BIG Data - BIG Gains? Empirical Evidence on the Link Between Big Data Anayltics and Innovation

Thomas Niebel*, Fabienne Rasel, Steffen Viete
Centre for European Economic Research (ZEW), Mannheim

Seville, 27-29 September 2017

6th European Conference on Corporate R&D and Innovation - CONCORDi 2017

Motivation

- Big Data is expected to enable firms from all industries to create new products and services, improve existing ones, and to develop new business models (Manyika et al. 2011, OECD 2015)

- Standard definition of big bata: (1) volume, (2) variety, (3) velocity

- Possible innovation-enabling mechanisms through big data:

  - Through better informed decision-making during the R&D and innovation process $\rightarrow$ big data alters the sources and types of information available to decision-makers (Constantiou and Kallinikos 2015)
  - Through big data technologies at the core of the innovation itself $\rightarrow$ Combining different information facilitates the development of new personalized services (Varian 2010)

$\Rightarrow$ Despite the hype, little empirical evidence

Literature: Big Data and Firm Performance

- Performance advantage of using data-driven decision-making in U.S. firms (Brynjolfsson et al. 2011; Tambe 2014; Brynjolfsson and McElheran 2016)
- ICT as an enabler for innovation (e.g. Brynjolfsson and Saunders 2010; Spiezia 2011)

This Paper and Contribution

- Research Questions:
  - Does the use of big data analytics increase the likelihood of realizing a product innovation?
  - Does big data analytics also increase the innovation intensity in terms of the market success of new product innovations?
- Contribution:
  - First large scale empirical evidence based on firm-level data on the role of big data for product innovation of manufacturing and service firms
  - Enhances knowledge of the value of large scale data usage to support firm performance

Empirical Methods: Innovation Propensity

Relationship of interest: Contribution of big data to realizing a product innovation

Knowledge production function framework (Griliches 1979):

$$y^*_{1i} = \beta_1 bigdata_i + \gamma_1' \mathbf{c}_{1i} + \epsilon_{1i} \tag{1}$$

$$y_{1i} = \mathbf{1}[y^*_{1i} > 0] \tag{2}$$

- y_{1i}: innovation success, i.e. the event of introducing a new product or service to the market
- $\mathbf{c}_{1i}$: vector of control variables
- ϵ_{1i}: idiosyncratic error term, $NID(0, \sigma_1^2)$

→ Probit Model

Empirical Methods: Innovation Intensity

Relationship of interest: Contribution of big data to innovation intensity of new products or services:

$$y_{2i}^* = \beta_2 bigdata_i + \gamma_2' \mathbf{c}_{2i} + \epsilon_{2i} \tag{3}$$

$$y_{2i} = \mathbf{1}[y_{2i}^* > 0] y_{2i}^*. \tag{4}$$

- y_{2i}: sales share of new products or services
- $\mathbf{c}_{2i}$: vector of control variables
- ϵ_{1i}: idiosyncratic error term, $NID(0, \sigma_1^2)$

$\rightarrow$ Tobit Model

Also: Fractional Logit, Heckman Selection Model

Data: ZEW ICT Survey 2015

- Firm-level survey conducted by Centre for European Economic Research (ZEW)
- Telephone interviews of around 4,400 firms located in Germany, stratified on a sectoral and size class
- Focus on diffusion and use of ICT: different ICT measures, e.g. computer and Internet use, enterprise software, big data
- General firm characteristics: number of employees, total turnover, export activity, innovation activity, workplace practices

Innovation Measures

- Product innovation realization: binary variable
- Product innovation performance: share of sales due to new products or services

	N	Mean	Median	SD	Min	Max
Product Innovation	2727	0.48	0	0.50	0	1
% of Sales New Product	2727	0.084	0	0.15	0	1

	By Big Data Use					
	No Big Data		Big Data		Total	
	N	Mean	N	Mean	N	Mean
Product Innovation	2134	0.45	593	0.60	2727	0.48
% of Sales New Product	2134	0.07	593	0.12	2727	0.08

Big Data Analytics and other IT Variables (1)

