Industrial Policy at Work: Evidence from Romania's Income Tax Break for Workers in IT

Isabela Manelici* and Smaranda Pantea†

February 16, 2021

Appendices for Online Publication

These online appendices supplement the paper "Industrial Policy at Work: Evidence from Romania's Income Tax Break for Workers in IT" with the following material:

- Online Appendix A provides context on the policies relevant to the IT sector in Romania. In particular, we describe the 2001 policy that grants a full personal income tax break to programmers and its subsequent amendments.

- Online Appendix B includes additional firm-level evidence (e.g., descriptive statistics for the Amadeus and administrative firm-level data and robustness checks on the firm-level effects of the 2013 reform to the income tax break law).

- Online Appendix C includes additional sector-level evidence (e.g., descriptive statistics with sector-level data on firm entries and exit, and robustness checks on the sector-level cross-country effects of the 2001 tax break).

- Online Appendix D describes the data construction for the sector-level synthetic control analysis.

*Department of Economics (Princeton University). Email: manelici@princeton.edu

†Directorate of Policy, Analysis, and Research in Public Economics (Ministry of Public Finance of Romania) and the Departments of Entrepreneurship and of International Economic Relations (Prague University of Economics and Business). Email: smaranda.pantea@vse.cz

Online Appendix A Context on Policies Relevant to the IT Sector in Romania

Online Appendix A.1 Details on Romania's Income Tax Break for Workers in IT

Table A1: 2001 Income Tax Break and Its Subsequent Amendments: Eligibility Criteria for the Income Tax Break (1/2)

Order Number	4079/ 268/ 1480/ 2001	661/ 444/ 2196/ 2001	250/ 189/ 748/ 2004	539/ 225/ 1479/ 2013	217/ 4172/ 1348/ 835/ 2015	872/ 5932/ 2284/ 2903/ 2016
Effective date	08/05/2001	11/22/2001	06/29/2004	09/10/2013	07/30/2015	01/09/2017
NACE Code	7220 (Rev 1)	7220 (Rev 1)	**7221, 7222 (Rev 1.1)**	**5821, 5829, 6201, 6202, 6209 (Rev 2)**	5821, 5829, 6201, 6202, 6209 (Rev 2)	5821, 5829, 6201, 6202, 6209 (Rev 2)
Occupations	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects **Computer System Programmer**	Analyst, Programmer, Computer Systems Designer, Engineer or Programmer of IT Systems, Database Manager, Software Engineer and Manager of IT Projects **Computer System Programmer**
Unit within the Firm	Unit Specialized in IT	Unit Specialized in IT	Unit Specialized in IT	Unit Specialized in IT	Unit Specialized in IT	Unit Specialized in IT

Extra Eligibility Criteria for the Income Tax Break, in Addition to Those from Table A1 (2/2)

Order Number: 4079/ 268/ 1480/ 2001

Eligible Major during Higher Education for Exempted Worker: Automation, Computers, Computer Science, Cybernetics, Mathematics, Electronics.

Minimum Annual Revenue from Software Development: Annual income of at least 10,000 U.S. dollars per employee benefiting from the income tax break.

Balance for Software Development Income: A balance is required, in which the income from software development needs to be explicitly reported.

Order Number: 661/444/2196/2001

Eligible Major during Higher Education for Exempted Worker: Automation and Industrial Computer Science; Computers, Electrical Engineering and Computers; Electronics; Applied Electronics, Electronics and Telecommunications, Communications; Mathematics, Mathematical Computer Science; Computer Science, Computer Science and Economics, Applied Computer Science, Cybernetics and Computer Science and Economics, Cybernetics and Economic Prediction, Accounting and Computer Science and Management.

Minimum Annual Revenue from Software Development: Annual income, in preceding year, of at least 10,000 U.S. dollars per employee benefiting in a given year from the income tax break.

Balance for Software Development Income: A balance is required, in which the income from software development needs to be explicitly reported.

Order Number: 250/189/748/2004

Eligible Major during Higher Education for Exempted Worker: Automation and Industrial Computer Science; Computers, Electrical Engineering and Computers; Electronics; Applied Electronics, Electronics and Telecommunications, Communications; Mathematics, Mathematical Computer Science; Computer Science, Computer Science and Economics, Applied Computer Science, Cybernetics and Computer Science and Economics, Cybernetics and Economic Prediction, Accounting and Computer Science and Management.

Minimum Annual Revenue from Software Development: Annual income, in preceding year, of at least 10,000 U.S. dollars per employee benefiting in a given year from the income tax break.

Balance for Software Development Income: A balance is required, in which the income from software development needs to be explicitly reported.

Order Number: 539/225/1479/2013

Eligible Major during Higher Education for Exempted Worker: Automation and Industrial Computer Science; Computers, Electrical Engineering and Computers; Electronics; Applied Electronics, Electronics and Telecommunications, Communications; Mathematics, Mathematical Computer Science; Computer Science, Computer Science and Economics, Applied Computer Science, Cybernetics and Computer Science and Economics, Cybernetics and Economic Prediction, Accounting and Computer Science and Management. **Newly eligible majors:** Industrial Computer Science, Applied Computer Science in Electrical Engineering, Applied Computer Science in Material Engineering, Mathematics and Applied Computer Science in Engineering; Cybernetics and Economics; Physics and Computer Sci-

ence; Chemistry and Computer Science; Automation and Applied Computer Science, Equipment for Modeling, Simulation and Computerized Warfare, Engineering of Multimedia Systems; Technologies and Telecommunication Systems, Remote Controls and Electronics in Transportation; Transmissions and Military Electronic Equipment.

Minimum Annual Revenue from Software Development: Annual income, in preceding year, of at least 10,000 U.S. dollars per employee benefiting in a given year from the income tax break.

Balance for Software Development Income: A balance is required, in which the income from software development needs to be explicitly reported.

Order Number: 217/4172/1348/835/2015 and 872/5932/2284/2903/2016

Eligible Major during Higher Education for Exempted Worker: A diploma issued after a form of higher education, irrespective of major.

Minimum Annual Revenue from Software Development: Annual income, in preceding year, of at least 10,000 U.S. dollars per employee benefiting in a given year from the income tax break.

Balance for Software Development Income: A balance is required, in which the income from software development needs to be explicitly reported. **New:** New firms or firms undergoing a restructuring during that fiscal year are exempted from this requirement.

Online Appendix A.2 Other Policies Relevant for the IT sector

Online Appendix A.2.1 State Aid Program

Between 2011 and 2016, Romania introduced several state aid programs supporting job creation.[1] The most relevant for the sectors studied was the program created by Government Decision 797/2012. It supported large investments in new technologies with an IT component and job creation of at least of 200 new jobs. While firms in most manufacturing, energy and service sectors were eligible, mainly firms in high-tech knowledge-intensive (HTKI) sectors benefited from it. We drop from our sample of analysis firms that have benefited from this State Aid. The program created by Government decision 332/2014 (meant to support large investments, job creation, and regional development) also benefited several firms in HTKI sectors. Firms in HTKI sectors were also eligible for programs supporting SMEs and start-ups, such as "Start-up Nation," but these programs were smaller and less likely to affect major investments or job creation (e.g., "Start-up Nation" had an upper limit of approximately 44,000 euros).

Online Appendix A.2.2 Other Tax Exemptions

Programmers are not the only category of workers exempted from the personal income tax in Romania. Two other categories of workers exempted from the income tax could be employed by IT firms, without being programmers: workers with serious disabilities[2] and since 2016 workers in research and development (defined broadly, with no requirement to work in software development).[3] For companies with at least 50 employees, it was compulsory that at least 4% of their workers have disabilities. When

[1] Government decisions 797/2012, 322/2014 and 807/2014.
[2] See Law 448/2006 for details.
[3] See Order 4947/899/2018/1840/906/2016 published in September 2016 for details.

firms could not comply, they had to pay a given amount to support the inclusion of people with disabilities. Despite these other exemptions, the vast majority of exempted employees in eligible sectors were exempted due to the tax break for workers in IT. In October 2017,[4] workers benefiting from the law under study in this article represented 96% of the exempted employees in eligible sectors.

Table A2: Correspondence Table between NACE sector Codes Rev 2 and Rev 1.1

NACE Rev 2	Description Rev 2	NACE Rev 1.1	Description Rev 1.1	Comments
5821	Publishing of computer games	7221	Publishing of software	Publishing of computer games
5821	Publishing of computer games	724	Database activities	On-line computer games publishing
5829	Other software publishing	7221	Publishing of software	All software publishing, except computer games publishing
5829	Other software publishing	724	Database activities	All on-line software publishing, except computer games on-line publishing
6201	Computer programming activities	7221	Publishing of software	Software programming
6201	Computer programming activities	7222	Other software consultancy and supply	Includes: Analysis, design and programming of systems ready to use: - development, production, supply and documentation of made-to-order software based on orders from specific users - writing of programs following directives of the user - web page design
6201	Computer programming activities	724	Database activities	Designing of structure and content of database
6202	Computer consultancy activities	721	Hardware consultancy	All
6202	Computer consultancy activities	7222	Other software consultancy and supply	Analysis, design and programming of systems ready to use: - analysis of the user's needs and problems, consultancy on the best solution
6209	Other information technology and computer service activities	3002	Manufacture of computers and other information processing equipment	Installation of personal computers and peripheral equipment
6209	Other information technology and computer service activities	7222	Other software consultancy and supply	Software installation services
6209	Other information technology and computer service activities	726	Other computer related activities	NACE 1.1 class 72.60 was an "empty class".

Notes: Source Eurostat.

[4] In 2017 we can observe the reason of the income tax exemption, which we cannot observe in previous years.

Online Appendix B Additional Firm-Level Evidence

Online Appendix B.1 Descriptive Statistics with Firm-Level Data

Online Appendix B.1.1 Amadeus Data: 1997 - 2005

Table B1: Descriptive Statistics for Firms in Treated and Comparison Sectors in Year 2000

Variable	# Firms	Mean	Median	SD
Eligible sector				
Revenues	825	128.73	27	321.65
Workers	825	7.28	3	12.10
Assets	825	67.64	12	191.91
Baseline comparison sectors				
Revenues	2,117	133.61	27	369.44
Workers	2,117	7.07	3	12.89
Assets	2,117	64.10	10	188.12

Notes: Table B1 reports descriptive statistics for the baseline sample used to study the firm-level impact of the introduction in 2001 of the income tax break for workers in IT (see results in Table 1). The data source is Amadeus. These statistics pertain to the year 2000, the year prior to the introduction of the policy of interest. The upper panel reports summary statistics for the firms in the eligible sector (722, NACE Rev 1) and for the three outcome variables of interest (operating revenues, number of workers, and total assets). The lower panel reports summary statistics for the firms that belong to the baseline set of comparison sectors (i.e., high-tech knowledge-intensive service sectors, as classified by Eurostat). The unit of measure for the operating revenues and assets variables is thousands of euros.

Online Appendix B.1.2 Administrative and Amadeus Data: 2011 - 2015

Table B2: Correlation b/n Administrative and Amadeus Data in the Amadeus Sample for 2012

Sector eligible Exempted after 2013	Not eligible Not exempted	Eligible Not exempted	Eligible Exempted
Revenues	0.974	0.972	0.985
Employees	0.977	0.995	0.989
Assets	1.000	1.000	1.000

Notes: This table reports correlations for 2012 (the year before the expansion of the income tax break law), for three categories of firms and for the three outcome variables of interest (revenues, number of workers, and total assets). The correlations are between values for the same variable and firm in 2012, where the values come from two different data sources: the administrative data and Amadeus data. The set of firms for which we compute these correlations is the set of firms found in both the administrative data and the Amadeus data in 2012. The first category of firms includes firms in HTKI sectors comparable to the IT sector. These comparable sectors were, however, not targeted by the income tax break law. The second and third categories contain firms in the eligible sectors. The second category contains firms with less than 5% of employees exempted from the income tax in each year of our 2011 to 2015 sample. The third category contains firms that had less than 5% of exempted employees in 2011 and 2012, but jumped to over 20% of exempted employees after 2013. The same correlations for 2011, 2013, 2014, and 2015 are similar to those for 2012.

Table B3: Descriptive Statistics for 2012, by Firm Category

	Administrative data. Administrative sample								
Sector eligible	Not eligible			Eligible			Eligible		
Exempted after 2013	Not exempted			Not exempted			Exempted		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Revenues	1,436.32	40.49	24,878.45	196.58	30.80	961.93	317.03	34.97	2,638.64
Employees	20.22	2.00	214.06	3.91	1.00	13.25	6.38	2.00	21.33
Assets	2,113.67	39.89	33,545.48	226.35	32.01	1,731.82	203.07	27.38	1,914.87
	Administrative data. Amadeus sample								
Sector eligible	Not eligible			Eligible			Eligible		
Exempted after 2013	Not exempted			Not exempted			Exempted		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Revenues	197.34	46.30	559.74	128.12	33.62	399.88	148.82	40.73	367.53
Employees	6.65	2.00	12.51	3.62	1.00	8.55	4.90	2.00	9.52
Assets	202.68	38.54	635.34	105.12	31.91	325.19	82.13	28.30	173.95
	Amadeus data. Amadeus sample								
Sector eligible	Not eligible			Eligible			Eligible		
Exempted after 2013	Not exempted			Not exempted			Exempted		
stats	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Revenues	187.11	41.37	556.14	120.14	31.50	359.71	137.44	36.96	348.39
Employees	6.53	2.00	12.67	3.54	1.00	8.46	4.67	2.00	9.22
Assets	202.74	38.52	635.40	105.12	31.91	325.19	82.12	28.30	173.95

Notes: This table reports descriptive statistics for 2012 (the year before the expansion of the income tax break law), for three categories of firms and for the three outcome variables of interest (revenues, number of workers, and total assets). The first category includes firms in HTKI sectors comparable to the IT sector. These comparable sectors were, however, not targeted by the income tax break law. The second and third categories contain firms in the eligible sectors. The second category contains firms with less than 5% of employees exempted from the income tax in each year of our 2011 to 2015 sample. The third category contains firms that had less than 5% of exempted employees in 2011 and 2012, but jumped to over 20% of exempted employees after 2013. We present the mean, median, and standard deviation for the three categories of firms for three samples. The sample in the upper panel is the baseline sample for the administrative data. The central panel uses administrative data but on the sample of firms found in the Amadeus data. The lower panel uses Amadeus data on the sample of firms found in the Amadeus data. The unit of measure for the revenues and assets variables is thousands of euros.

