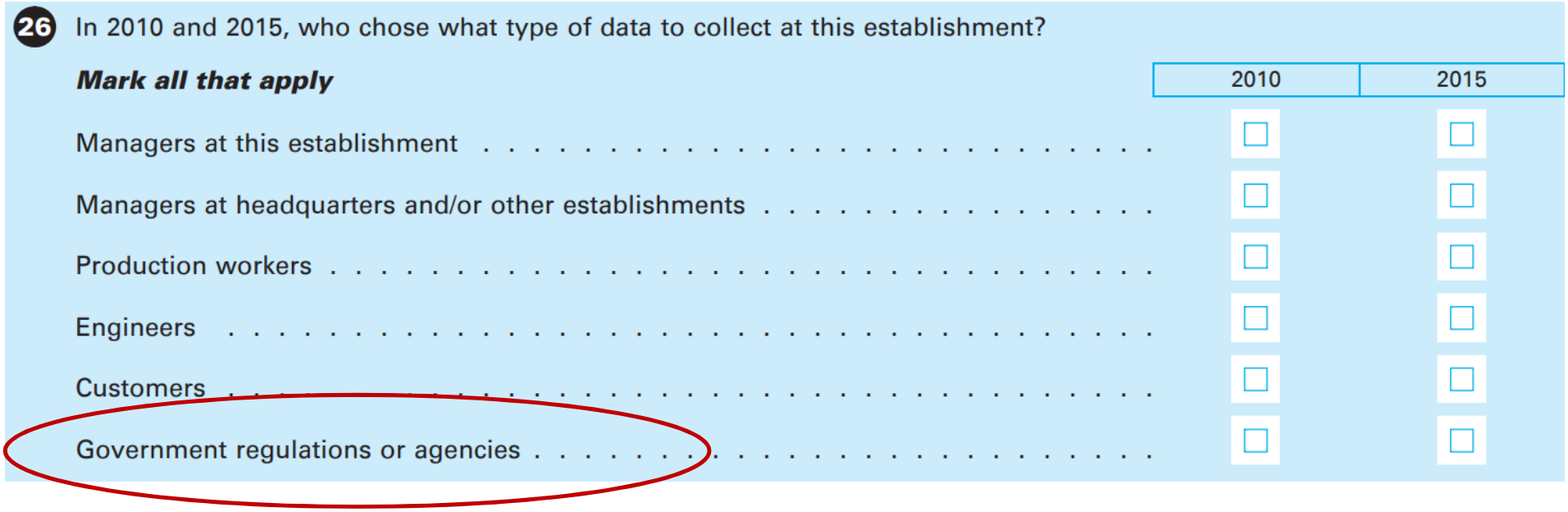
DATA

Identifying variation is largely cross-sectional

Variables	Definition	Sample	2010	2015
PA Adoption	Indicator for plants that adopted predictive analytics	0.74 (0.44)	0.73 (0.44)	0.80 (0.40)
PA Use Frequency	Index for highest frequency of PA use (1=Yearly, 2=Monthly, 3=Weekly, 4=Daily)	1.12 (1.06)	1.09 (1.05)	1.27 (1.12)
Log sales	Total value of shipment in log ($Thousands)	10.37 (1.52)	10.68 (1.39)	10.86 (1.37)
Log TE	Total number of employees in log	4.56 (1.17)	4.79 (1.09)	4.88 (1.09)
Log IT Capital Stock	IT capital stock in log terms ($Thousands)	5.16 (2.41)	5.58 (2.25)	5.62 (2.18)
Worker Education	Percentage of employees with a bachelor's degree	0.15 (0.14)	0.15 (0.13)	0.16 (0.14)
Flow-Efficient Production	Continuous Flow or Cellular (vs. Batch or Job-Shop) Production	0.63 (0.17)	0.60 (0.16))	0.68 (0.15)
High Managerial Capacity	Indicator for plants that tracked more than 10 KPIs (*temporary*)	0.44 (0.50)	0.37 (0.48)	0.56 (0.50)
Mandated Data Collection	Indicator for plants that are required to collect data by government regulations or agencies	0.25 (0.43)	N/A	N/A
Sample		Baseline MOPS	Balanced 2010	Balanced 2015
N		51,000	18,000	18,000

30,000+ unique plants

GOVERNMENT MANDATE TO COLLECT DATA



26 In 2010 and 2015, who chose what type of data to collect at this establishment?

Mark all that apply

	2010	2015
Managers at this establishment	☐	☐
Managers at headquarters and/or other establishments	☐	☐
Production workers	☐	☐
Engineers	☐	☐
Customers	☐	☐
Government regulations or agencies	☐	☐

