

Soon, Jan-Jan; Lim, Hock-Eam

**Working Paper**

## Driver of returns to schooling: Education-related policies or family background?

GLO Discussion Paper, No. 1471

**Provided in Cooperation with:**

Global Labor Organization (GLO)

*Suggested Citation:* Soon, Jan-Jan; Lim, Hock-Eam (2024) : Driver of returns to schooling: Education-related policies or family background?, GLO Discussion Paper, No. 1471, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/300729>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**Driver of returns to schooling:  
Education-related policies or family background?**

**Jan-Jan Soon\***  
**Hock-Eam Lim‡**

*Version: August 2024*

**Abstract**

‘Does schooling pay off?’ – a seemingly straightforward question, but it is in fact a puzzle among economists. Answers would differ based on how the returns to schooling are estimated. Among the top concerns is whether such estimations have any causal connotation between the amount of schooling and its returns or earnings. The endogeneity issue arises due to ability bias, where ability is typically related with years of schooling. The impact of schooling would be confounded by ability, hence the difficulty in isolating schooling’s causal impact on earnings. To address the concern, we conduct a meta-analysis of 74 empirical studies from which we retrieve returns to schooling coefficients estimated using both the causal instrumental variable and non-causal naïve estimation approaches. Key findings from our meta-analysis suggest an overall impact of 0.898, meaning an additional year of schooling is associated with a 8.98% increase in earnings, on average. We also find that over the years, returns to schooling exhibit an upward trend in general. Probing deeper, our analyses provide statistical evidence that education-related policy factors are driving the results more than family background factors.

**Keywords**

Returns to schooling, meta-analysis, family background, education-related policy

**JEL Codes**

C50, I26, I28

---

\* Corresponding author; Associate Professor, School of Economics, Finance & Banking, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia. Fellow, Global Labor Organization (<https://glabor.org/user/janjansoon/>); Associate Research Fellow, Economic & Financial Policy Institute (ECOFI). Email: [soon@uum.edu.my](mailto:soon@uum.edu.my). Acknowledgements: This work was supported by the Malaysian Ministry of Higher Education Fundamental Research Grant Scheme [FRGS/1/2016/SS08/UUM/02/9], and UUM University Grant [SO Code: 21425].

‡ Professor, School of Economics, Finance & Banking, Universiti Utara Malaysia.

## **Overview and selected literature**

Does schooling pay off? There is no short of empirical evidence that during the Covid-19 pandemic, the more educated have better job security. Worldwide school enrolment rates have stagnated well ahead of the pandemic. Statistics from World Bank show that globally, net enrolment rates of primary schools have stagnated at 89% between 2009 and 2018, while the rates of secondary schools have remained at 66% between 2014 and 2018. World Bank statistics have also revealed that the global average of government expenditure on education as a percentage of total government expenditure has declined since 2015 from 14.5%, to 12.7% in 2021. Currently, more than 243 million youths have gig employment, and most of them are high school graduates; these youths do not consider technical skills as important for gig jobs (World Bank 2023). The primary aim of this meta-analysis is to provide evidence to address the questions: (i) Does schooling pay off? (ii) What are the key factors contributing to heterogeneity in RtS estimates? (iii) What are the implications of estimating RtS using different types of IV (i.e. education-based policies vs socioeconomic family background)?

The returns to schooling (RtS henceforth) notion has its theoretical underpinnings from Becker's (1962) theory of human capital investment. Schooling or education is viewed as an investment, and it goes hand in hand with high earnings. Becker's theory finds its empirical representation in the form of Mincer's (1974) wage equation. A basic Mincerian wage equation takes form as  $\ln(Wage) = \beta_0 + \beta_1 Schooling + u$ . The basic Mincer wage equation has log earnings the dependent variable, years of schooling as its key regressor, and an error term component. The coefficient to the schooling,  $\beta_1$ , is known as the RtS.