Table B4: Percentage of Firms with Exempted Employees in Different Samples

Percentage of firms with at least one exempted employee, raw dataset

Year	HTKI	ICT	Eligible
2011	18%	19%	35%
2012	18%	19%	34%
2013	20%	21%	37%
2014	22%	23%	40%
2015	23%	24%	42%

Percentage of firms with at least one exempted employee, clean dataset used as main sample

Year	HTKI	ICT	Eligible
2011	20%	21%	36%
2012	19%	20%	35%
2013	21%	22%	39%
2014	23%	24%	42%
2015	25%	26%	45%

Percentage of firms with 20% or more exempted employees, clean dataset used as main sample

Year	HTKI	ICT	Eligible
2011	16%	16%	28%
2012	16%	17%	29%
2013	17%	18%	31%
2014	19%	19%	34%
2015	20%	21%	36%

Notes: Table B4 reports the share of firms in a given sample that have at least one exempted employee or that have more than 20% of their employees exempted from the income tax for each year between 2011 and 2015. By construction, all firms in ineligible sectors have less than 5% of their employees exempted from the income tax (employees who are likely to have a disability, another criterion on which this tax break can be granted). The sample of firms in "HTKI" services (HTKI stands for "high-tech knowledge-intensive") from the lower two panels ("cleaned dataset used as main sample") is the baseline sample whose results are reported in Table 3, columns (1)-(3). The sample of firms in "ICT" services sectors from the lower two panels ("cleaned dataset used as main sample") is the robustness check sample whose results are reported under the "alternative control group 1" header in Table 3, columns (4)-(6). The sample of firms in "eligible" sectors from the two lower panels ("cleaned dataset used as main sample") contains only firms in sectors whose IT workers are eligible to the income tax break and is the robustness check sample whose results are reported under the "alternative control group 3" header in Table 4 (Online Appendix B.2).

Table B5: Predictors of Firm-level Exemption Performance

	1 if more than 20% exempted workers			Share of exempted workers		
	OLS	Probit	Logit	OLS	Probit	Logit
	(1)	(2)	(3)	(4)	(5)	(6)
Foreign-owned	0.105***	0.101***	0.100***	0.068***	0.063***	0.062***
	(0.022)	(0.022)	(0.021)	(0.015)	(0.014)	(0.014)
Small	0.085***	0.083***	0.083***	0.030***	0.029***	0.030***
	(0.017)	(0.017)	(0.017)	(0.011)	(0.010)	(0.010)
Medium	0.218***	0.215***	0.213***	0.098***	0.090***	0.089***
	(0.034)	(0.035)	(0.035)	(0.021)	(0.020)	(0.020)
Large	0.242***	0.239***	0.238***	0.129***	0.119***	0.117***
	(0.075)	(0.079)	(0.079)	(0.044)	(0.042)	(0.040)
Young	-0.004	-0.003	-0.003	0.017**	0.018**	0.018**
	(0.011)	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)
County % exempt.	0.573***	0.569***	0.573***	0.406***	0.403***	0.406***
	(0.062)	(0.061)	(0.061)	(0.044)	(0.044)	(0.043)
d_{2011}	-0.011*	-0.011*	-0.012*	-0.013***	-0.014***	-0.014***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
d_{2013}	0.020***	0.021***	0.021***	0.012***	0.013***	0.013***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
d_{2014}	0.030***	0.032***	0.032***	0.017***	0.018***	0.018***
	(0.007)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)
d_{2015}	0.047***	0.048***	0.048***	0.023***	0.024***	0.024***
	(0.008)	(0.008)	(0.008)	(0.005)	(0.006)	(0.006)
Adjusted/Pseudo R^2	0.042	0.033	0.033	0.030	0.019	0.019
# Observations	21,601	21,601	21,601	21,601	21,601	21,601
# Firms	4,986	4,986	4,986	4,986	4,986	4,986

Notes: Table B5 explores the predictors of the firm-level share of income tax-exempt workers. We only focus on the sample of firms in eligible sectors, as only in those sectors workers can benefit from an income tax break based on their IT-related occupation. Columns (1) and (4) report the results of OLS models, columns (2) and (5) of probit (fractional response for the share of exempted workers) models, and columns (3) and (6) of logit (fractional response for the share of exempted workers) models. Columns (1)-(3) use as the dependent variable a dummy variable that takes value 1 for firm i in year t if firm i has more than 20% of its workers exempted from the income tax in year t. Columns (4)-(6) use as the dependent variable the share of workers of firm i who are exempted from the income tax in year t. The sample used in this table is different from the "alternative control 3" sample in Table 4 in that the former keeps *all* firms in eligible sectors, whereas the latter brings additional restrictions on the firm-level share of workforce exemption before and after 2013. The dependent variables and sample used in this table are the same as those used in Table 2, which explores the first stage effects of the 2013 reform on the firm-level share of workforce exemption from the income tax. The explanatory variables of size and age are contemporaneous. Data on foreign ownership in 2016 comes from Amadeus. We assume that if a firm is foreign-owned in 2016 it has been foreign-owned throughout 2011 to 2015. The reference category contains firms that are domestically-owned, have a micro size, and are older than five years. The reference year in 2012.

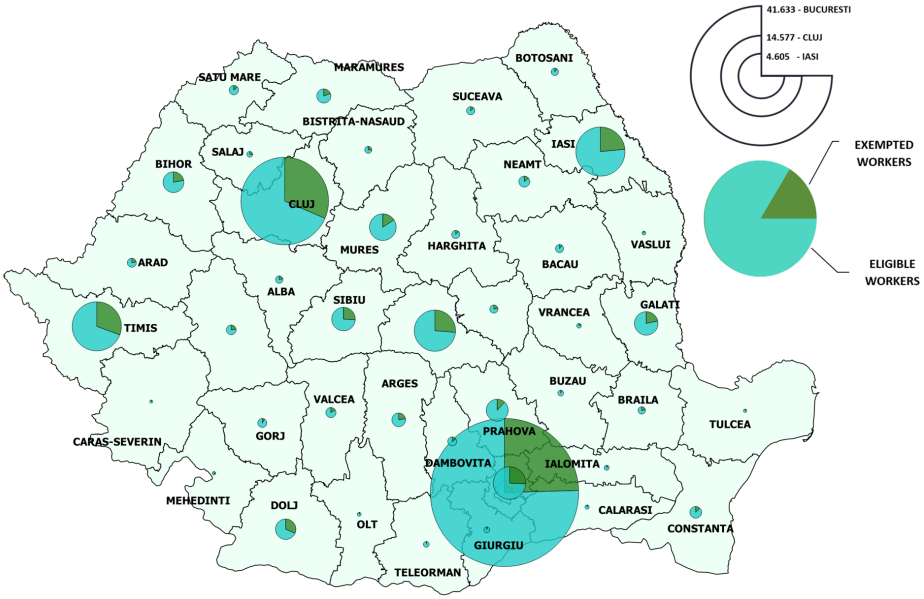

Figure B1: County-Level Number of Employees in the Eligible IT Sector in 2015 and the Share of these Employees Actually Exempted from the Income Tax

Notes: The legend reports the total number of employees in the eligible IT sectors in 2015 in the three counties with the largest number of such employees, i.e., București, Cluj, and Iași. This map uses the sample from the 2011 to 2015 administrative dataset kept for the econometric analysis, as described in Section 3.2.

Figure B1 shows the spatial distribution in 2015 of the number of workers in the sectors eligible for the income tax exemption and the share of these workers who are actually benefiting from this exemption. The unit of analysis is a county (*judeţ*). This map highlights the striking spatial concentration of the IT sector in Romania's leading counties: București (the capital), Cluj, Iași, and Timiș.

This map also draws attention to heterogeneity in the share of employees in the IT sector who actually benefit from the tax exemption. This share reaches a peak in Timiș (around 46%) and a low in Teleorman (under 5%). This heterogeneity is likely to reflect differences across workers and firms in their ability to meet the requirements of the tax break (firms in less developed counties are more likely to hire workers with a profile that does not meet the educational criteria, such firms are also less likely to earn more than 10,000 U.S. dollars per eligible employee etc.). Discrepancies are also likely to reflect differential abilities across firms to fill in the necessary paperwork for a worker to be granted the tax break. From Table B5 we learn that IT firms located in counties with a higher share of exemption of the total workforce in IT are more likely to have a higher share of workforce exemption themselves. This finding points to potentially localized knowledge spillovers on the necessary procedures to benefit from the tax break and the importance of a strong local pool of skilled programmers.

We also compute the county-level number of workers in the eligible sectors per 100,000 inhabitants. While Bucureşti remains the leader in the country with 2,257 workers in the eligible sectors per 100,000 inhabitants, Cluj continues to stand out in a comparable position with 2,078 such workers per 100,000 inhabitants. The following three performers are Timiş, Braşov, and Iaşi, each having between 580 to 660 workers in IT per 100,000 inhabitants. The lowest-ranked 30 counties are home to less than 100 workers in the eligible sectors per 100,000 inhabitants.

Online Appendix B.2 Additional Robustness Checks on the Firm-Level Effects of the 2013 Reform to the Income Tax Break

Table B6: Difference-in-Differences Design Around 2013 Reform: Robustness of the Baseline Results to the Threshold Choice

Administrative data	Threshold 10%			Threshold 15%			Threshold 20% (Baseline)			Threshold 25%			Threshold 30%		
	Revenues (1)	Workers (2)	Assets (3)	Revenues (4)	Workers (5)	Assets (6)	Revenues (7)	Workers (8)	Assets (9)	Revenues (10)	Workers (11)	Assets (12)	Revenues (13)	Workers (14)	Assets (15)
$\beta_{DiD,2011}$	-0.069*	0.020	-0.048	-0.054	0.031	-0.022	-0.054	0.031	-0.005	-0.043	0.038*	0.003	-0.007	0.048**	0.018
	(0.038)	(0.018)	(0.043)	(0.041)	(0.019)	(0.046)	(0.043)	(0.019)	(0.046)	(0.044)	(0.020)	(0.046)	(0.041)	(0.021)	(0.046)
$\beta_{DiD,2013}$	0.176***	0.147***	0.088***	0.168***	0.147***	0.080**	0.151***	0.135***	0.067*	0.120**	0.128***	0.060	0.103**	0.118***	0.042
	(0.047)	(0.022)	(0.033)	(0.049)	(0.023)	(0.035)	(0.049)	(0.024)	(0.037)	(0.050)	(0.024)	(0.038)	(0.052)	(0.025)	(0.039)
$\beta_{DiD,2014}$	0.273***	0.169***	0.167***	0.255***	0.160***	0.155***	0.235***	0.137***	0.148***	0.198***	0.118***	0.149***	0.187***	0.100***	0.135***
	(0.052)	(0.029)	(0.042)	(0.054)	(0.031)	(0.043)	(0.054)	(0.031)	(0.045)	(0.056)	(0.031)	(0.047)	(0.058)	(0.032)	(0.050)
$\beta_{DiD,2015}$	0.237***	0.139***	0.180***	0.222***	0.122***	0.164***	0.203***	0.102***	0.170***	0.153**	0.075**	0.163***	0.150**	0.065*	0.172***
	(0.062)	(0.035)	(0.051)	(0.065)	(0.036)	(0.053)	(0.067)	(0.037)	(0.055)	(0.068)	(0.037)	(0.057)	(0.070)	(0.039)	(0.059)
β_{DiD}	0.259***	0.143***	0.165***	0.239***	0.130***	0.142***	0.220***	0.112***	0.130***	0.177***	0.091***	0.122***	0.151***	0.075***	0.107**
	(0.048)	(0.026)	(0.039)	(0.050)	(0.027)	(0.041)	(0.050)	(0.027)	(0.042)	(0.050)	(0.027)	(0.043)	(0.051)	(0.028)	(0.044)
# Observations	27,950	27,950	27,950	27,666	27,666	27,666	27,491	27,491	27,491	27,346	27,346	27,346	27,144	27,144	27,144
Amadeus data	Threshold 10%			Threshold 15%			Threshold 20% (Baseline)			Threshold 25%			Threshold 30%		
$\beta_{DiD,2008}$	-0.125*	-0.073*	-0.013	-0.112	-0.064	-0.001	-0.092	-0.059	0.026	-0.051	-0.026	0.063	-0.019	-0.013	0.057
	(0.065)	(0.041)	(0.070)	(0.070)	(0.044)	(0.075)	(0.074)	(0.046)	(0.079)	(0.074)	(0.046)	(0.081)	(0.078)	(0.049)	(0.086)
$\beta_{DiD,2009}$	-0.034	-0.043	-0.030	-0.016	-0.046	-0.021	-0.003	-0.038	-0.000	0.050	-0.015	0.034	0.083	-0.014	0.035
	(0.056)	(0.036)	(0.058)	(0.060)	(0.038)	(0.062)	(0.063)	(0.039)	(0.065)	(0.063)	(0.040)	(0.066)	(0.066)	(0.042)	(0.069)
$\beta_{DiD,2010}$	-0.066	-0.015	0.002	-0.020	0.007	0.029	-0.028	0.003	0.035	-0.002	0.009	0.034	0.014	0.005	0.034
	(0.055)	(0.031)	(0.050)	(0.056)	(0.032)	(0.052)	(0.060)	(0.034)	(0.055)	(0.061)	(0.035)	(0.057)	(0.063)	(0.036)	(0.059)
$\beta_{DiD,2011}$	-0.062	-0.001	-0.008	-0.042	0.012	0.015	-0.057	0.009	0.012	-0.014	0.026	0.027	0.006	0.037	0.034
	(0.046)	(0.023)	(0.042)	(0.049)	(0.025)	(0.045)	(0.051)	(0.026)	(0.047)	(0.050)	(0.027)	(0.047)	(0.050)	(0.028)	(0.047)
$\beta_{DiD,2013}$	0.155***	0.105***	0.108***	0.147***	0.108***	0.098***	0.147***	0.110***	0.098***	0.147***	0.106***	0.098**	0.132***	0.102***	0.080**
	(0.036)	(0.021)	(0.033)	(0.038)	(0.022)	(0.035)	(0.038)	(0.023)	(0.036)	(0.040)	(0.024)	(0.038)	(0.042)	(0.025)	(0.039)
$\beta_{DiD,2014}$	0.236***	0.126***	0.164***	0.221***	0.115***	0.143***	0.215***	0.110***	0.143***	0.210***	0.100***	0.148***	0.194***	0.087***	0.133***
	(0.046)	(0.028)	(0.042)	(0.049)	(0.029)	(0.043)	(0.050)	(0.030)	(0.045)	(0.053)	(0.031)	(0.047)	(0.055)	(0.033)	(0.050)
$\beta_{DiD,2015}$	0.203***	0.108***	0.201***	0.178***	0.084**	0.174***	0.172***	0.080**	0.190***	0.158**	0.059	0.188***	0.155**	0.058	0.199***
	(0.058)	(0.035)	(0.053)	(0.062)	(0.037)	(0.055)	(0.064)	(0.038)	(0.058)	(0.067)	(0.040)	(0.060)	(0.070)	(0.042)	(0.061)
β_{DiD}	0.243***	0.134***	0.159***	0.210***	0.116***	0.128***	0.206***	0.113***	0.124***	0.171***	0.088***	0.110**	0.142***	0.079***	0.101**
	(0.043)	(0.026)	(0.043)	(0.044)	(0.027)	(0.044)	(0.046)	(0.028)	(0.047)	(0.047)	(0.028)	(0.048)	(0.048)	(0.029)	(0.049)
# Observations	33,301	33,301	33,301	32,948	32,948	32,948	32,715	32,715	32,715	32,504	32,504	32,504	32,257	32,257	32,257