Big data analytics as a binary variable based on the following question:

> *"Up next a question about so called big data, i.e. the processing of large amounts of data. Does your company systematically analyze large amounts of data to support business operations?"*

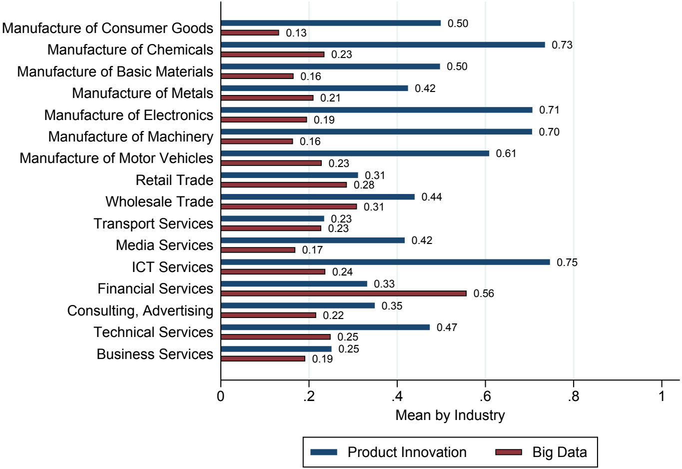
Other IT variables:

- Share of employees using the PC predominantly
- Share of employees using the Internet
- Enterprise Software (binary variable)

Big Data Analytics and other IT Variables (2)

	N	Mean	Median	SD	Min	Max
Big Data	2727	0.22	0	0.41	0	1
% of Emp. Predom. Using PC	2727	0.45	0.33	0.34	0	1
% of Emp. Using Internet	2727	0.57	0.50	0.37	0	1
Enterprise Software	2727	0.56	1	0.50	0	1

Industry Means of Product Innovation and Big Data



2727 Observations

Dependent Variable: Dummy for Product Innovation - Probit Regression - Average Marginal Effects

	(1) Full Sample	(2) Manufacturing	(3) Services
Big Data	**0.067*** **	**0.066***	**0.068** **
	(0.023)	(0.035)	(0.029)
% of Emp. Predom. Using PC	-0.000	-0.089	0.058
	(0.042)	(0.074)	(0.051)
% of Emp. Using Internet	0.080**	0.080	0.074
	(0.035)	(0.049)	(0.052)
Enterprise Software	0.086***	0.115***	0.064**
	(0.020)	(0.030)	(0.026)
% of R&D Expenses	0.912***	1.123***	0.776***
Employees (in logs)	0.010	0.015	0.009
	(0.012)	(0.017)	(0.015)
Investment (in logs)	0.024***	0.019*	0.029***
	(0.007)	(0.011)	(0.010)
Exporter	0.165***	0.145***	0.183***
	(0.021)	(0.029)	(0.032)
% Highly Qualified Employees	0.159***	0.372***	0.043
	(0.061)	(0.123)	(0.079)
% Medium Qualified Employees	-0.040	-0.017	-0.097
	(0.043)	(0.055)	(0.069)
Further Controls	Yes	Yes	Yes
Pseudo R^2	0.207	0.182	0.212
Observations	2727	1415	1312

Standard errors in parentheses, * p<0.10, ** p<0.05, *** p<0.01.

All models include an intercept.

Dependent Variable: Share of New Products in Turnover - Tobit/FracReg Regressions