Meta-analysis was first developed in educational research by a pioneering meta-analyst since mid-1970s to synthesise research findings from different studies working on similar topics (Glass 1976). Meta-analysis is a type of quantitative literature review, and a study of primary studies. Applications of meta-analysis in the field of economics were pioneered by Stanley and Jarrell, notably in labour economics (Stanley & Jarrell 1989). Since then, there have been several influential meta-analyses on RtS (Ashenfelter et al. 1999; Harmon et al. 2003; Card 2001). All the four meta-analyses examine differences between OLS and IV estimates of RtS, using between 7 and 27 primary studies respectively. They collectively find evidence of publication selection, heterogeneity in RtS estimates, and larger IV estimates compared to OLS estimates.

### **Methodology**

Observations in meta-analyses are the key statistics and studies' characteristics extracted from the set of primary studies included based on a pre-determined set of inclusion criteria. We extract RtS estimates from each of the included primary studies along with their associated standard errors (or t-statistics); these are the two crucial statistics. Table 1 shows the list of essential criteria for primary studies or publications to be included in our meta-analysis.

We begin our search for relevant primary studies by first combing through the list of primary studies included in the previous 3 RtS meta-analyses. Next, we search through the Scopus and Google Scholar databases using keywords such as 'returns to schooling', 'returns to education', 'instrument', 'socioeconomic', 'family background', 'parents' education', 'compulsory schooling law', and 'college proximity'. We also search the reference lists of our included primary studies, to round up any remaining relevant studies. Our search strategy gives us a final working meta-

dataset of 257 RtS estimates from 70 primary studies. Meta-analyses can determine the sources of heterogeneity in estimates across primary studies, i.e. factors which make them differ in terms of estimates' magnitude, sign, and statistical significance. It is received wisdom in conventional meta-analyses to model heterogeneity using meta-regressions, which are basically analogous to multiple linear regressions. Meta-regressions help identify why RtS estimates differ between the primary studies. Our meta-regression is as follows:  $RtS_i = \beta_0 + \beta_1 se_i + \sum_{k=1}^K \delta_k \mathbf{W}_{ik} + u_i$ .

[Table 1 about here]

### **Discussion of findings**

Figure 1 shows the distributions of the two types of RtS estimates, i.e. estimates obtained through OLS and IV estimations in the primary studies. The topmost panel shows a somewhat right-skewed distribution, indicating a larger proportion of RtS estimates with values less than the mean of 0.089. It appears that such a skewed distribution is driven by IV estimates, judging by its similar histogram shape. We also notice larger proportions of IV estimates with values higher than their OLS counterparts – this finding is supported by numerous empirical evidence in the RtS literature.

[Figure 1 about here]

The scatterplots in Figure 2 also show that generally, IV estimates have larger values than OLS estimates. Both types of estimates exhibit a somewhat slight upward trend by publication and dataset years. The RtS empirical literature seems to collectively solidify the positive wage-schooling relationship. Figure 2 reveals another interesting detail – after more than 3 decades of

research on RtS since 1991, one would expect the RtS estimates to converge to a narrower range of values. The scatterplots, however, do not display clear traces of convergence. If anything, the divergence seems more pronounced among IV estimates, with a much wider range of values compared to the OLS estimates. Both Figure 2a and 2b show similar patterns of divergence regardless of whether we use publication year or dataset year.

[Figure 2a & 2b about here]

Table 2 uses simple univariate descriptive statistics to summarize the RtS estimates and characteristics of primary studies. We divide Table 2 into three vertical panels, by types of RtS estimates. Panel A reports the statistics for all RtS estimates, i.e. both types of RtS estimates obtained using OLS and IV estimation methods in the primary studies. Here, the mean RtS estimate is 0.0898, i.e. based on all the RtS estimates extracted from our set of 70 primary studies – an additional year of schooling is associated with an 8.98% increase in wages, on average. The mode for dataset year is 1995, and 2002 for publication year. The years of dataset used in the primary studies range from 1966 to 2017. The primary studies included in our meta-analysis are published between 1991 and 2022.