Notes: In this exercise, we check the robustness of the baseline results from Table 3 on the firm-level impact of the 2013 amendment to the income tax break law of 2001. The samples in this table differ in two ways. The upper panel uses the administrative data, while the lower panel uses the Amadeus data. Columns differ in the threshold above which we consider a firm to be treated by the 2013 amendment, i.e., instead of allowing $Exempted_{isk}$ to become 1 only when the firm passes above the 20% threshold of exempted employees, we change this threshold to 10, 15, 25 and 30%. In each sample firms that have above 5% exempted employees post-2013 but under the given threshold are excluded. Robust standard errors, clustered at the firm-level, in parenthesis. All specifications include firm and sector-by-year fixed effects.***,**,* denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B7: Difference-in-Difference Around 2013 Reform Tax Break: Robustness to the Data Source for the Same Sample of Firms

Data source	Administrative			Amadeus			Administrative			Amadeus		
Sample	Administrative			Amadeus			Amadeus			Amadeus		
Regr. period	2011–2015			2008-2015			2011–2015			2011–2015		
Dep. var.	Revenues	Workers	Assets	Revenues	Workers	Assets	Revenues	Workers	Assets	Revenues	Workers	Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\beta_{DiD,2008}$				-0.092	-0.059	0.026						
				(0.074)	(0.046)	(0.079)						
$\beta_{DiD,2009}$				-0.003	-0.038	-0.000						
				(0.063)	(0.039)	(0.065)						
$\beta_{DiD,2010}$				-0.028	0.003	0.035						
				(0.060)	(0.034)	(0.055)						
$\beta_{DiD,2011}$	-0.054	0.031	-0.005	-0.057	0.009	0.012	-0.008	0.030	0.014	-0.035	0.010	-0.003
	(0.043)	(0.019)	(0.046)	(0.051)	(0.026)	(0.047)	(0.041)	(0.020)	(0.048)	(0.050)	(0.026)	(0.046)
$\beta_{DiD,2013}$	0.151***	0.135***	0.067*	0.147***	0.110***	0.098***	0.181***	0.133***	0.094***	0.147***	0.123***	0.076**
	(0.049)	(0.024)	(0.037)	(0.038)	(0.023)	(0.036)	(0.046)	(0.026)	(0.036)	(0.036)	(0.023)	(0.034)
$\beta_{DiD,2014}$	0.235***	0.137***	0.148***	0.215***	0.110***	0.143***	0.259***	0.133***	0.140***	0.230***	0.152***	0.126***
	(0.054)	(0.031)	(0.045)	(0.050)	(0.030)	(0.045)	(0.056)	(0.032)	(0.045)	(0.047)	(0.029)	(0.043)
$\beta_{DiD,2015}$	0.203***	0.102***	0.170***	0.172***	0.080**	0.190***	0.225***	0.103**	0.182***	0.195***	0.137***	0.159***
	(0.067)	(0.037)	(0.055)	(0.064)	(0.038)	(0.058)	(0.069)	(0.040)	(0.057)	(0.061)	(0.037)	(0.054)
Adjusted R^2	0.848	0.910	0.905	0.792	0.855	0.849	0.822	0.887	0.886	0.831	0.891	0.885
# Observations	27,491	27,491	27,491	32,715	32,715	32,715	22,459	22,459	22,459	22,459	22,459	22,459
# Firms	6,146	6,146	6,146	4,837	4,837	4,837	4,837	4,837	4,837	4,837	4,837	4,837

Notes: In this exercise we study the firm-level impact of the 2013 expansion to the income tax break law of 2001. We present both the baseline results on the administrative in columns (1)-(3) and their robustness to an alternative data source and sample of firms. The outcome variables used are log revenues, log number of workers, and log total assets. There are three differences between each group of four columns (columns (1)-(3), (4)-(6), (7)-(9), (10)-(12)). The first difference stems from the source of the data over which we estimate the model in Equation (2). Columns (1)-(3) and (7)-(9) use the administrative data described in Section 3.2. Results from columns (1)-(3) are the same as those from columns (1)-(3) in Table 3. Columns (4)-(6) and (10)-(12) use data from Amadeus, a Bureau van Dijk product. Results from columns (4)-(6) are the same as those from columns (4)-(6) in Table 3. The second difference stems from the sample of firms used in the estimation. Columns (1)-(3) use the larger sample of 6,146 firms (called the "administrative sample"). Columns (4)-(12) use the sub-sample of these 6,146 firms (4,837 firms) which were found in the Amadeus data (called the "Amadeus sample"). The third difference stems from the years over which the regression is run. Columns (1)-(3) and (7)-(12) use the 2011–2015 period, whereas columns (4)-(6) use the 2008–2015 period. Firms in the baseline control group are part of high-tech knowledge-intensive service sectors (as classified by the Eurostat). Robust standard errors, clustered at the firm-level, in parenthesis. All specifications include firm and sector-by-year fixed effects. ***,**,* denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B8: Difference-in-Difference Around 2013 Reform Tax Break: Robustness of the Baseline Results to the Definition of Treatment. Here Treatment Based on Firm Sector

DiD specification	2013, Eq. (2)			2001, Eq. (1)			2001, Eq. (1)		
Threshold restr.	Baseline 2013			Baseline 2013			Under 5% pre-2013		
	Revenues	Workers	Assets	Revenues	Workers	Assets	Revenues	Workers	Assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\beta_{DiD,2011}$	-0.054	0.031	-0.005	-0.006	0.000	-0.013	-0.019	-0.006	-0.025
	(0.043)	(0.019)	(0.046)	(0.021)	(0.009)	(0.021)	(0.020)	(0.009)	(0.021)
$\beta_{DiD,2013}$	0.151***	0.135***	0.067*	0.039	0.040***	-0.022	0.059**	0.054***	-0.010
	(0.049)	(0.024)	(0.037)	(0.025)	(0.011)	(0.018)	(0.024)	(0.011)	(0.018)
$\beta_{DiD,2014}$	0.235***	0.137***	0.148***	0.096***	0.081***	0.007	0.125***	0.104***	0.025
	(0.054)	(0.031)	(0.045)	(0.031)	(0.015)	(0.024)	(0.030)	(0.015)	(0.024)
$\beta_{DiD,2015}$	0.203***	0.102***	0.170***	0.074**	0.096***	0.006	0.102***	0.123***	0.021
	(0.067)	(0.037)	(0.055)	(0.035)	(0.017)	(0.029)	(0.034)	(0.017)	(0.028)
Adjusted R^2	0.848	0.910	0.905	0.847	0.910	0.904	0.848	0.908	0.904
# Observations	27,491	27,491	27,491	27,491	27,491	27,491	28,856	28,856	28,856
# Firms	6,146	6,146	6,146	6,146	6,146	6,146	6,437	6,437	6,437

Notes: In columns (4) to (9) of this table, we study the effects of the 2013 expansion to the income tax break law of 2001 using the same DiD design used to study the initial introduction of the law in 2001. Specifically, treatment is assigned to firms based on whether their sector was eligible for the income tax break law or not, interacted with a dummy activated when $t \geq 2013$ (as in Equation (1), for year 2001), as opposed to their workforce exemption status. Columns (1)-(6) use the same baseline sample as in columns (1)-(3) from Table 3. In columns (7)-(9), to ensure comparability with the baseline sample, we only keep firms with under 5% exemption rate before 2013 – irrespective of their post-2013 exemption rate. For reference, in columns (1)-(3), we reproduce columns (1)-(3) of Table 3, estimated using the specification in Equation (2). In this table, we use the baseline control group, which includes firms in high-tech knowledge-intensive service sectors (as classified by the Eurostat). Also, in this table, we focus on the administrative data. The outcome variables used are log revenues, log number of workers, and log total assets. Robust standard errors, clustered at the firm-level, in parenthesis. All specifications in columns (4)-(9) include firm, year, and sector fixed effects. Note that sector-by-year fixed effects cannot be included in those columns, as in Equation (2). ***,**,* denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B9: Difference-in-Differences Design Around 2013 Reform: Robustness of the Results to the Definition of Treatment, Sample and Data Source Used. Outcome Variable: Log Number of Workers

Data source	Administrative data				Amadeus data			
DiD specification	2001, Eq. (1)	2001, Eq. (1)	2001, Eq. (1)	2013, Eq. (2)	2001, Eq. (1)	2001, Eq. (1)	2001, Eq. (1)	2013, Eq. (2)
Threshold restr.	None	Under 5% pre-2013	Baseline 2013	Baseline 2013	None	Under 5% pre-2013	Baseline 2013	Baseline 2013
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\beta_{DiD,2008}$					-0.044**	0.027	0.056**	-0.059
					(0.018)	(0.023)	(0.024)	(0.046)
$\beta_{DiD,2009}$					-0.033**	0.022	0.045**	-0.038
					(0.016)	(0.020)	(0.020)	(0.039)
$\beta_{DiD,2010}$					-0.016	0.029*	0.042**	0.003
					(0.014)	(0.017)	(0.017)	(0.034)
$\beta_{DiD,2011}$	-0.013	-0.006	0.000	0.031	-0.009	0.000	0.010	0.009
	(0.008)	(0.009)	(0.009)	(0.019)	(0.010)	(0.013)	(0.013)	(0.026)
$\beta_{DiD,2013}$	0.049***	0.054***	0.040***	0.135***	0.034***	0.043***	0.039***	0.110***
	(0.010)	(0.011)	(0.011)	(0.024)	(0.009)	(0.011)	(0.011)	(0.023)
$\beta_{DiD,2014}$	0.093***	0.104***	0.081***	0.137***	0.073***	0.100***	0.087***	0.110***
	(0.013)	(0.015)	(0.015)	(0.031)	(0.012)	(0.015)	(0.015)	(0.030)
$\beta_{DiD,2015}$	0.120***	0.123***	0.096***	0.102***	0.090***	0.122***	0.104***	0.080**
	(0.015)	(0.017)	(0.017)	(0.037)	(0.015)	(0.018)	(0.019)	(0.038)
Adjusted R^2	0.917	0.908	0.910	0.910	0.839	0.856	0.854	0.855
# Observations	38,899	28,856	27,491	27,491	60,640	34,464	32,715	32,715
# Firms	8,837	6,437	6,146	6,146	10,424	5,089	4,837	4,837

Notes: For brevity, all columns in this table report regression results for the outcome variable log number of workers. In all columns for which the "DiD design" line states "2001, Eq. (1)", we study the effects of the 2013 expansion to the income tax break law of 2001 using the same DiD specification used to study the initial introduction of the law in 2001. Specifically, in those columns, treatment is assigned to firms based on whether their sector was eligible for the income tax break law or not, interacted with a dummy activated when $t \geq 2013$ (as in Equation (1), for the year 2001), as opposed to their workforce exemption status. Columns (4) and (8) reiterate the results from implementing the 2013 DiD specification (see Equation (2)) on the administrative data and the Amadeus data (these columns are the same as columns (2) and (5) in Table 3). Columns (1)-(4) have as data source the administrative data. Columns (5)-(8) have as data source Amadeus, Bureau van Dijk. The threshold restriction line reports whether we have imposed any threshold restriction (for the firm-level exemption rate of employees from the personal income tax) on the sample or not. Columns (1) and (5) have imposed no threshold restriction. In columns (2) and (6), to ensure comparability with the baseline samples, we only keep firms with under 5% exemption rate before 2013 – irrespective of their post-2013 exemption rate. Columns (3) and (4) ((7) and (8)) use the baseline administrative (Amadeus) data sample, which imposes that after 2013 treated firms jump over to an over 20% exemption rate of their employees. All specifications in columns (1) to (3) and (5) to (7) include firm, year, and sector fixed effects. Note that sector-by-year fixed effects cannot be included in those columns (in contrast to Equation (2)). Columns (4) and (8) include firm and sector-by-year fixed effects (as they implement Equation (2)). ***,**,* denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

Online Appendix C Additional Sector-Level Evidence

Online Appendix C.1 Descriptive Statistics with Sector-Level Data

Online Appendix C.1.1 Sector Code Switches of Existing Firms to the IT Sector

One channel through which the income tax break law could have boosted the growth of the Romanian IT sector is by changing incentives of firms (workers) to operate in the eligible sector. Using a sector-level transition matrix from the National Bank of Romania, we study transitions into the eligible sector (NACE Rev 1 code 722) before and after the introduction of the 2001 income tax break. While this evidence is descriptive, it highlights abnormal trends in transitions after 2001.