	Full Sample		Manufacturing		Services	
	(1)	(2)	(3)	(4)	(5)	(6)
	Tobit	FracReg	Tobit	FracReg	Tobit	FracReg
Big Data	**0.025*** **	**0.029*** **	**0.027*** **	**0.033*** **	**0.025*** **	**0.028*** **
	(0.006)	(0.008)	(0.009)	(0.011)	(0.008)	(0.010)
% of Emp. Predom. Using PC	0.006	0.009	-0.006	-0.000	0.018	0.022
	(0.011)	(0.013)	(0.019)	(0.022)	(0.014)	(0.015)
% of Emp. Using Internet	0.018*	0.016	0.022*	0.023	0.015	0.008
	(0.010)	(0.011)	(0.013)	(0.015)	(0.014)	(0.018)
Enterprise Software	0.020***	0.018***	0.031***	0.029***	0.013**	0.012
	(0.005)	(0.006)	(0.007)	(0.008)	(0.007)	(0.008)
% of R&D Expenses	0.253***	0.195***	0.319***	0.240***	0.199***	0.157***
	(0.020)	(0.024)	(0.035)	(0.049)	(0.023)	(0.024)
Employees (in logs)	-0.007**	-0.014***	-0.004	-0.010*	-0.009**	-0.019***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)
Investment (in logs)	0.007***	0.008***	0.003	0.002	0.011***	0.014***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Exporter	0.037***	0.030***	0.034***	0.028***	0.039***	0.030***
	(0.005)	(0.007)	(0.007)	(0.009)	(0.008)	(0.010)
% Highly Qualified Employees	0.036**	0.027	0.055**	0.025	0.016	0.028
	(0.016)	(0.019)	(0.028)	(0.032)	(0.022)	(0.023)
% Medium Qualified Employees	-0.015	-0.018	-0.019	-0.033	-0.019	-0.001
	(0.012)	(0.015)	(0.015)	(0.020)	(0.019)	(0.021)
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.363	0.092	0.403	0.069	0.330	0.126
Observations	2727	2727	1415	1415	1312	1312
Censored	1441		636		805	
Uncensored	1286		779		507	

Heckman Selection Model with exclusion restriction, Marginal Effects

	Full Sample		Manufacturing		Services	
	(1)	(2)	(3)	(4)	(5)	(6)
	1st	2nd	1st	2nd	1st	2nd
Big Data	**0.066*** **	**0.022*** **	**0.066* **	**0.026** **	**0.066** **	**0.023* **
	(0.022)	(0.008)	(0.034)	(0.010)	(0.029)	(0.013)
% of Emp. Predom. Using PC	-0.002	0.002	-0.094	0.016	0.056	-0.008
	(0.043)	(0.017)	(0.071)	(0.022)	(0.053)	(0.027)
% of Emp. Using Internet	0.080**	-0.004	0.079	0.009	0.076	-0.026
	(0.036)	(0.014)	(0.050)	(0.016)	(0.053)	(0.028)
Enterprise Software	0.086***	-0.007	0.115***	-0.003	0.064**	-0.011
	(0.020)	(0.008)	(0.030)	(0.010)	(0.026)	(0.013)
% of R&D Expenses	0.950***	0.221***	1.230***	0.282***	0.796***	0.179***
	(0.112)	(0.025)	(0.202)	(0.039)	(0.128)	(0.037)
Employees (in logs)	0.011	-0.020***	0.016	-0.011**	0.009	-0.031***
	(0.011)	(0.004)	(0.017)	(0.006)	(0.015)	(0.007)
Investment (in logs)	0.024***	0.004	0.018*	-0.004	0.029***	0.014***
	(0.008)	(0.003)	(0.011)	(0.004)	(0.010)	(0.005)
% Highly Qualified Employees	0.160***	0.003	0.378***	-0.041	0.041	0.046
	(0.061)	(0.023)	(0.113)	(0.034)	(0.081)	(0.040)
% Medium Qualified Employees	-0.037	-0.011	-0.015	-0.027	-0.097	0.034
	(0.043)	(0.017)	(0.056)	(0.019)	(0.069)	(0.037)
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2727	2727	1415	1415	1312	1312
$\hat{\sigma}_{12}$	-0.238		-0.282		-0.291	
LR-Test $H_0 : \sigma_{12} = 0$ $[\chi^2(1)]$, p-Val	0.127		0.113		0.203	
Log Likelihood	-977.274		-411.139		-533.017	

Subsequent analyses

- Split sample regressions for low vs. high general human capital as well as low vs. high IT-skills:
  Relationship between big data use and the likelihood to innovate does ...
  - not depend on general human capital
  - depend IT-specific knowledge and skills.
- In-depth analysis on how firms use big data:

  Firms in the manufacturing and service sector applying big data rely on different sources of digital information and different data-related firm practices to reap the benefits of big data analytics.