[Table 2 about here]

Panel B reports on descriptive statistics according to the types of RtS estimates, i.e. OLS and IV estimates. The p-value column indicates if there are any statistically significant differences between the two means or proportions. For example, at the 5% significance level, IV estimates are

larger than OLS estimates. This finding is consistent with the mainstream prevailing evidence from the RtS literature. The typically higher IV estimates are due to the change in RtS of a ‘marginal’ person (or ‘compliers’, or those affected by the change in IV), such as those who would have more years of schooling due to compulsory schooling laws or due to having better-educated parents. Panel C categorises the IV estimates into those obtained using socioeconomic family background instruments (socio-IV) and those using education-based policy instruments (policy-IV). The means between socio-IV and policy-IV estimates do not have any statistically significant difference, as suggested by a p-value of 0.1425.

[Table 3 about here]

Table 3 reports results of our meta-regressions, i.e. an approach used to pinpoint factors that contribute to estimates’ heterogeneity. All the 5 specifications in suggest that – the types of sample (twin or female) and whether a primary study is published in a top journal – do not significantly explain heterogeneity in the RtS estimates. All remaining meta-regressors are statistically significant in Specification (1). However, when we run the meta-regressions by types of RtS estimates, a different story emerges. Comparing OLS and IV estimates in Specification (2) and (3), the countries of study do not contribute to heterogeneity in IV estimates. The use of surveys in collecting data, compared to using ready administrative data, has a statistically significant impact on heterogeneity in IV estimates but not in OLS estimates. Among IV estimates, comparing between RtS estimates obtained using family background IVs and education-based policy IVs in Specification (4) and (5), we find the use of education-based policy IVs such as compulsory

schooling laws to be a significant factor in influencing estimates' heterogeneity. The use of socioeconomic IVs such as parents' education, does not contribute to estimates' heterogeneity.

### **Conclusion and policy implications**

Using a meta-analysis of 70 primary studies and 257 RtS estimates, this paper addresses the fundamental question of whether schooling pays off. Our key findings suggest that yes, schooling does pay off. We find upward trends of RtS throughout the years of publication and dataset. The RtS estimates are predominantly positive; there are only 5 estimates showing a negative relationship of wages and schooling, all of which however are statistically insignificant. This provides strong empirical support to the human capital theory of investment and thereby the positive effects of schooling on wages. Comparing between socio-IV and policy-IV estimates, we find that the use of education-based policy IVs such as compulsory schooling laws to be a significant factor in influencing estimates' heterogeneity. The use of socioeconomic family background IVs such as parents' education, however, has no statistically significant impact on estimates' disparity.

What are the implications of our findings? They suggest consequential policy implications of estimating RtS using different types of IV. Our meta-analytic evidence indicates that policy interventions have stronger impact on schooling than socioeconomic family background. The upward trends of RtS over the years suggest that schooling still pays off. That is, for governments and individuals alike, educational investments are still worth it. Due to the issues of wage-schooling causality and endogeneity of schooling, our meta-analysis emphasises the importance of not making any policy decisions based on naïve OLS estimations of RtS. It is because such non-



causal and biased RtS estimates would be misleading and could have detrimental policy repercussions. Our findings on policy-IV estimates suggest that education-based policies would most likely affect schooling decisions of those who would not have continued with schooling without such policies. This finding implies that authorities in education-related matters should prescribe more targeted policies for subgroups who would yield the highest RtS. That is, policies should not be a blanket one-size-fits-all type.

### **Reference list**

- Ashenfelter, O., Harmon, C. & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, 6(4), 453-470.
- Becker, G.S. (1962). Investment in Human Capital: A Theoretical Analysis *Journal of Political Economy*. 70, 9-49.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127-1160.
- Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Harmon, C., Oosterbeek, H. & Walker, I. (2003). The returns to education: Microeconomics. *Journal of Economic Surveys*, 17(2), 115-155.
- Mincer, J. (1974). *Schooling, experience and earnings*. New York: Columbia University Press.
- Stanley, T.D., Jarrell, S.B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 19(3), 299-308.
- World Bank (2023). *Working without borders: The promise and peril of online gig work, Short Notes Series #1: Spotlight on youth*. Washington DC: The World Bank.