Figure C1 plots the absolute number of firms alive in year $(t-1)$ and with an economic activity different from 722, which switched in year t to the NACE Code 722. Figure C2 plots the year t share of firms in NACE Code 722 alive in year $(t-1)$ coming from firms with a different economic activity in $(t-1)$ that have switched their year t economic activity to 722. We notice that while switching one's sector towards 722 was a trend occurring both before and long after the passing of the 2001 law, immediately after 2001, switching to NACE Rev 1 sector 722 became a visibly more prevalent practice. From 2004 onward, the momentum of this switching practice is lost, which suggests that most justifiable switches occurred immediately after the passage of the law.

One caveat to this exercise is that we cannot pin down the exact reasons behind these switches in firms' main sector code. Hence, this type of growth of the sector may be to some extent artificial, if firms switching their main sectoral activity were already conducting most of their activity in the eligible sector, but had a misassigned main sector. As the law made the income tax break eligibility dependent on a firm's main sector code, the law may have incentivized corrections in firms' sector codes.

Most switches in firms' main sector code occurred from sector codes complementary to the eligible sector (e.g., 721 - "Consulting in the field of computing equipment / hardware" and 726 - "Other activities related to computer science"). We conjecture that these switching firms were multi-activity firms that decided to focus on software development once the income tax break for programmers became available. One might be concerned that these sector code switches were meant to deceit tax authorities. We do not believe that this concern is justified because the sector of the firm was only one of several strict requirements for a worker to qualify for the exemption. In particular, the firm had to provide thorough evidence that revenues of at least 10,000 U.S. dollars per exempted employee came from software development and that employees involved in this software development had eligible bachelor degrees.

For these reasons, the evidence in Figures C1 and C2 suggests that switches were likely to be motivated by an actual switch of focus of the firm on software development. Even if part of these switches may not have resulted in a stronger focus on software development (had some of these firms had a previously misassigned main sector), it is plausible to expect that the income tax break law has strengthened incentives to focus on software development. Moreover, the stark jump in switches around 2001 is a convincing evidence of the salience of the law.

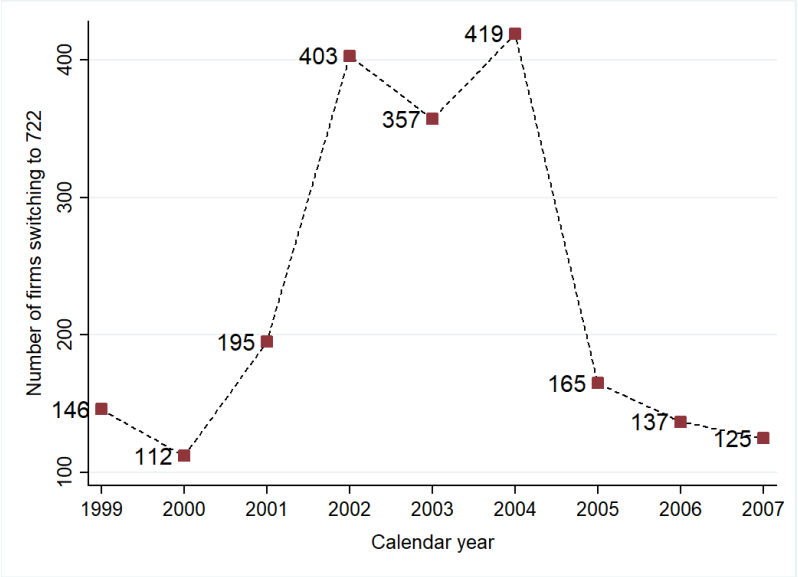

Figure C1: Romania: Number of firms alive in year $(t-1)$ and switching their economic activity in t towards the beneficiary sector, 722

Notes: Data source: National Bank of Romania. This graph plots the number of firms alive in year $(t-1)$ (with an economic activity different from 722) that switched in year t to the NACE Code 722 (7221 and 7222, software consultancy and supply). We notice two peaks in this practice in 2002 and 2004, the years after the first and second laws of income tax exemption for workers in sector 722 were passed in Romania.

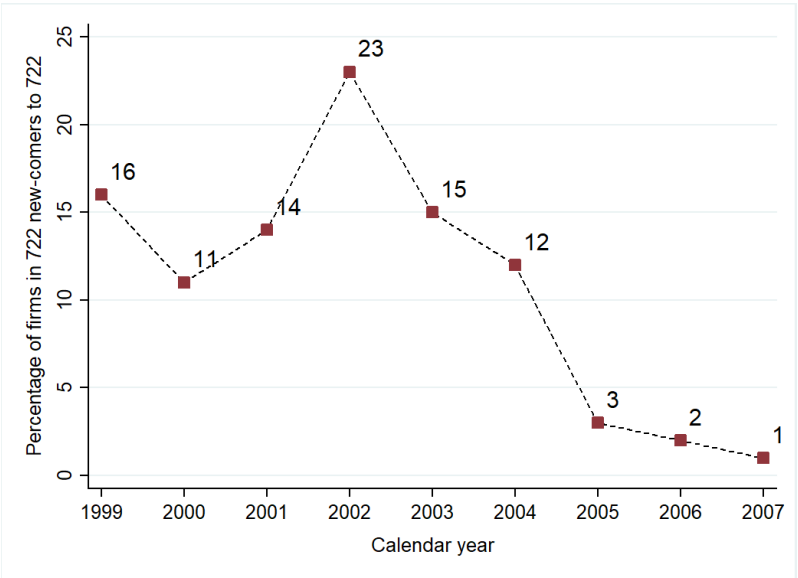

Figure C2: Romania: Among firms in NACE code 722 in year t and alive in year $(t-1)$ (with any economic activity), the percentage of new-comers in t to 722

Notes: Data source: National Bank of Romania. This graph plots the year t share of firms in NACE Code 722 (7221 and 7222, software consultancy and supply) alive in year $(t-1)$ coming from firms with a different economic activity in $(t-1)$ that have switched their year t economic activity to 722. We notice a clear peak in this practice in 2002, the year after the first law of income tax exemption for workers in sector 722 was passed in Romania.

Online Appendix C.1.2 Firm Entries and Exits in the IT Sector

The income tax break may have also affected the growth of the software sector by increasing the birth rate of new firms in the sector and/or decreasing firm death rates. In this section, we provide descriptive evidence on the evolution of firm birth and death rates in the sector (722, in NACE Rev 1.1 classification) in the years before and after the introduction of the policy. The birth/entry (death/exit) rate at time t is defined as the number of firm births/entries (deaths/exits) relative to the population of active firms at the beginning of time t. Our analysis is based on sector-level data from the Eurostat Business Demography database. Figures C3 and C4 show the evolution of birth and death rates in the software sector between 2000 and 2007 for Romania and four other comparable Central and Eastern European (CEE) countries. These other countries were chosen based on data availability for the entire 2000 to 2007 period and on their similarity to Romania (i.e., they are transition economies, at similar stages of development, and with similar economic structures).

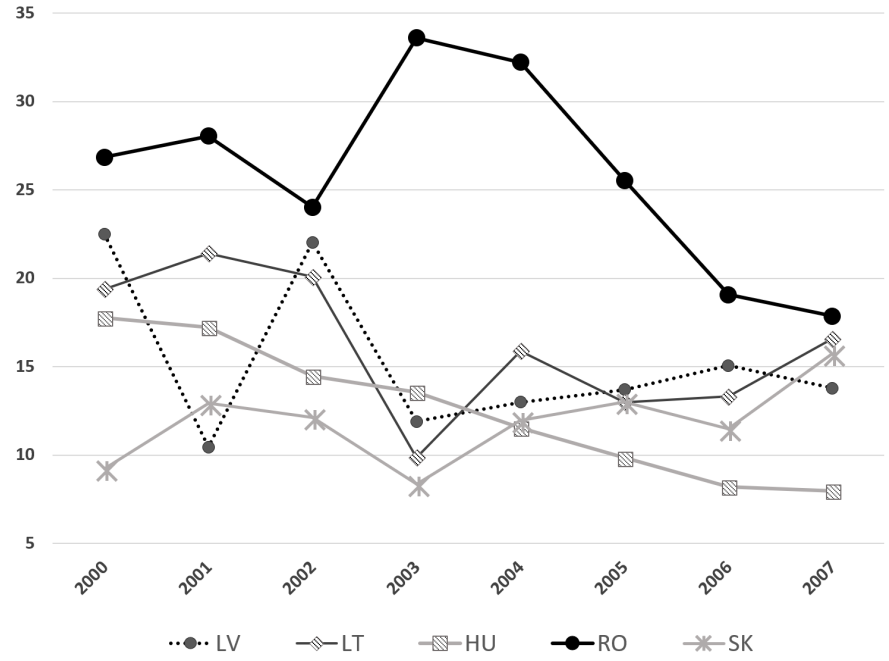

Figure C3: Entry Rates of Firms in the "Software Consultancy and Supply" Sector in Romania and Comparable Countries from Central and Eastern Europe

Notes: Data source: Sector-level data from the Eurostat Business Demography database. The entry rate in year t in a given country is defined as the number of firm entries (births) relative to the population of active firms in that country at the beginning of t.

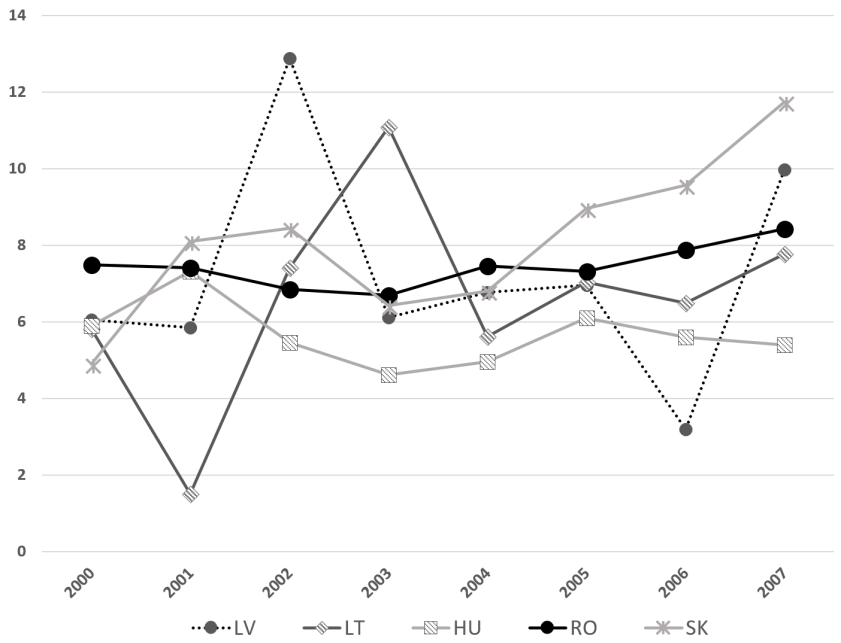

Figure C4: Death Rates of Firms in the "Software Consultancy and Supply" Sector in Romania and Comparable Countries from Central and Eastern Europe

Notes: Data source: Sector-level data from the Eurostat Business Demography database. The exit rate in year t in a given country is defined as the number of firm exits (deaths) relative to the population of active firms in that country at the beginning of t.

Figure C3 shows that the birth rate of firms in the IT sector in Romania experienced a notable increase in 2003 and 2004 relative to its pre-tax break levels.[5] No similar increase can be observed for any of the other four CEE countries. There is a two-year lag in the peak of the firm birth rate in Romania, which is consistent with a need for entrepreneurs to be reassured that the tax break was not to be reverted in the short-term. Figure C4 shows the evolution of firm death rates in the IT sector in Romania and in comparable CEE countries. While the death rate in Romania does not seem to be affected by the introduction of the tax break, throughout the period of analysis Romanian firm death rates remain less volatile than those in comparable countries. The combined trends in firm birth and death rates lead to an increasing trend in the stock of firms in the IT sector in Romania.

Online Appendix C.1.3 Firm Entry through Foreign Direct Investment in IT

In terms of foreign ownership in the IT sector of Romania, the early 2000s were the turning point. Between 2000 and 2004, the first multinational firms decided to offshore part of their operations to Romania (Pruna and Soleanicov, 2012). Moreover, during the same period, there were also smaller foreign companies that acquired Romanian firms, e.g., Adobe Systems Inc. who acquired InterAKT (Pruna and Soleanicov, 2012). Ever since, Romania has continued attracting steady FDI inflows into the IT sector, to the point that in 2017 the biggest players in the IT sector of Romania were multinationals (Oracle, IBM, Ericsson, and Endava). Bitdefender is the only Romanian firm with comparable operating

[5] As shown in Online Appendix C.1.3, from 2003 to 2016, the stock of FDI in the IT sector of Romania has been multiplied by twenty. A sizable share of the firm entry into the IT sector is likely to be due to foreign firm entry.

revenue from its operations in Romania.[6] In 2017 foreign-owned companies produced 73% of the gross revenues and employed 59% of the total number of workers in the Software and IT Services of Romania.