Conclusion

- Significantly positive relationship between use of big data analytics and innovation performance, for manufacturing and for service firms
  $\Rightarrow$ Big data analytics supports innovation activities

- Employees with appropriate IT-skills are required to reap the benefits of big data.
- Manufacturing and services firms differ with respect to how they use big data analytics.

Thank you for your attention!

niebel@zew.de

Back up

References I

Brynjolfsson, E., Hitt, L. M. and Kim, H. H. (2011), 'Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?'. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.1819486`.

Brynjolfsson, E. and McElheran, K. (2016), Data in Action: Data-Driven Decision Making in U.S. Manufacturing, Working Papers 16-06, Center for Economic Studies, U.S. Census Bureau.

Brynjolfsson, E. and Saunders, A. (2010), *Wired for Innovation: How IT is Reshaping the Economy*, The MIT Press.

Constantiou, I. D. and Kallinikos, J. (2015), 'New Games, New Rules: Big Data and the Changing Context of Strategy', *Journal of Information Technology* **30**(1), 44–57.

References II

Griliches, Z. (1979), 'Issues in Assessing the Contribution of R&D to Productivity Growth', *Bell Journal of Economics* **10**(1), 92–116.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute. Available at `http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation`. Last Accessed: 26 September 2017.

References III

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, Available at http://dx.doi.org/10.1787/9789264229358-en.

Spiezia, V. (2011), 'Are ICT Users More Innovative?: An Analysis of ICT-Enabled Innovation in OECD Firms', *OECD Journal: Economic Studies* **2011**(1), 1–21.

Tambe, P. (2014), 'Big Data Investment, Skills, and Firm Value', *Management Science* **60**(6), 1452–1469.

Varian, H. R. (2010), 'Computer Mediated Transactions', *The American Economic Review* **100**(2), 1–10.

Descriptives - Estimation Sample

	N	Mean	Median	SD	Min	Max
Product Innovation	2727	0.48	0	0.50	0	1
% of Sales New Product	2727	0.084	0	0.15	0	1
Big Data	2727	0.22	0	0.41	0	1
% of Emp. Predom. Using PC	2727	0.45	0.33	0.34	0	1
% of Emp. Using Internet	2727	0.57	0.50	0.37	0	1
Enterprise Software	2727	0.56	1	0.50	0	1
% of R&D Expenses	2727	0.051	0.0059	0.11	0	1
Employees	2727	93.6	25	263.2	5	4500
Employees (in logs)	2727	3.44	3.22	1.31	1.61	8.41
Investment in Mill. Euro	2727	0.91	0.10	4.68	0.00050	130
Investment (in logs)	2727	-2.02	-2.30	1.84	-7.60	4.87
Exporter	2727	0.45	0	0.50	0	1
% Highly Qualified Employees	2727	0.19	0.10	0.24	0	1
% Medium Qualified Employees	2727	0.63	0.70	0.27	0	1
% of Employees < Age 30	2727	0.24	0.20	0.17	0	1
% of Employees > Age 50	2727	0.27	0.25	0.19	0	1
East Germany	2727	0.24	0	0.43	0	1
Age (in logs)	2727	3.17	3.14	0.92	0	6.39
Group	2727	0.30	0	0.46	0	1
Multinational	2727	0.095	0	0.29	0	1