Table 1: Inclusion criteria

Characteristics of primary studies	Inclusion criteria
Dependent variable	Logarithm of wages (hourly, weekly, monthly, annual wages)
Key explanatory variable of interest	Years of schooling
Key estimate	Returns to schooling (i.e. estimated coefficient to years of schooling)
Standard error estimate	Standard error or t-statistic associated with the reported RtS estimate
Estimation model	IV estimation of Mincerian wage equations
Instrumental variable	Socioeconomic family background, education-based policies
Type of data structure	Cross-section, pooled, panel
Type of data level	Individual or micro-level
Geographic coverage of data	Any countries
Publication cutoff year	From Angrist and Krueger's 1991 work onwards
Publication type	Publications in journals and working paper series
Language	Studies published in English

Figure 1: Distribution of RtS estimates by type

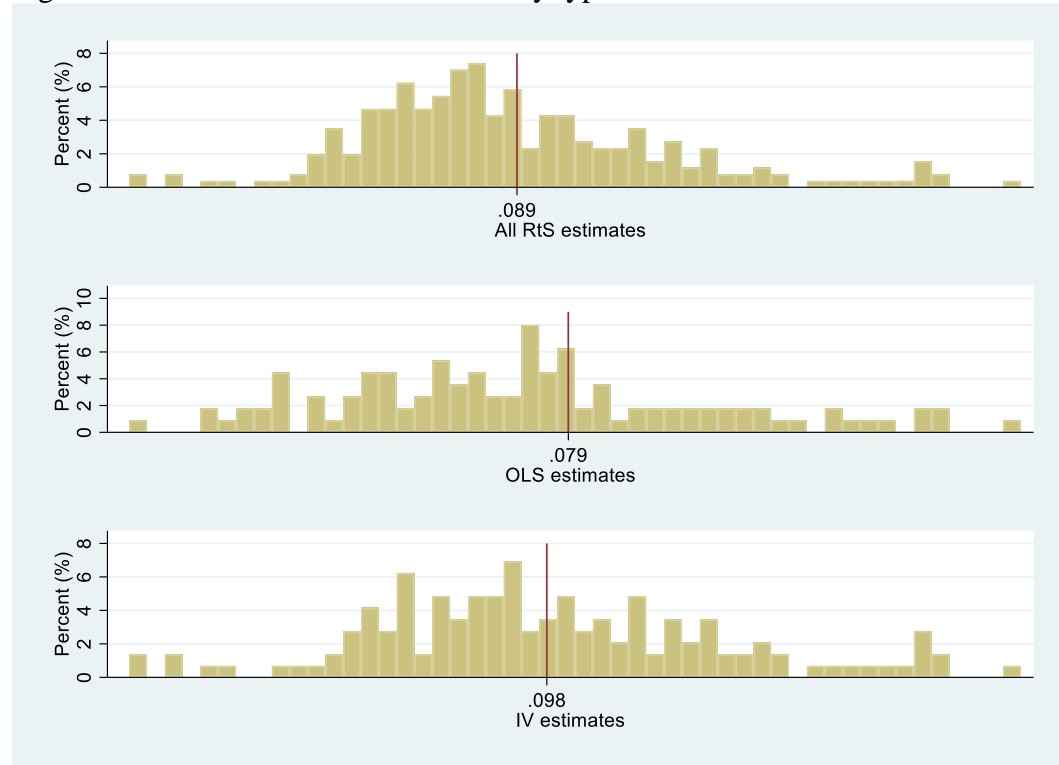


Figure 2a: Trend of RtS estimates by publication year

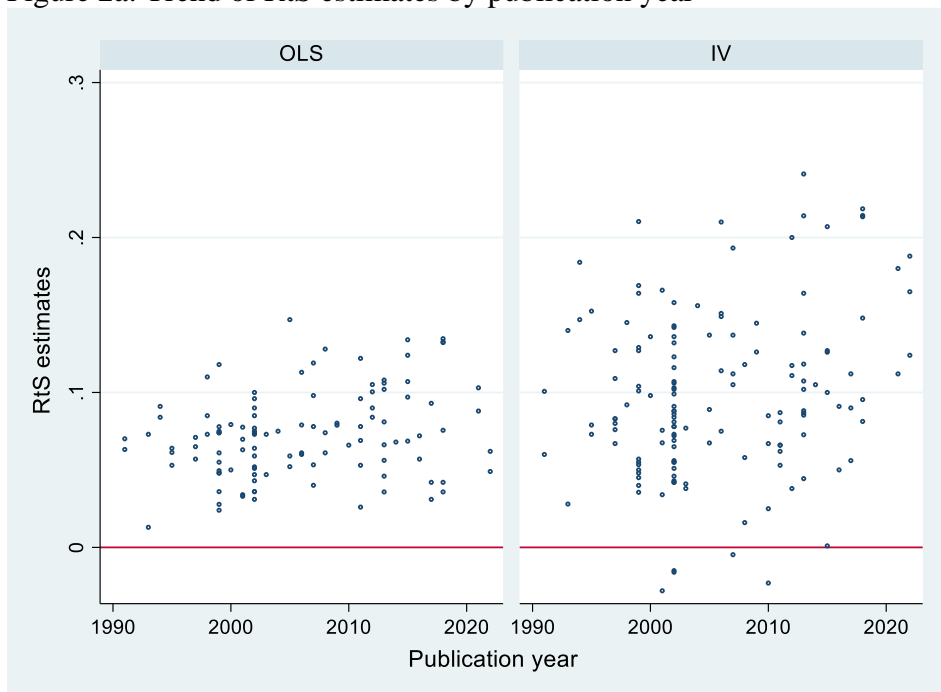


Figure 2b: Trend of RtS estimates by dataset year

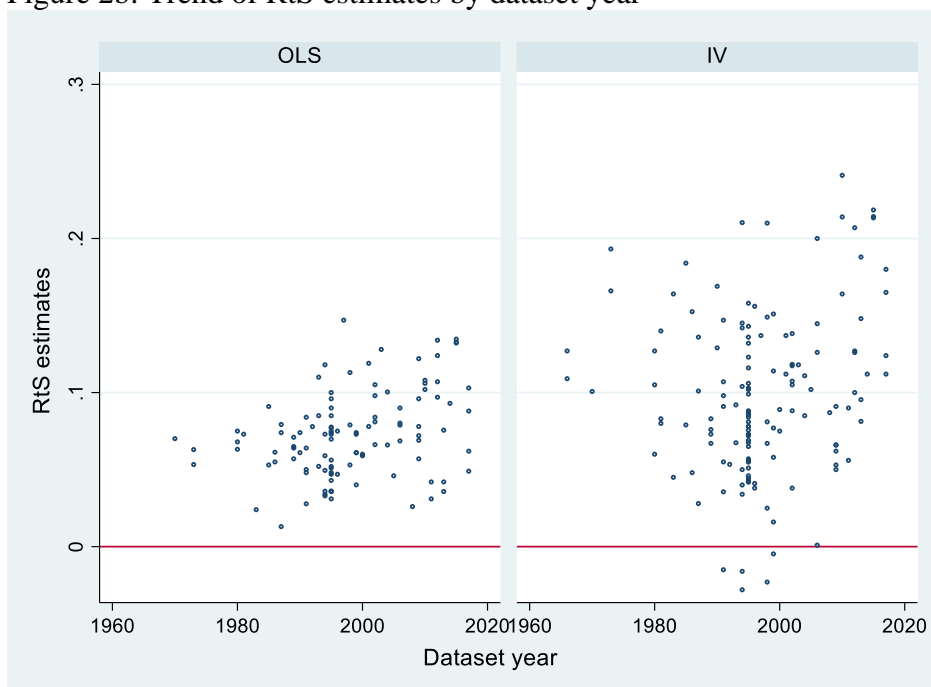


Table 2: Descriptive statistics

	Panel A:	Panel B: Types of RtS estimates			Panel C: Types of IV estimates		