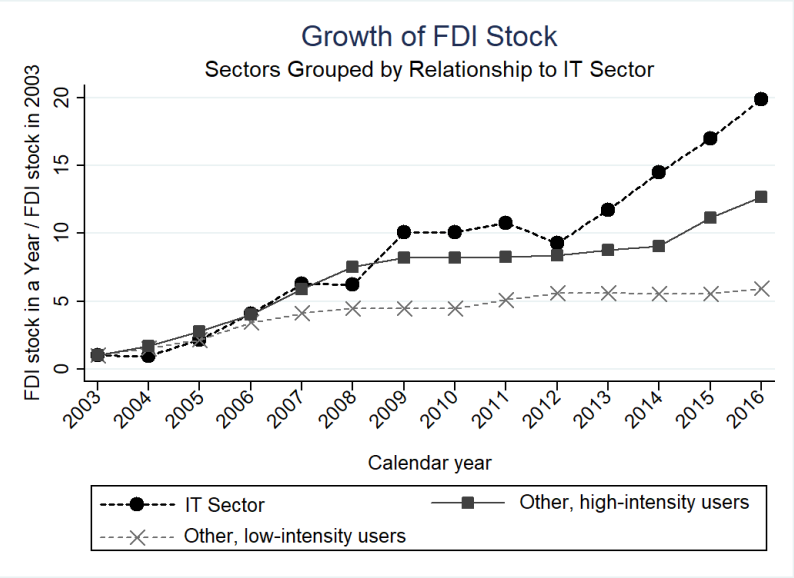

Figure C5: Growth of the Stock of FDI in Romania, by Relationship to the IT Sector

Notes: Data sources: National Bank of Romania (for historical sector-level FDI stocks) and OECD (for the Romanian input-output table for the year 2000, at the two-digit NACE Rev 1 level). This figure plots the growth of the FDI stock in Romania between 2003 and 2016 (defined as the yearly value of the FDI stock divided by the relevant FDI stock value in 2003) for three categories of sectors. The first category is the sector 72 ("Computer and related activities") itself. The second category called "Other, high-intensity users" contains all two-digit NACE Rev 1 sectors that are in the top quarter of the distribution of sector-level shares of input purchases from sector 72 in a sector's total input expenditure (except sector 72). These shares are computed based on Romania's input-output table for the year 2000. The last category of sectors called "Other, low-intensity users" contains all other sectors for which IT inputs are less important. The following are the NACE Rev 1 codes and names of the sectors in the "Other, high-intensity users" category: 50-52 (Wholesale and retail trade; repairs), 60-63 (Transport and storage), 70 (Real estate activities), 71 (Renting of machinery and equipment), 73-74 (R&D and other business activities), 75 (Public administration and defence; compulsory social security), 80 (Education), and 85 (Health and social work).

An ideal dataset to study the causal effects of the income tax break for workers in IT on Foreign Direct Investment (FDI, henceforth) into the IT sector of Romania would have to start before 2001, include not only Romania but also comparable countries, and be disaggregated at the sector-level. Unfortunately, we have not found such a dataset. For Romania, the longest sector-level FDI time series (from the National Bank of Romania) starts in 2003. With this data, we study the growth of the FDI stock for three categories of sectors: the IT sector itself, the group of sectors that are "high-intensity users" of IT services, and the "low-intensity users".[7] Figure C5 plots the growth of the FDI stock in Romania between 2003 and 2016 (defined as the yearly value of the FDI stock divided by the relevant FDI stock value in 2003) for these three categories of sectors. The figure shows that sectors with the highest share of IT services in their input expenditure are those experiencing the highest growth of their FDI stock since 2003. The IT sector itself is experiencing an even starker growth in its FDI stock. The importance of foreign revenues has also been growing steadily over the years, to the point that in 2017 total foreign revenues were three times larger than domestic revenues.[8]

[6] See December 2017 article from the *Romanian Journal*.
[7] See Section 4.2 for details on how we assign sectors into either the high- or low-usage groups.
[8] Source of last two sets of statistics: the Association of Employers in the Software and Services Industry or ANIS.

Online Appendix C.2 Additional Robustness Checks on the Sector-Level Cross-Country Effects of the 2001 Tax Break

Online Appendix C.2.1 IT Sector Vs. Rest of the Economy

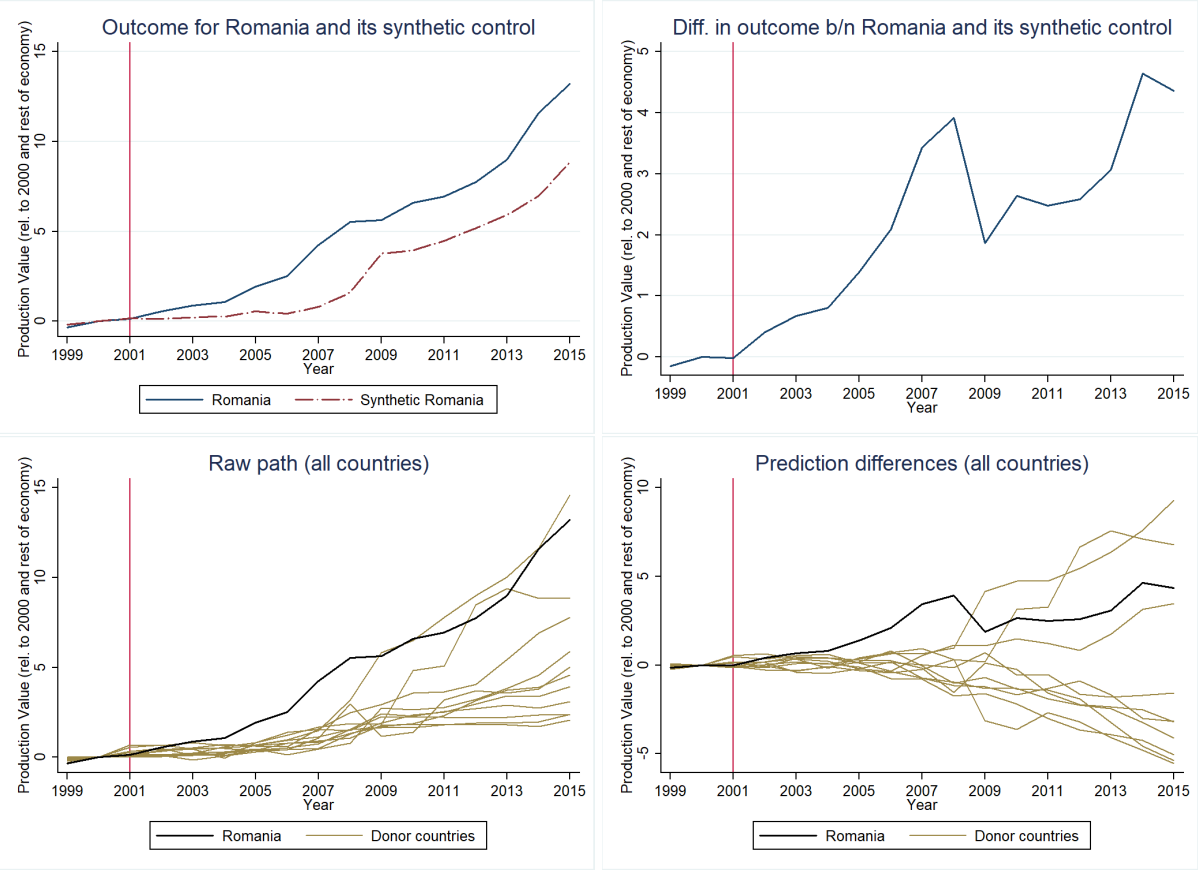

Figure C6: IT Sector Vs. Rest of Economy. SCM with Outcome Variable: "Production Value" - Million Euros (Normalized)

Notes: In this exercise we use the synthetic control method introduced in Section 4.1 to study the sector-level direct effects of the introduction of the 2001 law granting an income tax break to workers in IT. All figures have as dependent variable the country-level (normalized) " Production value - million euros" The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sector is K72 ("Computer and related activities," including "Software consultancy and supply " and "Publishing of software"). We use as comparison sectors all other sectors in the economy (all except K72). The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

Table C1: Post-treatment Results: Effects, p-values, Standardized p-values. Outcomes: Gross Revenues, Employment, and Production Value

	Gross Revenues			**Employment**			**Production Value**		
Post-treatment year	$\hat{\alpha}_{1t}$	p-values	Standardized p-values	$\hat{\alpha}_{1t}$	p-values	Standardized p-values	$\hat{\alpha}_{1t}$	p-values	Standardized p-values
2001	0.13	0.27	0.55	0.14	0.27	0.00	-0.01	1.00	1.00
2002	0.69	0.00	0.27	0.26	0.09	0.00	0.41	0.10	0.50
2003	1.13	0.00	0.27	0.67	0.09	0.00	0.67	0.00	0.80
2004	1.37	0.00	0.27	1.13	0.00	0.00	0.80	0.00	0.30
2005	1.95	0.00	0.36	1.46	0.00	0.00	1.39	0.00	0.40
2006	2.57	0.00	0.36	1.75	0.00	0.00	2.10	0.00	0.50
2007	3.17	0.00	0.36	2.10	0.00	0.00	3.43	0.00	0.30
2008	3.21	0.00	0.27	1.90	0.00	0.00	3.92	0.00	0.40
2009	4.14	0.09	0.27	1.17	0.27	0.00	1.87	0.20	0.70
2010	4.49	0.09	0.27	1.03	0.45	0.00	2.64	0.30	0.70
2011	2.92	0.36	0.27	1.12	0.36	0.00	2.48	0.40	0.70
2012	3.32	0.45	0.36	1.19	0.18	0.00	2.59	0.40	0.90
2013	4.52	0.36	0.36	1.32	0.18	0.00	3.06	0.50	0.90
2014	6.15	0.36	0.36	1.54	0.18	0.00	4.64	0.30	0.70
2015	6.52	0.27	0.36	1.83	0.18	0.00	4.36	0.50	0.80

Notes: In this exercise we use the synthetic control method (SCM) introduced in Section 4.1 to study the sector-level direct effects of the introduction of the 2001 law granting an income tax break to workers in IT. This table presents the SCM estimates of the treatment effects ($\hat{\alpha}_{1t}$) of the 2001 policy on three outcome variables (relative gross revenues, relative employment, and the relative production value) in Romania (where 1=Romania). This table corresponds to the treatment effects plotted in Figures 3, 4, and C6. Treatment effects, $\hat{\alpha}_{1t}$, for Romania in post-treatment years t (2001 to 2015) are estimated as the difference between the year t outcomes for Romania and those for synthetic Romania (for details see Equation (3) in Section 4.1). The p-values are constructed using permutation tests. To calculate a p-value for each post-treatment period, the **synth_runner** package for Stata (Quistorff and Galiani, 2017) finds the proportion of placebo effects that are as large as the main effect. The standardized p-values are the studentized p-values. See Table C2 for the combination of countries and weights that comprise the synthetic control.

Table C2: Predictor balance, averaged between 1999 and 2000. Synthetic Romania = 0.104 Czech Republic and 0.896 Slovakia (Gross Revenues); = 0.534 Bulgaria and 0.466 Czech Republic (Employment); = 0.339 Bulgaria and 0.661 Lithuania (Production Value)

	Romania	Synthetic Romania		
		Gross Revenues	Employment	Prod. Value
GDP per capita (constant LCU, $)	9,663	9,796	15,972	5,341
Services (% of GDP)	53.23	60.32	60.07	62.59
High-tech manufacturing (% mfg. value added)	23.85	35.69	34.05	23.85

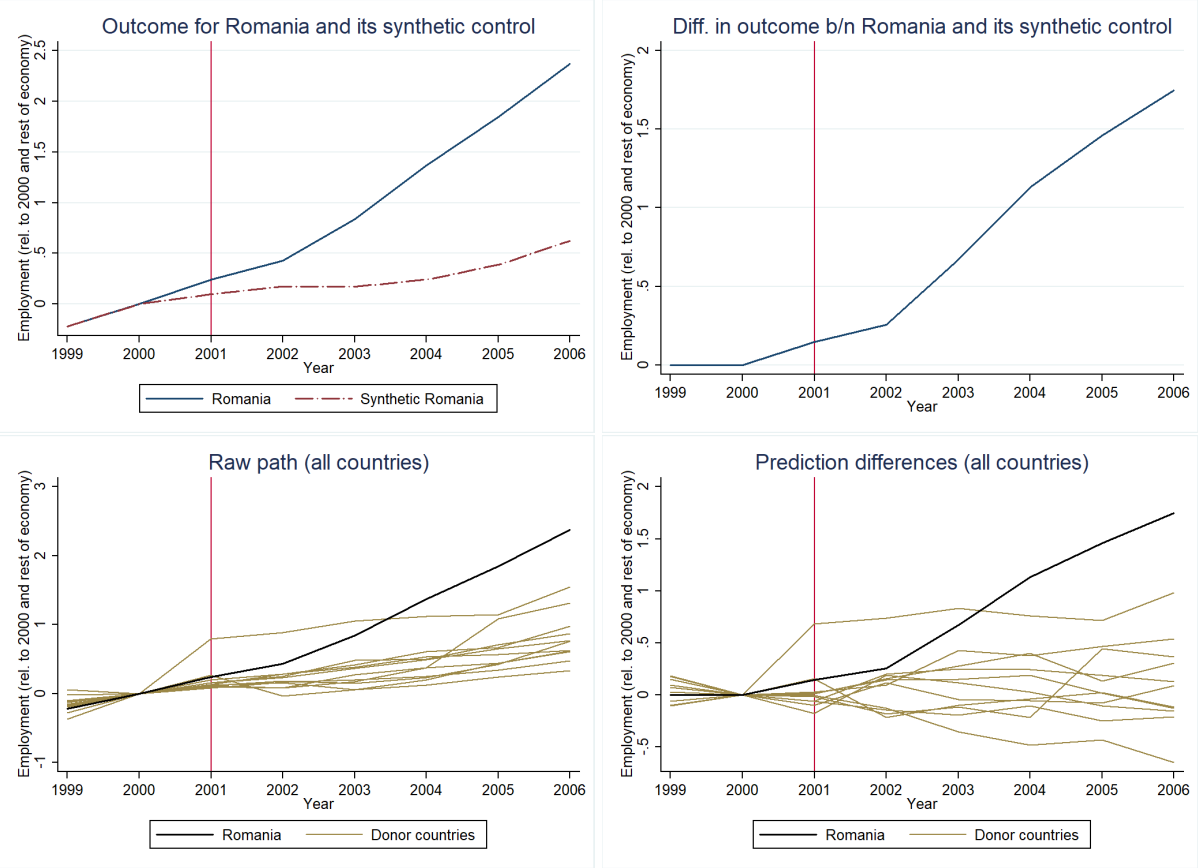

Figure C7: IT Sector Vs. Rest of Economy. SCM with Outcome Variable: "Employees - Number" (Normalized). Robustness to the Correspondence between NACE Sector Classifications