- **High plant-level variation (even conditioning on industry, productivity, other observables)**
- **Inspections and reporting are quasi-random in many jurisdictions (Levine, Toffel, and Johnson, *Science,* 2012)**
- **Robust empirical and anecdotal evidence for similar mechanisms ("Porter Hypothesis", "Worker Safety at Alcoa," etc.)**

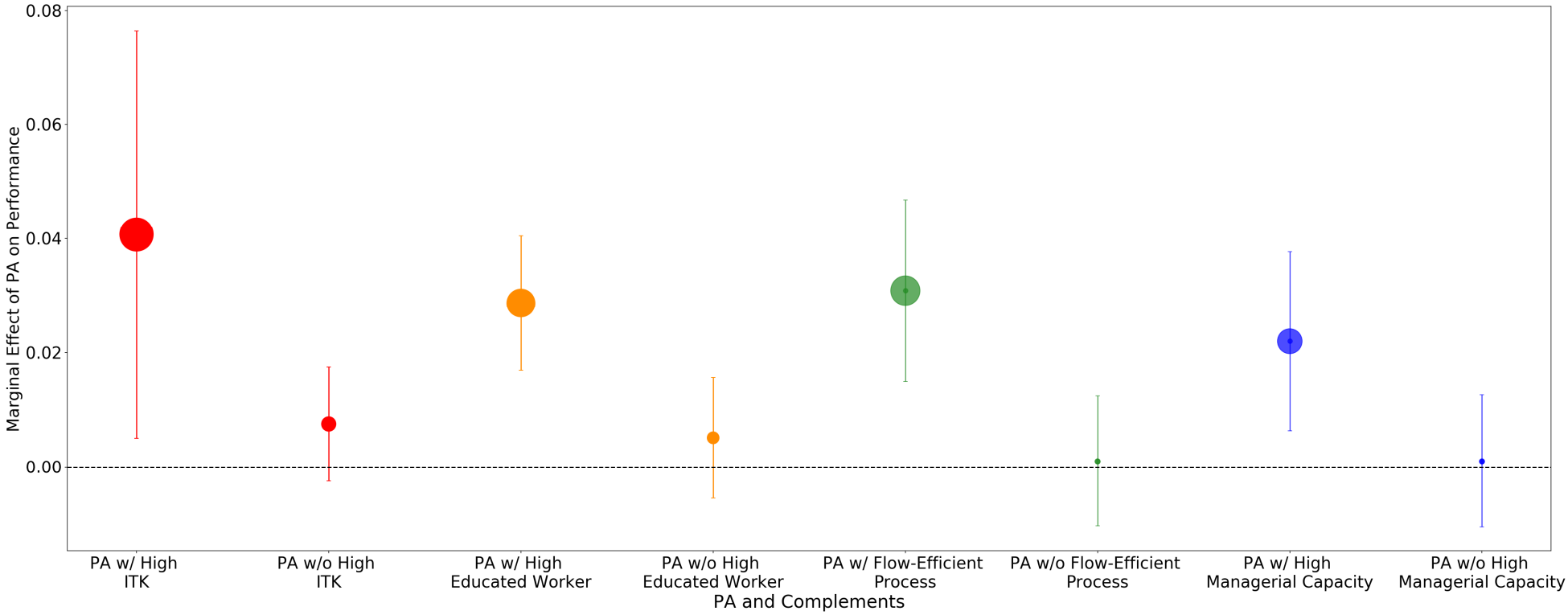
RESULTS: BASELINE PRODUCTIVITY

Model	(1) Baseline	(2) Full	(3) Frequency	(4) IV
Dependent Variable		Log Sales		
Predictive Analytics (Adoption)	0.029***	0.015***		
	(0.005)	(0.005)		
Predictive Analytics (Frequency)			0.009***	0.051***
			(0.002)	(0.016)
Government Mandated Data Collection (First stage)				0.322***
				(0.017)
Other controls (L, M, Non-ITK, HQ, and MU)	Y	Y	Y	Y
Other controls (Complements)	N	Y	Y	Y
Industry x Year FX	Y	Y	Y	Y
N			51,000	
R2	0.931	0.933	0.940	0.879