	All RtS	OLS	IV	p-value	socio-IV	policy-IV	p-value
Mean RtS	0.0898	0.0791	0.0982	0.0188	0.1029	0.0892	0.1425
<i>Characteristics of primary studies:</i>							
D_inhourW	157	0.584	0.631	0.4366	0.747	0.408	0.0000
D_schys	257	-	-	-	-	-	-
D_ivcoeff	144	-	-	-	-	-	-
D_zbinary	84	0	0.583	0.0000	0.505	0.734	0.0079
D_signif5	234	0.991	0.847	0.0000	0.915	0.714	0.0013
D_coeffnegative	5	0	0.034	0.0457	0	0.102	0.0014
D_US	48	0.185	0.187	0.9731	0.231	0.102	0.0598
D_China	45	0.176	0.173	0.9439	0.221	0.081	0.0366
D_cross	128	0.522	0.479	0.4961	0.463	0.510	0.5954
D_survey	26	0.115	0.090	0.5153	0.136	0	0.0064
D_twinsample	15	0.061	0.055	0.8291	0.084	0	0.0368
D_femalesample	56	0.203	0.229	0.6230	0.284	0.122	0.0287
D_topjournal	39	0.159	0.145	0.7664	0.115	0.204	0.1571
D_zsocio	95	0	0.645	0.0000	-	-	-
D_zpolicy	49	0	0.340	0.0000	-	-	-
D_parentedu	46	0	0.319	0.0000	0.484	0	0.0000
D_CSL	27	0	0.194	0.0000	0	0.551	0.0000
Dataset year	1995						
Publication year	2002						

Table 3: Meta-regressions

	(1) All	(2) OLS	(3) IV	(4) IV-socio	(5) IV-policy
datayear	0.0011*** (0.0002)	0.0012*** (0.0003)	0.0007 (0.0008)	0.0012 (0.0011)	-0.0015 (0.0014)
pubyear	0.0007* (0.0004)	0.0008 (0.0005)	-0.0001 (0.0011)	-0.0009 (0.0020)	0.0017 (0.0014)