Notes: In this exercise we use the synthetic control method introduced in Section 4.1 to study the sector-level direct effects of the introduction of the 2001 law granting an income tax break to workers in IT. This exercise differs from the one presented in Figure 4 in the period of analysis. Instead of considering the entire 1999 to 2015 period, we cut the analysis in 2006 (the last year where the data is available in the NACE Rev 1 classification). This tests the robustness of our results to the correspondence we develop between the NACE Rev 1 and NACE Rev 2 classifications. All figures have as dependent variable the country-level (normalized) "Employees - number." The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sector is K72 ("Computer and related activities," including "Software consultancy and supply " and "Publishing of software"). We use as comparison sectors all other sectors in the economy (all except K72). The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

Online Appendix C.2.2 IT-Using Sectors Vs. Non-IT Using Sectors

Table C3: sectors with Consistently-Available Data in Eurostat, by Quarter of IT-use Intensity

Under Quartile 1	**Between Quartiles 1-2**
Wood and products of wood and cork (20)	Manufacture of textiles and textile products (17)
Basic metals (27)	Manufacture of wearing apparel; dressing, dyeing of fur (18)
Fabricated metal products (28)	Manufacture of pulp, paper and paper products (21)
Electricity, gas and water supply (40-41)	Publishing, printing and reproduction of recorded media (22)
	Manufacture of machinery and equipment n.e.c. (29)
	Manufacture of furniture; manufacturing n.e.c. (36)
	Recycling (37)
	Construction (45)
Between Quartiles 2-3	**Over Quartile 3**
Manufacture of chemicals and chemical products (24)	Wholesale and retail trade (50-52)
Manufacture of rubber and plastic products (25)	Transport and storage (60-63)
Manufacture of other non-metallic mineral products (26)	Real estate activities (70)
Manufacture of office machinery and computers (30)	Renting of machinery and equipment (71)
Manufacture of electrical machinery and apparatus (31)	Computer and related activities (72)
Manufacture of radio, television and communication (32)	R&D and other business activities (73-74)
Manufacture of medical, precision, optical instruments (33)	
Manufacture of motor vehicles, trailers, semi-trailers (34)	
Manufacture of other transport equipment (35)	
Hotels and restaurants (55)	

Notes: This table groups sectors with consistently-available data in Eurostat (across variables, years, and countries - see Online Appendix D for details) based on their dependence on the IT sector for inputs. To establish this dependence, we start from the input-output table (I-O table, henceforth) of Romania for the year 2000. We use the harmonized I-O table provided by the OECD, that tracks the flows of goods and services between all two-digit NACE Rev 1 sectors in Romania. We then compute the share of the total input expenditures of a given sector purchased from the IT sector (NACE Rev 1 sector 72, "Computer and related activities"). It is based on this distribution of shares that we compute the quartiles mentioned in this table. In the main specification of the inter-sector SCM analysis we exclude sector 72 altogether from the analysis, as this is the treated sector itself. We also assign sectors over quartile three into the high-intensity category and all other sectors into the low-intensity. Note that the sectors actually used in the analysis and presented in this table are a subset of all sectors in the economy, as not all sectors had consistently-available data. Nevertheless, the grouping of these sectors by quarters was based on the full I-O matrix, which includes all sectors in the economy.

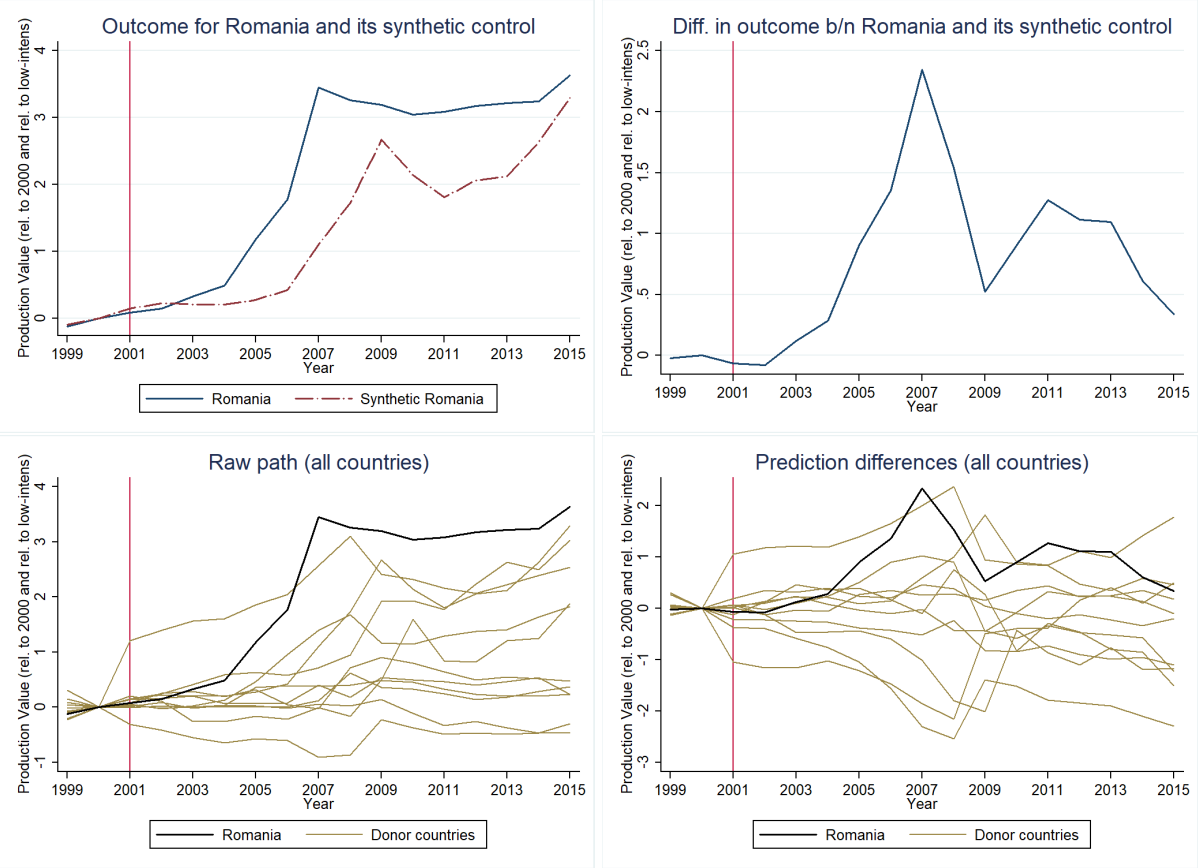

Figure C8: IT-Using Sectors Vs. Non-IT Using Sectors. SCM with Outcome Variable: "Production Value" - Million Euros (Normalized)

Notes: In this exercise we use the synthetic control method introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. All figures have as dependent variable the country-level (normalized) " Production value - million euros." The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sectors are those that use K72 ("Computer and related activities," including "Software consultancy and supply" and "Publishing of software") services at high-intensity. We exclude K72 itself from this category. Sectors that have a low-intensity of use of K72 services serve as comparison sectors. The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

Table C4: Post-treatment Results: Effects, p-values, Standardized p-values. Outcomes: Gross Revenues, Employment, and Production Value

	Gross Revenues			**Employment**			**Production Value**		
Post-treatment year	$\hat{\alpha}_{1t}$	p-values	Standardized p-values	$\hat{\alpha}_{1t}$	p-values	Standardized p-values	$\hat{\alpha}_{1t}$	p-values	Standardized p-values
2001	-0.04	0.80	0.50	0.08	0.22	0.22	-0.07	0.64	0.64
2002	-0.02	1.00	0.50	0.09	0.22	0.11	-0.08	0.91	0.64
2003	0.18	0.30	0.50	0.27	0.22	0.11	0.12	0.82	0.55
2004	0.28	0.00	0.40	0.43	0.22	0.11	0.29	0.55	0.45
2005	0.54	0.00	0.40	0.68	0.00	0.11	0.90	0.27	0.27
2006	0.90	0.00	0.40	0.86	0.00	0.11	1.36	0.27	0.18
2007	1.71	0.00	0.40	0.93	0.00	0.11	2.34	0.00	0.18
2008	1.88	0.00	0.40	0.98	0.11	0.11	1.53	0.36	0.36
2009	1.46	0.00	0.40	0.77	0.00	0.11	0.52	0.45	0.27
2010	1.33	0.00	0.40	0.68	0.00	0.11	0.90	0.09	0.27
2011	1.34	0.00	0.40	0.73	0.00	0.11	1.27	0.09	0.27
2012	1.30	0.10	0.40	0.68	0.00	0.11	1.11	0.18	0.27
2013	0.68	0.30	0.40	0.67	0.00	0.11	1.10	0.09	0.27
2014	0.73	0.30	0.50	0.61	0.22	0.11	0.61	0.45	0.36
2015	0.75	0.50	0.50	0.61	0.11	0.11	0.34	0.73	0.55

Notes: In this exercise we use the synthetic control method (SCM) introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. This table presents the SCM estimates of the treatment effects ($\hat{\alpha}_{1t}$) of the 2001 policy on three outcome variables (relative gross revenues, relative employment, and the relative production value) in Romania (where 1=Romania). This table corresponds to the treatment effects plotted in Figures 5, 6, and C8. Treatment effects, $\hat{\alpha}_{1t}$, for Romania in post-treatment years t (2001 to 2015) are estimated as the difference between the year t outcomes for Romania and those for synthetic Romania (for details see Equation (3) in Section 4.1). The p-values are constructed using permutation tests. To calculate a p-value for each post-treatment period, the **synth_runner** package for Stata (Quistorff and Galiani, 2017) finds the proportion of placebo effects that are as large as the main effect. The standardized p-values are the studentized p-values. See Table C5 for the combination of countries and weights that comprise the synthetic control.

Table C5: Predictor balance, averaged between 1999 and 2000. Synthetic Romania = 0.341 Bulgaria and 0.659 Lithuania (Gross Revenues); = 0.363 Bulgaria + 0.002 Hungary + 0.625 Lithuania (Employment) ; = Bulgaria (Production Value)

	Romania	Synthetic Romania		
		Gross Revenues	Employment	Prod. Value
GDP per capita (constant LCU, $)	9,663	5,342	9,616	5,771
Services (% of GDP)	53.23	62.59	62.53	60.85
High-tech manufacturing (% mfg. value added)	23.85	23.87	24.15	30.53

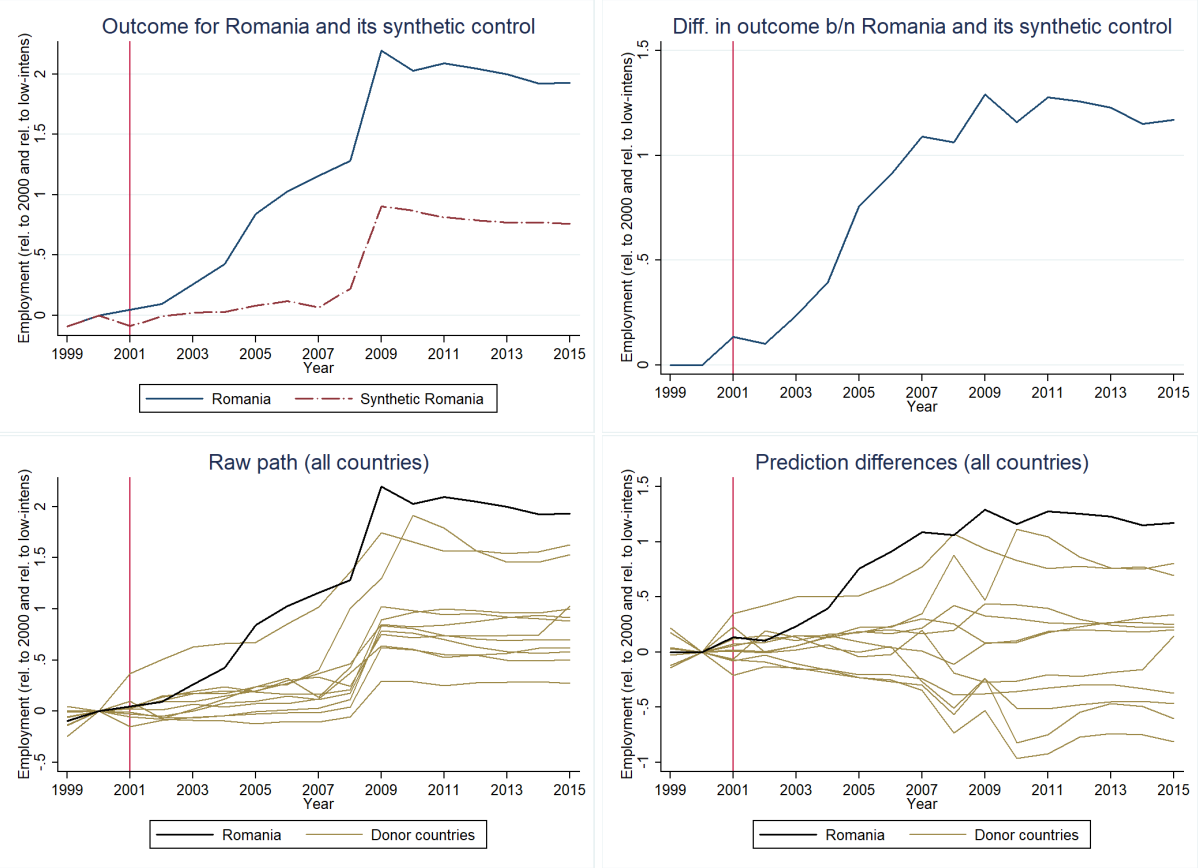

Figure C9: IT-Using Sectors Vs. Non-IT Using Sectors. SCM with Outcome Variable: "Employees - Number" (Normalized). Robustness to Exclusion of Sectors Comparable to IT from High-Intensity Sectors Category