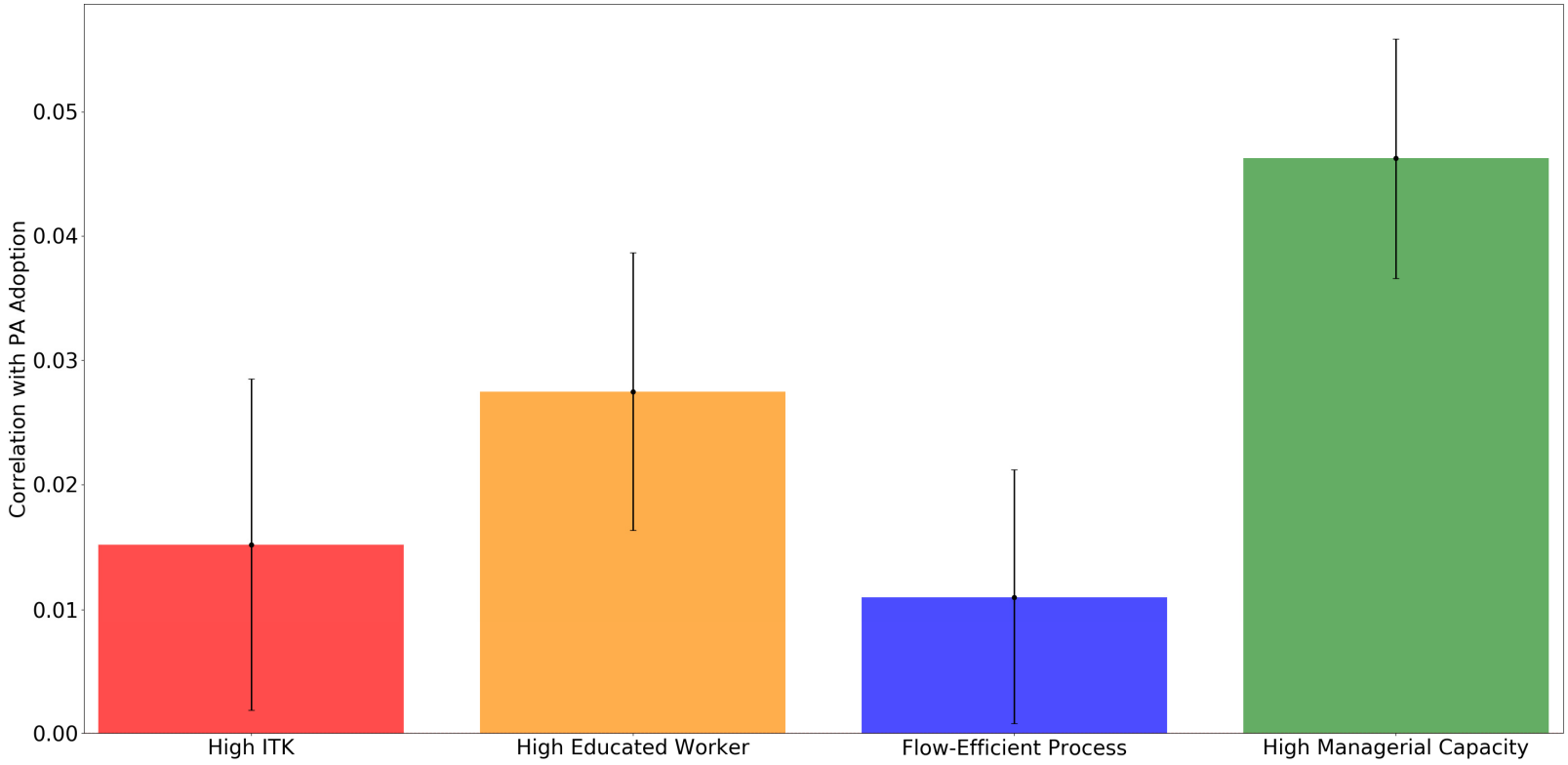
PERFORMANCE TEST FOR COMPLEMENTS

Model	(1) High IT Capital Stock	(2) High Educated Workers	(3) High-Flow	(4) Top KPI Tracking
Dependent Variable		Log Sales		
Predictive Analytics (Adoption)	0.008 (0.005)	0.005 (0.005)	0.001 (0.006)	-0.002 (0.008)
Complement	0.110*** (0.017)	0.037*** (0.010)	0.016* (0.009)	0.002 (0.009)
Predictive Analytics x Complements	0.033* (0.019)	0.023** (0.011)	0.031*** (0.009)	0.024** (0.010)
Other controls (L, M, Non-ITK, HQ, and MU)	Y	Y	Y	Y
Ind x Year FX	Y	Y	Y	Y
N		51,000		
R2	0.933	0.933	0.933	0.933