Summary Statistics by Big Data Use of Firms: Estimation Sample

	not		Big Data		Total	
	N	Mean	N	Mean	N	Mean
Product Innovation	2134	0.45	593	0.60	2727	0.48
% of Sales New Product	2134	0.07	593	0.12	2727	0.08
Big Data	2134	0.00	593	1.00	2727	0.22
% of Emp. Predom. Using PC	2134	0.42	593	0.55	2727	0.45
% of Emp. Using Internet	2134	0.55	593	0.65	2727	0.57
Enterprise Software	2134	0.51	593	0.78	2727	0.56
% of R&D Expenses	2134	0.04	593	0.07	2727	0.05
Employees	2134	65.73	593	193.88	2727	93.60
Employees (in logs)	2134	3.24	593	4.18	2727	3.44
Investment in Mill. Euro	2134	0.57	593	2.16	2727	0.91
Investment (in logs)	2134	-2.28	593	-1.07	2727	-2.02
Exporter	2134	0.44	593	0.49	2727	0.45
% Highly Qualified Employees	2134	0.19	593	0.21	2727	0.19
% Medium Qualified Employees	2134	0.63	593	0.61	2727	0.63
% of Employees < Age 30	2134	0.23	593	0.26	2727	0.24
% of Employees > Age 50	2134	0.28	593	0.26	2727	0.27
East Germany	2134	0.25	593	0.22	2727	0.24
Age (in logs)	2134	3.13	593	3.30	2727	3.17
Group	2134	0.26	593	0.43	2727	0.30
Multinational	2134	0.08	593	0.15	2727	0.09

Heckman Selection Model (no exclusion restriction), Marginal Effects

	Full Sample		Manufacturing		Services	
	(1)	(2)	(3)	(4)	(5)	(6)
	1st	2nd	1st	2nd	1st	2nd
Big Data	0.065***	0.022***	0.066*	0.025**	0.066**	0.023*
	(0.022)	(0.008)	(0.034)	(0.010)	(0.029)	(0.013)
% of Emp. Predom. Using PC	-0.003	0.002	-0.095	0.018	0.056	-0.010
	(0.043)	(0.017)	(0.071)	(0.023)	(0.053)	(0.028)
% of Emp. Using Internet	0.079**	-0.004	0.079	0.009	0.076	-0.026
	(0.036)	(0.014)	(0.050)	(0.016)	(0.053)	(0.028)
Enterprise Software	0.086***	-0.007	0.115***	-0.003	0.064**	-0.013
	(0.020)	(0.008)	(0.030)	(0.010)	(0.026)	(0.013)
% of R&D Expenses	0.956***	0.221***	1.244***	0.282***	0.800***	0.178***
	(0.111)	(0.026)	(0.201)	(0.039)	(0.127)	(0.038)
Employees (in logs)	0.010	-0.020***	0.015	-0.011**	0.009	-0.031***
	(0.011)	(0.004)	(0.017)	(0.006)	(0.015)	(0.008)
Investment (in logs)	0.024***	0.004	0.018	-0.004	0.029***	0.014**
	(0.008)	(0.003)	(0.011)	(0.004)	(0.010)	(0.005)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2727	2727	1415	1415	1312	1312
$\hat{\sigma}_{12}$	-0.273		-0.316		-0.340	
LR-Test $H_0 : \sigma_{12} = 0$ $[\chi^2(1)]$, p-Val	0.089		0.067		0.143	
Log Likelihood	-976.942		-410.712		-532.725	

Dependent Variable: Dummy for Product Innovation - Probit Regression by Skill Group - Average Marginal Effects

	General Human Capital		IT-specific skills	
	(1)	(2)	(3)	(4)
	low	high	low	high
Big Data	0.075**	0.066**	0.044	0.096***
	(0.033)	(0.030)	(0.034)	(0.029)
% of Emp. Predom. Using PC	-0.128**	0.089	-0.006	-0.033
	(0.064)	(0.058)	(0.059)	(0.061)
% of Emp. Using Internet	0.092*	0.010	0.110**	0.041
	(0.049)	(0.050)	(0.046)	(0.056)
Enterprise Software	0.112***	0.044	0.099***	0.046
	(0.028)	(0.028)	(0.026)	(0.030)
% of R&D Expenses	0.684***	0.995***	0.805***	0.960***
	(0.204)	(0.220)	(0.190)	(0.270)
Employees (in logs)	0.006	0.011	0.017	-0.009
	(0.017)	(0.016)	(0.016)	(0.017)
Investment (in logs)	0.031***	0.021**	0.017*	0.035***
	(0.010)	(0.011)	(0.009)	(0.012)
Pseudo R^2	0.192	0.239	0.186	0.215
Observations	1394	1312	1491	1215
Log likelihood	-765.502	-688.597	-813.189	-647.612

Robust standard errors in parentheses, * p<0.10, ** p<0.05, *** p<0.01.