Notes: In this exercise we use the synthetic control method introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. This exercise differs from the one presented in Figure 6 in its exclusion of sectors K73 and K74 from the analysis. Both sectors are otherwise part of the list of sectors that rely heavily on IT services (top 25% users of IT services). We used these sectors as comparison sectors in the firm-level analysis. All figures have as dependent variable the country-level (normalized) "Employees - number." The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sectors are those that use K72 ("Computer and related activities," including "Software consultancy and supply" and "Publishing of software") services at high-intensity. We exclude K72 itself from this category. Sectors that have a low-intensity of use of K72 services serve as comparison sectors. The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

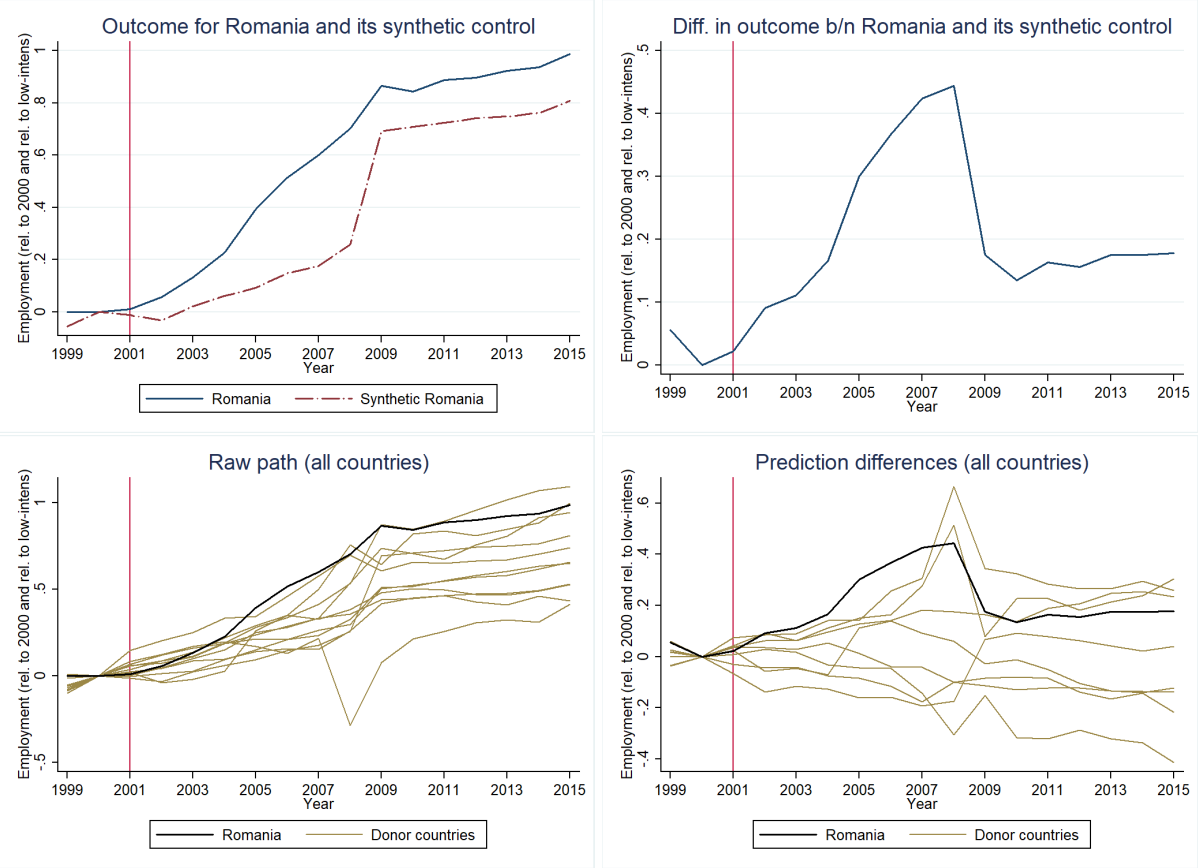

Figure C10: IT-Using Sectors Vs. Non-IT Using Sectors. SCM with Outcome Variable: "Employees - Number" (Normalized). Robustness to Grouping of Quarters into High- and Low-Intensity

Notes: In this exercise we use the synthetic control method introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. This exercise differs from the one presented in Figure 6 in the grouping of quarters into high- and low-intensity IT users. The main results compare the top 25% sectors in terms of IT services use (high-intensity category) with the bottom 75% (low-intensity category). In this exercise we compare the top 50% (high-intensity category) to the bottom 50% (low-intensity category). All figures have as dependent variable the country-level (normalized) "Employees - number." The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sectors are those that use K72 ("Computer and related activities," including "Software consultancy and supply" and "Publishing of software") services at high-intensity. We exclude K72 itself from this category. Sectors that have a low-intensity of use of K72 services serve as comparison sectors. The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

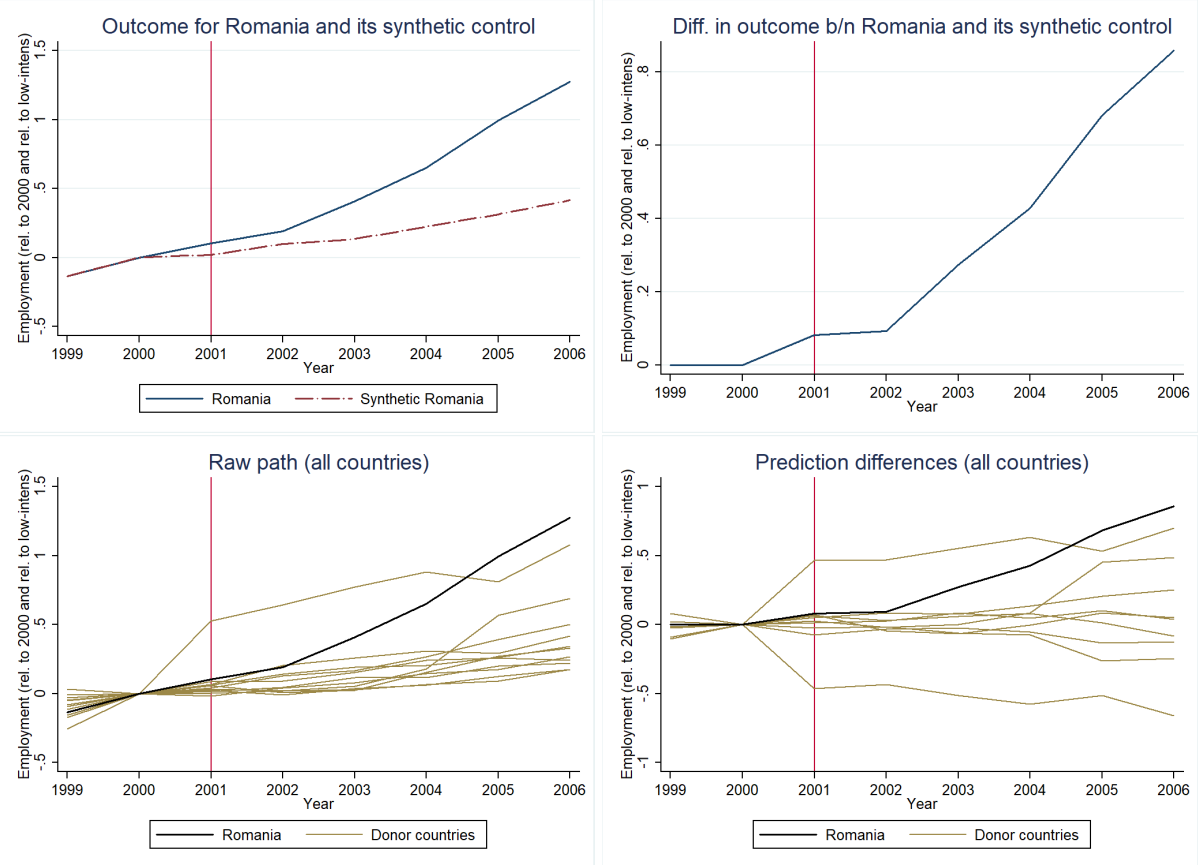

Figure C11: IT-Using Sectors Vs. Non-IT Using Sectors. SCM with Outcome Variable: "Employees - Number" (Normalized). Robustness to the Correspondence between NACE Sector Classifications

Notes: In this exercise we use the synthetic control method introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. This exercise differs from the one presented in Figure 6 in the period of analysis. Instead of considering the entire 1999 to 2015 period, we cut the analysis in 2006 (the last year where the data is available in the NACE Rev 1 classification). This tests the robustness of our results to the correspondence we develop between the NACE Rev 1 and NACE Rev 2 classifications. All figures have as dependent variable the country-level (normalized) "Employees - number." The yearly absolute value of the dependent variable in the treated sector is divided by its value in 2000, the year prior to the introduction of the income tax break in Romania. From these resulting yearly ratios we subtract the corresponding ratios for the comparison sectors. The treated sectors are those that use K72 ("Computer and related activities," including "Software consultancy and supply" and "Publishing of software") services at high-intensity. We exclude K72 itself from this category. Sectors that have a low-intensity of use of K72 services serve as comparison sectors. The data source for the dependent variable is Eurostat, Structural Business Statistics, Annual detailed enterprise statistics on services (NACE Rev 1.1). Data for the predictors comes from the World Bank, World Development Indicators. We use as predictors the "GDP per capita (constant LCU)," "Medium and high-tech sector (% manufacturing value added)" and "Services, etc., value added (% of GDP)." All figures are an output of the **synth_runner** package for Stata (Quistorff and Galiani, 2017), with the *nested* option specified.

Table C6: Post-treatment Results: Effects, p-values, Standardized p-values. Outcome: (Goods Export) Trade Values

Post-treatment year	$\hat{\alpha}_{1t}$	p-values	Standardized p-values
2001	0.02	0.90	0.90
2002	0.03	0.80	0.70
2003	0.16	0.80	0.40
2004	0.53	0.10	0.10
2005	1.03	0.00	0.20
2006	1.60	0.10	0.10
2007	2.30	0.10	0.10
2008	3.64	0.10	0.10
2009	4.76	0.00	0.10
2010	5.96	0.00	0.10
2011	7.46	0.00	0.10
2012	6.79	0.00	0.10
2013	7.69	0.00	0.10
2014	7.70	0.00	0.10

Notes: In this exercise we use the synthetic control method (SCM) introduced in Section 4.2 to study the sector-level downstream effects of the introduction of the 2001 law granting an income tax break to workers in IT. This table presents the SCM estimates of the treatment effects ($\hat{\alpha}_{1t}$) of the 2001 policy on the relative (goods export) trade values in Romania (where 1=Romania). This table corresponds to the treatment effects plotted in Figure 7. Treatment effects, $\hat{\alpha}_{1t}$, for Romania in post-treatment years t (2001 to 2015) are estimated as the difference between the year t outcomes for Romania and those for synthetic Romania (for details see Equation (3) in Section 4.1). The p-values are constructed using permutation tests. To calculate a p-value for each post-treatment period, the **synth_runner** package for Stata (Quistorff and Galiani, 2017) finds the proportion of placebo effects that are as large as the main effect. The standardized p-values are the studentized p-values. See Table C7 for the combination of countries and weights that comprise the synthetic control.

Table C7: Predictor balance, averaged between 1996 and 2000. Synthetic Romania = 0.706 Czech Republic, 0.240 Lithuania, and 0.053 Slovakia

	Romania	
	Actual	Synthetic
GDP per capita (constant LCU, $)	9,771	20,689
Services (% of GDP)	47.89	58.41
High-tech manufacturing (% mfg. value added)	24.33	32.60

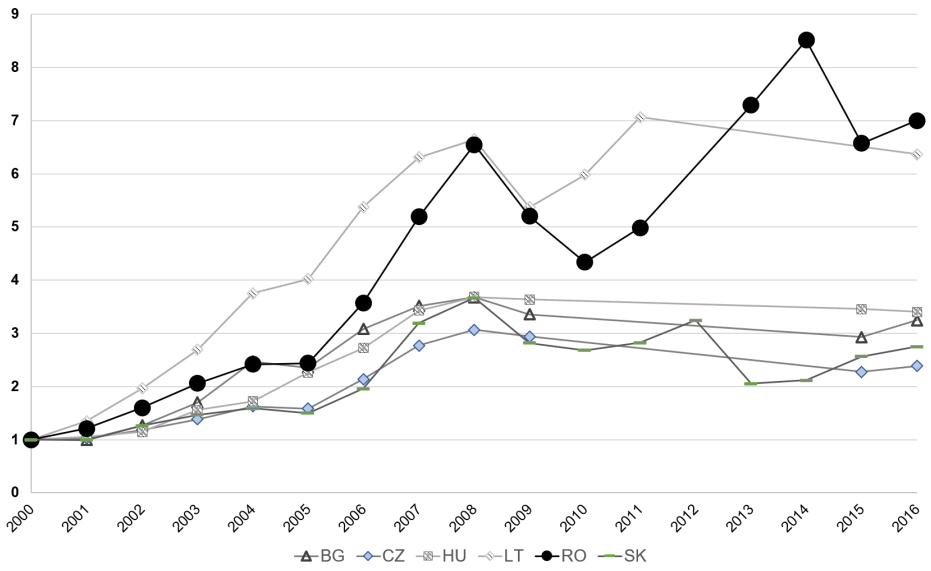

Figure C12: Trends in Export Value of IT-using Service Sectors

Notes: This figure plots the evolution of export trade values for service sectors relying more on IT services. These sectors are NACE Rev 1 sectors 51 (Wholesale trade and commission trade), 60 (Land transport), 62 (Air transport), 63 (Cargo handling and storage), 74 (Other business activities), and 75 (Public administration and defence). In Romania, most of the impressive growth experienced between 2001 and 2016 is explained by sectors under NACE Rev 1 sector 74 (e.g., call centers, advertising, business and management consultancy, secretarial and translation activities etc.). The figure includes Romania and five other countries that appear in SCM exercises as part of synthetic Romania, i.e. Bulgaria, the Czech Republic, Hungary, Lithuania, and Slovakia. This evolution is with respect to the export value in year 2000. This figure uses UN Comtrade data, EBOPS 2002. A similar graph can be made comparing the growth of IT-using service sectors to the growth of less IT-using service sectors. Romania continues to stand out in a relative comparison as well.

