

# Returns to Education in the Russian Federation: Application of a machine learning instrumental variable technique for policy development in priority regions

Ekaterina Melianova\* | Suhas Parandekar\* | Harry Patrinos\*  
| Artëm Volgin\*

\* Education Global Practice, Europe and  
Central Asia

## Correspondence

Email: [sparandekar@worldbank.org](mailto:sparandekar@worldbank.org)

## Data and Code

Thanks are due to the Higher School of  
Economics, Moscow for making the Russian  
Longitudinal Monitoring Study (RLMS)  
Household data readily available for  
researchers around the world. The code  
used for this paper is made freely available  
for all researchers at [https://bitbucket.  
org/zagamog/edreru/src/master/](https://bitbucket.org/zagamog/edreru/src/master/)

This is a generic template designed for use by multiple journals, which includes several options for customization. Please consult the author guidelines for the journal to which you are submitting in order to confirm that your manuscript will comply with the journal's requirements. Please replace this text with your abstract.

## KEYWORDS

keyword 1, keyword 2, keyword 3, keyword 4, keyword 5,  
keyword 6, keyword 7

## 1 | MOTIVATION FOR THIS PAPER

### 1.1 | Stylized fact 1: Returns trend and looking at cohorts

### 1.2 | Stylized fact 2: Priority regions and policy options for education

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

---

**Acknowledgments:** Country Director: Renaud Seligman; Regional Director: Fadia Saadah; Practice Manager: Harry Patrinos; Program Leader: Dorota Nowak; Peer Reviewers: Cristian Aedo; Ruslan Yemtsov; Husein Abdul-Hamid; Team members: Polina Zavalina; Zhanna Terlyga. Thanks to seminar participants at the World Bank Moscow office on Jan. 29, 2020 for useful feedback. Any errors are a responsibility of the authors.

Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

First level:

$$\text{LogWage}_{ij} = b_{0j} + b_{1j} \cdot \text{Educ} + b_{2j} \cdot \text{Exp} + b_{3j} \cdot \text{Exp}^2 + b_{4j} \cdot \text{Gender} + \epsilon_{ij} \tag{1}$$

Second Level:

$$b_{0j} = \gamma_{00} + \gamma_{0n} \cdot Z + u_{00}; \qquad b_{1j} = \gamma_{10} + \gamma_{1n} \cdot Z + u_{10}; \qquad b_{ij} = \gamma_{i0} \quad \text{for } i \neq 0 \tag{2}$$

where an individual  $i$  is nested withing a region  $j$ ,  $\text{LogWage}$  is a logarithm of monthly wage,  $\text{Educ}$  stands for highest attained level of education,  $\text{Exp}$  and  $\text{Exp}^2$  reflect the years of working experience and its quadratic term respectively,  $\text{Gender}$  is a dummy variable for gender,  $Z$  is an  $n \times i$  matrix of regional characteristics,  $\epsilon$  and  $u_{00}$ ,  $u_{10}$  are the first- and second-level errors respectively.

We refer to `card_070_1999` and `belloni052.2011`.

## 2 | LITERATURE REVIEW: USE OF INSTRUMENTAL VARIABLES IN RETURNS TO EDUCATION STUDIES

General Literature

Russia Specific

## 3 | LASSO AND POST-LASSO INSTRUMENTAL VARIABLES: OVERVIEW OF ESTIMATION METHOD

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis.

Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

## 4 | EMPIRICAL STUDIES USING POST-LASSO IV METHODOLOGY

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

## 5 | FINDINGS 1: TRADITIONAL INSTRUMENTAL VARIABLE TECHNIQUE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula,

eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

## 6 | FINDINGS 2: POST LASSO IV RESULTS

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

## 7 | CONCLUSIONS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae

risus porta vehicula.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

## Appendix

**TABLE A1** Results of Estimating Human Capital Depreciation for the Female sample, RLMS

	1994	1998	2003	2006	2012	2018
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	9.725*** (0.381)	3.786*** (0.322)	5.464*** (0.301)	6.946*** (0.247)	8.133*** (0.186)	8.767*** (0.242)
Educ, years ( <i>S</i> )	0.122*** (0.025)	0.153*** (0.022)	0.158*** (0.020)	0.118*** (0.016)	0.087*** (0.012)	0.066*** (0.015)
Educ X Exper ( <i>T</i> <i>S</i> )	−0.002* (0.001)	−0.002*** (0.001)	−0.002** (0.001)	−0.0002 (0.001)	−0.0001 (0.0005)	0.0004 (0.001)
Exper ( <i>T</i> )	0.074*** (0.019)	0.080*** (0.016)	0.055*** (0.015)	0.013 (0.013)	0.020** (0.010)	0.020* (0.011)
Exper squared ( <i>T</i> <sup>2</sup> )	−0.001*** (0.0002)	−0.001*** (0.0002)	−0.001*** (0.0002)	−0.0003** (0.0001)	−0.0005*** (0.0001)	−0.001*** (0.0001)
Observations	1,645	1,667	2,093	2,630	4,057	3,312
R <sup>2</sup>	0.051	0.089	0.110	0.139	0.104	0.092
Adjusted R <sup>2</sup>	0.049	0.087	0.108	0.138	0.103	0.091
Residual Std. Error	0.853	0.728	0.731	0.664	0.641	0.597
F Statistic	22.179***	40.520***	64.342***	106.385***	117.366***	83.993***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**TABLE A2** Results of Estimating Human Capital Depreciation for the Male sample, RLMS

	1994	1998	2003	2006	2012	2018
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	10.357*** (0.433)	5.029*** (0.360)	7.334*** (0.282)	8.067*** (0.243)	8.771*** (0.157)	9.094*** (0.185)
Educ, years ( <i>S</i> )	0.136*** (0.028)	0.123*** (0.024)	0.080*** (0.019)	0.077*** (0.016)	0.077*** (0.010)	0.077*** (0.012)
Educ X Exper ( <i>TS</i> )	−0.002* (0.001)	−0.001 (0.001)	0.0004 (0.001)	−0.0003 (0.001)	−0.0004 (0.0005)	−0.001 (0.001)
Exper ( <i>T</i> )	0.054** (0.023)	0.032* (0.017)	0.002 (0.014)	0.007 (0.013)	0.035*** (0.009)	0.037*** (0.010)
Exper squared ( <i>T</i> <sup>2</sup> )	−0.001*** (0.0003)	−0.0004** (0.0002)	−0.0003* (0.0002)	−0.0003* (0.0001)	−0.001*** (0.0001)	−0.001*** (0.0001)
Observations	1,392	1,433	1,763	2,170	3,360	2,800
R <sup>2</sup>	0.057	0.070	0.078	0.074	0.153	0.110
Adjusted R <sup>2</sup>	0.054	0.067	0.076	0.072	0.152	0.108
Residual Std. Error	0.951	0.803	0.754	0.688	0.598	0.570
F Statistic	20.989***	26.879***	37.362***	43.281***	151.868***	86.125***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**TABLE A3** Results of Multilevel Modeling with Coverage by Vocational Education, Rosstat 2018

	Null model	Mincerian	Random Slope	Cross-Level Interaction
	(1)	(2)	(3)	(4)
Constant	10.178*** (0.034)	10.032*** (0.034)	10.056*** (0.036)	10.065*** (0.036)
Vocational		0.283*** (0.009)	0.279*** (0.021)	0.267*** (0.021)
Higher		0.638*** (0.009)	0.641*** (0.025)	0.622*** (0.025)
Coverage VE X Vocational				0.050** (0.025)
Coverage VE X Higher				0.083*** (0.030)
Experience		-0.026*** (0.002)	-0.027*** (0.002)	-0.027*** (0.002)
Experience squared		-0.065*** (0.002)	-0.065*** (0.002)	-0.065*** (0.002)
Females		-0.403*** (0.005)	-0.404*** (0.005)	-0.404*** (0.005)
Coverage VE			-0.101*** (0.039)	-0.142*** (0.043)
Variance of Intecept	0.09	0.08	0.09	0.09
Variance of Vocational			0.02	0.02
Variance of Higher			0.04	0.04
Residual Deviance	0.45	0.35	0.34	0.34
Observations	49,187	49,187	49,187	49,187
Log Likelihood	-59,755.060	-53,289.500	-53,094.620	-53,096.640
Akaike Inf. Crit.	119,516.100	106,595.000	106,217.200	106,225.300
Bayesian Inf. Crit.	119,542.500	106,665.400	106,340.500	106,366.100

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01