SIGNIFICANT PERFORMANCE GAINS ONLY IN PRESENCE OF COMPLEMENTS

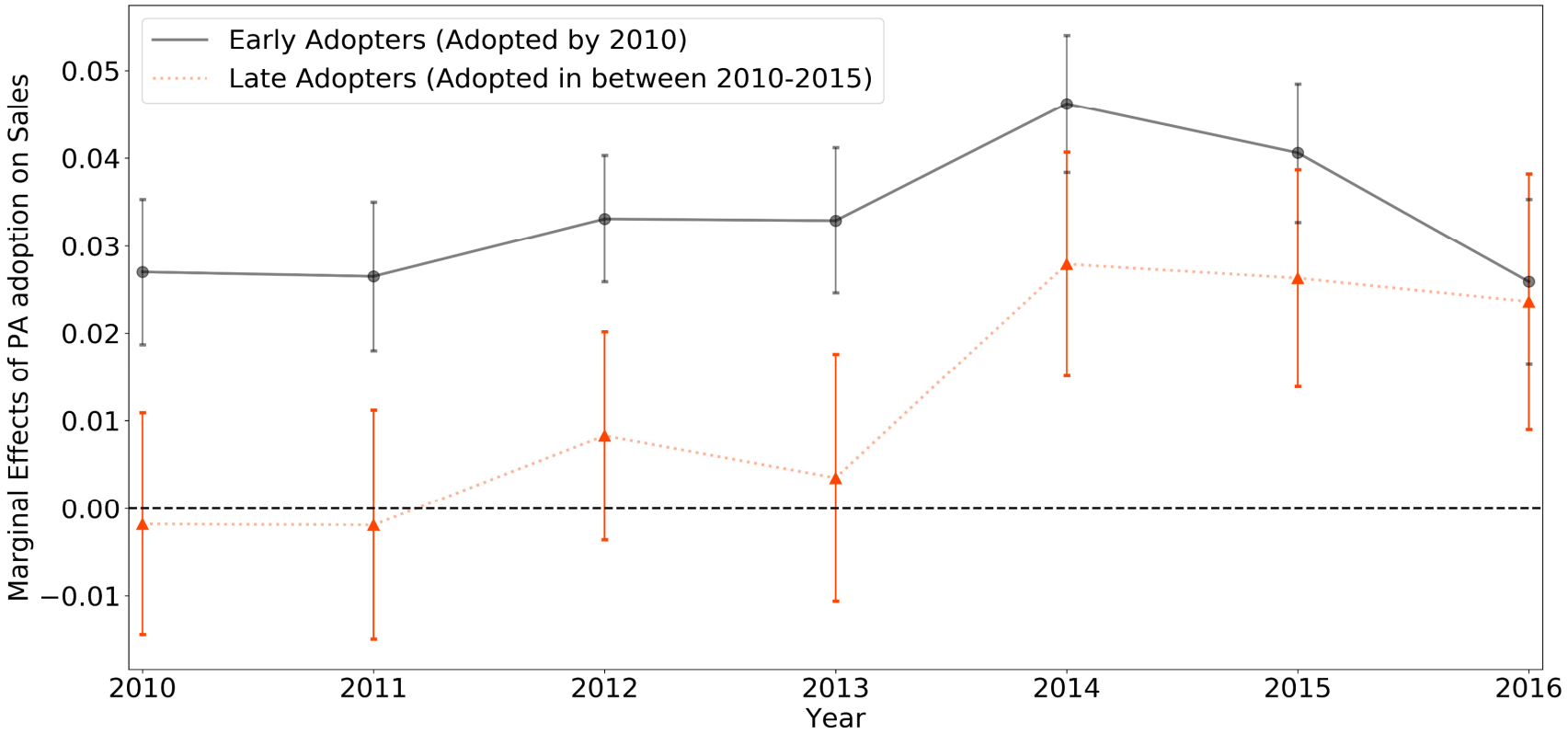


POSITIVE & SIGNIFICANT CORRELATED ADOPTION WITH HIGH COMPLEMENTS



UNIVERSITY OF TORONTO

PERFORMANCE GAINS SHOW UP AFTER ADOPTION



UNIVERSITY OF TORONTO

CONCLUSION AND IMPLICATIONS

1. **Wide adoption of predictive analytics in the U.S. manufacturing sector**
2. **Predictive analytics significantly improves performance**
   - Causal evidence from timing and from IV
3. **Complementary investments are key to performance gains**
   a) High accumulated IT capital
   b) College-educated labor
   c) Flow-efficient production strategy
   d) Managerial capacity & "Humans in the Loop"

"Known" but slow-changing

Strategic Commitment

Intangibles & "Future of Work" Optimism

Thank you!

Comments and Questions

k.mcelheran@utoronto.ca