Online Appendix D Data Construction for the Synthetic Control

Online Appendix D.1 Eurostat Data

Downloading the Data. The first step in this process was to download data from Eurostat, using the Structural Business Statistics: Annual Detailed Enterprise Statistics tables. The data were available to download in batches, after having selected the sectors, countries, years, and variables of interest. These selections are detailed below:

Sectors: Data were available for download at the one-, two-, three-, and four-digit sector level. For the main analysis, we selected and downloaded data at the two-digit sector level due to high levels of missing data at the three- and four-digit level. According to the Rev 1.1 sector classification, there were 59 possible two-digit sectors for which we might like to have data. Our process enabled us to find corresponding data as early as 1999 for 48 of these sectors. Missing sectors include A01 ("Agriculture, hunting and related service activities"), A02 ("Forestry, logging and related service activities"), B05 ("Fishing, fish farming and related service activities"), L75 ("Public administration and defence; compulsory social security"), M80 ("Education"), N85 ("Health and social work"), O90 ("Sewage and refuse disposal, sanitation and similar activities"), O91 ("Activities of membership organizations n.e.c."), O92 ("Recreational, cultural and sporting activities"), O93 ("Other service activities"), and P95 ("Activities of households as employers of domestic staff"). For each two-digit sector, data were downloaded in two

pieces: firstly as per the Rev 1.1 sector classification and secondly as per the Rev 2 sector classification. Since the transition from Rev 1.1 to Rev 2 in the year 2007 was not one-to-one, fractions of two-digit sectors were carefully linked across the two classifications according to the degree of mutually overlapping three-digit sectors between any pair of two-digit sectors. More detail on this crosswalk follows in the "Creation of Crosswalk between Rev 1.1 and Rev 2 sector Codes" section.

Countries: Data were downloaded for the following fourteen countries: Bulgaria, Cyprus, the Czech Republic, Estonia, Hungary, Ireland, Latvia, Lithuania, Malta, Poland, Portugal, Romania, the Slovak Republic, and Slovenia. Data were widely available for most countries, with frequent missing values for Cyprus and Malta across the sectors, variables, and years of interest.

Variables: Data were downloaded for the following five variables: Gross revenues (Turnover or gross premiums written) - million euros (V12110), Production value - million euros (V12120), Employees - number (V16130), Gross investment in tangible goods - million euros (V15110), and Enterprises - number (V11110).

Years: For each sector, country, and year, relevant data were downloaded for as many years as possible, ranging from 1995 to 2016. There was low data availability prior to 1999 and after 2015 across the set of countries, sectors, and variables of interest.

After having downloaded all available data for five variables from 1995 to 2016 across 14 countries and 48 sectors, there was a further selection of years, countries, and sectors to include in the final analysis based on having a sufficiently high level of non-missing data to complete the analysis. These selections are detailed in the "Selection of Years, Countries, and Sectors for Final Analysis" section.

Creation of Crosswalk between Rev 1.1 and Rev 2 Sector Codes. The SCM analysis compares pre-trends from before and after 2001, between Romania and synthetic Romania, for sectors with high- versus low-intensity of IT use. Since the Rev 1.1 sector classification was in effect for the treatment period of interest, our analysis seeks to classify sectors as either high- or low-intensity of IT use according to their Rev 1.1 sector classification. For this reason, the data downloaded as per Rev 2 sector classifications (for years 2007 onwards) needed to be stitched together with the data downloaded as per Rev 1.1 sector classifications (for years prior to 2007) according to a correspondence or crosswalk between the two sector classifications.

An extensive correspondence detailing equivalence between two-, three-, or four-digit sectors (as applicable and as dictated by the transition) between Rev 1.1 and Rev 2 was obtained from Eurostat and used for constructing a crosswalk that could link sectors across sector classifications at the two-digit level. The revision between Rev 1.1 and Rev 2 was in most cases not a one-to-one transition. For six of the 59 Rev 1.1 sectors, one two-digit Rev 2 sector corresponded to a single two-digit Rev 1.1 sector. Six of the 59 two-digit Rev 1.1 sectors were either split into two or more corresponding two-digit Rev 2 sectors or were part of a merge into one corresponding two-digit Rev 2 sector. The remaining 47 of 59 sectors were combinations of splits and merges, that is, some portion of the two-digit sectors as per Rev 1.1 were split into two or more sectors as per Rev 2, where they were merged with other portions of two-digit Rev 1.1 sectors that were similarly split among two or more Rev 2 sectors.

Two primary methods were used to create a crosswalk to link some fraction of each two-digit Rev

2 sector to a corresponding two-digit Rev 1.1 sector. In the first method ("gross revenues and count"), the average gross revenues across all 14 countries of interest was used as a proxy for the size of each Rev 1.1 sector at the three-digit level. Data for this portion of the task was similarly downloaded from Eurostat, using the Structural Business Statistics: Annual Detailed Enterprise Statistics tables. This information on sector size was used to calculate what portion of a two-digit Rev 2 sector stemmed from distinct three-digit Rev 1.1 sectors, and in turn, to which two-digit Rev 1.1 sectors these three-digit sectors corresponded. This involved calculating the average gross revenues in 2007 across all 14 countries of interest for each three-digit Rev 1.1 sector included in the correspondence table (in cases where the correspondence table indicated linkages at the four-digit level, these were aggregated up to the three-digit level so that all comparisons at this stage were made at the three-digit level). The size of corresponding three-digit sectors common to a pair of two-digit Rev 1.1 and Rev 2 sectors relative to the total size of the Rev 2 sector were used to assign fractions of that two-digit Rev 2 sector to corresponding Rev 1.1 sectors. To illustrate, suppose a particular two-digit Rev 2 sector was composed of five three-digit Rev 2 sectors that corresponded to five three-digit Rev 1.1 sectors, three of which corresponded to one two-digit Rev 1.1 sector (totaling average gross revenues of 200) and two of which corresponded to a second two-digit Rev 1.1 sector (totaling average gross revenues of 300). In this case, 200/500=40% of each variable value from the Rev 2 sector would be allocated to the first Rev 1.1 sector, and 300/500=60% of each variable value from the Rev 2 sector would be allocated to the second Rev 1.1 sector.

The second method, ("count only") used a similar approach. However, instead of calculating the fraction of each two-digit Rev 2 code to assign to each two-digit Rev 1.1 code using an estimate of the size of corresponding three-digit sectors, this method used the count of three-digit sectors common to a pair of two-digit Rev 1.1 and Rev 2 sectors relative to the total number of three-digit sectors that make up the two-digit Rev 2 sector to assign the fractions. In the same example as above, 3/5=60% of each variable value from the two-digit Rev 2 sector would be allocated to the first two-digit Rev 1.1 sector, and 2/5=40% of each variable value would be allocated to the second Rev 1.1 sector. For the first method above, in cases where gross revenues was not available in 2007 in order to make the necessary estimation of sector size, fractions were allocated based on the "count only" method.

The resulting versions of the crosswalk between the two sector classifications are similar, and there is no clear test that can be used to assess which of the two methods provides a more accurate representation of each Rev 1.1-equivalent sector from 1995 to 2016. Visual inspection of trend lines (focusing in particular on the gross revenues variable for Romania over time) seems to suggest that the "count only" method leads to a smoother correspondence across Rev 1.1 and Rev 2. While there remain some jumps in the times series according to both correspondences (which could reflect the similarity in timing of the transition from Rev 1.1 to Rev 2 along with the global financial crisis), the "count only" method appears to minimize these jumps. For this reason, the "gross revenues and count" method was disregarded in favor of moving forward with the "count only" method.

Data Cleaning. After downloading data for as many countries, variables, years, and sectors as possible and constructing a crosswalk to link the data across Rev 1.1 and Rev 2 sector classifications, the next step was to combine the data, clean the data, and select years, countries, and sectors to include in the final analysis. Data that was downloaded according to the Rev 2 sector classification (for 2007 onward)

was split fractionally to corresponding Rev 1.1 sectors as per the crosswalk and combined with the data that was downloaded according to the Rev 1.1 sector classification (for prior to 2007). In cases where data was available under both classifications for overlapping years (as was often the case for 2005, 2006, and 2007), the data from the Rev 1.1 classification was prioritized, then filled in with data from the Rev 2 classification when the Rev 1.1 data was missing.

Selection of Years, Countries, and Sectors for Final Analysis. Years, countries, and sectors to include in the final analysis were selected to maximize availability of data and minimize the incidence of missing observations in the resulting panel. Less than 70% of the data across all countries and sectors were available for years prior to 1999 and after 2015, so only the years from 1999 to 2015 inclusive were retained for the analysis sample. The 70% cutoff applies to all countries and all sectors; this percentage would be higher once restricting to the final set of countries and sectors used in the analysis.

Similarly, Malta and Cyprus were dropped from the analysis sample due to low data availability. The remaining 12 countries retained featured relatively high data availability: at least 75% of the desired data were non-missing for the 12 countries retained across all years, and at least 85% of the desired data were non-missing when restricting attention to the years 1999-2015. Again, these percentages become larger once we focus on the final sectors chosen for their sufficient level of data availability.

Finally, because of the importance of estimating pre-trends for Romania, sectors were dropped if data were missing for any of the seven variables in either 1999 or 2000 for Romania. This requirement eliminated 14 sectors from the analysis sample (C10, C11, C12, C13, C14, G51, G52, I60, I61, I62, I64, J65, J66, and K67). An additional four sectors were dropped for reasons of low data availability: considering only 1999-2015 for all seven variables and the 12 retained countries, these were four sectors for which more than 10% of the desired data was missing (D15, D16, D19, and D23).

After restricting the analysis sample to the years 1999-2015, dropping Cyprus and Malta, and restricting to the 48 sectors where data was available for all seven variables for Romania in 1999 and 2000 and where there was a sufficiently high level of non-missing data, remaining missing values were imputed as follows. In cases where there was a missing observation for a particular variable within a time series, the missing observation's value was estimated by a simple average of the value for the most recently available and next available year, including for cases where the gaps were greater than one year. For cases where 1999 and potentially several consecutive years were missing, all of those observations were set to match the value for the earliest available year. Similarly, in the rare situation where there were potentially several consecutive years missing data ending in 2015, all of those variable's values were set to match that variable's value for the latest available year.

The ultimate level of imputation of missing values required was minimal, ranging from 0.8% to 1.9% of desired year by country by sector observations across the seven variables.

Online Appendix D.2 UN Comtrade Data

Downloading the Data. For for the exports of goods and services, the data was downloaded from the UN Comtrade official website. We downloaded the data for the same set of twelve countries that constitute the pool of donor countries for the SCM exercise using Eurostat data: Bulgaria, the Czech Republic, Estonia, Hungary, Ireland, Latvia, Lithuania, Poland, Portugal, Romania, the Slovak Republic, and

Slovenia. For goods exports, we downloaded the data disaggregated by two-digit SITC Rev 1 commodity codes. For service exports, we downloaded the data disaggregated by two-digit (X.X) EBOPS2002 codes (except Financial Services which is just classified as 6).[9] The variable of interest was Trade Value (US$) for Trade Flow=Exports.

Data Cleaning and Selection of Years and Codes for Final Analysis. The dataset for goods exports had the following features:

- There was a high rate of missing data for BG in 1995 and HU, PT in 2015
- We had to remove some codes to make sure that the codes entering the aggregates (high- versus low-intensity of IT inputs use) are consistent across countries. Removing the following sectors leads to a (close to) consistent panel for 1996-2014. 34. Missing: BG (1995, 1998-2000, 2003), LV (1998), RO (1996-1997), SK (2005); 35. Missing: BG (1995-2006), HU (1995-1996, 2001-2002), IE (2010-2011), PT (1997-1998, 2015), RO (1996), SK (1995-1996, 2005); 41. Missing: RO (2003); 52. Missing: HU (1998-1999, 2006, 2009, 2015), IE (2001, 2003, 2005), LV (2009), SK (2007); 93. Missing: CZ (1996-2001, 2003) , EE (1995-1998, 2002), HU (1995-1996, 2015), IE (1996), PL (1996-1997), RO (2006), SK (1995-2003); 95. Missing: BG (1995, 1997-2001, 2003, 2007-2013), HU (1995, 2003, 2009, 2015), IE (1997, 1998, 2000), RO (1998-2005), SI (1996, 1999). Note that in 2008, SK does not have sector 12.

The download of the service export data allowed us to notice that the data is available only from 2000 to 2016. Even during these years, most countries present a lot of missing values across codes. In particular, for Romania the treated NACE Rev 1 sector (72) starts reporting data only in 2005, not allowing us to study the service exports of this sector. The resulting (unbalanced) panel does not allow for an SCM, as we only observe one year before the event in 2001. Moreover, the large extent of missing values delivers noisy estimates of treatment effects.

Creation of Crosswalk between EBOPS2002 Codes and NACE Rev 1.1 Sector Codes. Our categories of high- and low-intensity (of IT usage) sectors were constructed based on Romania's input-output Table for 2000, which is at the two-digit NACE Rev 1 sector level. Hence, we had to construct a mapping between two-digit SITC Rev 1 commodity codes and two-digit (X.X) EBOPS2002 service codes and the two-digit NACE Rev 1 codes. After commodity or service codes were mapped to NACE Rev 1 codes, we aggregated export values at the NACE Rev 1 level.

References

Madalina Pruna and John Soleanicov. Rethinking Romanian Industrial Policy. Technical report, MPA/ID Thesis, John F. Kennedy School of Government, Harvard University, 2012.

Brian Quistorff and Sebastian Galiani. The synth_runner Package: Utilities to Automate Synthetic Control Estimation Using **synth**, 2017. URL https://github.com/bquistorff/synth_runner. Version 1.6.0.

[9] EBOPS2002 includes the standard items, memorandum items, and supplementary items. Only standard items were used. For codebook, visit link.