Table A.4: Variable Descriptions for Supplementary Survey

Variable	Description/Question
Social Media and Online Content	
s_feedback	firm offers customers to evaluate products or services online
s_content	systematically search user-generated content about own products or services or about the company
s_ads	firm engages in online advertising
Data in Production of Goods and Services	
We use automated data recording, processing and transmission in order to...	
p_efficiency	...make internal processes more efficient, e.g. reduce material or energy consumption.
p_assistance	...provide our employees with digital assistance systems, e.g. in logistics, production, maintenance.
p_edi	...exchange information with suppliers and customers.
p_customize	...customize products/services to individual customer needs.
p_adapt	...adapt internal processes flexibly or remedy errors.
p_network	firm introduced integration of IT between different business processes or divisions during the previous two years
Digital Technology in Goods and Services	
g_sensor	firm produces products with embedded RFID-Chips, QR-codes, sensors
g_apps	firm offers mobile apps for products or services
g_service	firm offers product-supporting services, e.g. online platforms, software
g_partner_service	partner firms offer product-supporting services

Source: ZEW ICT-Survey 2015 and supplementary survey 2015/2016.

Table A.8: Controlled Correlations: Big Data and Social Media and Online Content

	s_feedback	s_content	s_ads
Manufacturing	0.077**	0.093**	0.054
	(0.039)	(0.041)	(0.040)
Services	0.049	0.030	0.107**
	(0.037)	(0.039)	(0.041)

NOTES: This table shows partial correlations between Big Data use and firms' utilization of various aspects of social media and online content. Parameter estimates of OLS regression analysis are shown. Included control variables are the share of employees working predominantly with PCs, an indicator for use of ERP software, employees (in logs), the share of highly and medium qualified employees, the share of employees who received IT-training, industry and regional indicators. Results are based on 872 obs. in manufacturing and 726 obs. in services. Robust standard errors in parentheses, * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.9: Controlled Correlations: Big Data and Data in Production of Goods and Services

	p_efficiency	p_assistance	p_edi	p_network	p_customize	p_adapt
Manufacturing	0.069	0.139***	0.088*	0.133***	0.106**	0.146***
	(0.042)	(0.039)	(0.045)	(0.043)	(0.044)	(0.043)
Services	0.020	0.070	0.049	0.012	0.094**	0.087*
	(0.043)	(0.043)	(0.043)	(0.042)	(0.045)	(0.044)

NOTES: This table shows partial correlations between Big Data use and various forms of firms' utilization of data in the production process. Parameter estimates of OLS regression analysis are shown. Included control variables are the share of employees working predominantly with PCs, an indicator for use of ERP software, employees (in logs), the share of highly and medium qualified employees, the share of employees who received IT-training, industry and regional indicators. Results are based on 872 obs. in manufacturing and 726 obs. in services. Robust standard errors in parentheses, * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.10: Controlled Correlations: Big Data and Digital Technology in Goods and Services

	g_sensor	g_apps	g_service	g_partner_service
Manufacturing	0.011	0.038	0.012	0.079**
	(0.033)	(0.026)	(0.041)	(0.037)
Services	0.056**	0.100***	0.080*	0.071*
	(0.026)	(0.037)	(0.042)	(0.043)

NOTES: This table shows partial correlations between Big Data use and various forms of digital technology embedded in final goods and services. Parameter estimates of OLS regression analysis are shown. Included control variables are the share of employees working predominantly with PCs, an indicator for use of ERP software, employees (in logs), the share of highly and medium qualified employees, the share of employees who received IT-training, industry and regional indicators. Results are based on 872 obs. in manufacturing and 726 obs. in services. Robust standard errors in parentheses, * $p<0.10$, ** $p<0.05$, *** $p<0.01$.