D_ivcoeff	0.0178*** (0.0049)				
D_zbinary	-0.0171*** (0.0057)		-0.0190** (0.0076)	-0.0201*** (0.0060)	-0.0208 (0.0297)
D_signif5	0.0661*** (0.0097)	0.0630*** (0.0175)	0.0685*** (0.0091)	0.0601*** (0.0140)	0.0843*** (0.0133)
D_coeffnegative	-0.0807*** (0.0109)		-0.0824*** (0.0145)		-0.0564** (0.0235)
D_US	0.0176*** (0.0036)	0.0197*** (0.0050)	0.0084 (0.0117)	0.0080 (0.0131)	-0.0458 (0.0381)
D_china	-0.0316*** (0.0099)	-0.0380** (0.0154)	-0.0109 (0.0112)	-0.0087 (0.0159)	0.0523** (0.0214)
D_cross	0.0126** (0.0058)	0.0136* (0.0073)	0.0122* (0.0066)	0.0101 (0.0112)	0.0206 (0.0172)
D_survey	0.0259*** (0.0084)	0.0148 (0.0258)	0.0385** (0.0149)	0.0346** (0.0159)	
D_twinsample	0.0033 (0.0071)	0.0081 (0.0089)	-0.0060 (0.0095)	-0.0042 (0.0096)	
D_femalesample	0.0074 (0.0075)	0.0070 (0.0110)	0.0063 (0.0054)	0.0083 (0.0062)	-0.0059 (0.0167)
D_topjournal	0.0049 (0.0042)	0.0078 (0.0068)	-0.0067 (0.0112)	-0.0094 (0.0127)	0.0036 (0.0270)
D_parentedu			0.0090 (0.0066)	0.0082 (0.0076)	
D_CSL			0.0260* (0.0138)		0.0339* (0.0189)
N	257	113	144	93	49

Notes: Standard errors in parentheses. Dependent variable, Y = RtS estimates. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$