Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

A Cross-Country Comparison Using PISA 2018 Data

Tony C. A. Tan

Thesis submitted for the degree of
Master in Assessment, Measurement and Evaluation
30 credits

Centre for Educational Measurement
Faculty of Educational Sciences

UNIVERSITY OF OSLO

Spring 2021

Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

A Cross-Country Comparison Using PISA 2018 Data

Tony C. A. Tan

Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

敬致父母

To my parents

Study hard what interests you the most in the most undisciplined, irreverent and original manner possible.

Richard P. Feynman

Contents

List of Tables

List of Figures

Acknowledgement

Thank-you goes to

Popular Abstract

This is a press release style abstract.

Abstract

Repeated economic crises in recent memory have exposed the harsh consequences of financial *illiteracy* shared by high proportions of the general population. Policy makers experienced little resistance when identifying youth as the most effective group for bringing about improvement in citizens' ability to engage with economic and financial matters, but opinions quickly diverge over the optimal approaches for achieving such targeted outcome. Existing literature frequently reports the importance of family environment in cultivating students' financial literacy through the process of "financial socialisation" – [definition goes here] (reference). Such practice, however, encounters interrogation by educators over equity concerns should families remain the main arena for financial literacy development. Schools play vital roles in alleviating inequality in accessing education and training in general but scarce research so far has been devoted into identifying the specific classroom factors that are most effective in advancing students' financial literacy outcomes. The current study therefore attempts to contribute to this enquiry by investigating the relationship between school climate variables and students' financial literacy achievement with an aim of stimulating policy debate about the levers and instruments available to education interventionists for the purpose of improving young people's financial literacy and preparedness as they step into an increasingly uncertain world. Using the 2018 PISA dataset, this paper employs a three-level hierarchical model to conduct cross-country comparisons to highlight school climate variables that are most strongly associated with high financial literacy outcomes.

Chapter 1 Conceptual Framework

1.1 In-depth definitions of “financial literacy”

1.1.1 Every term my readers need in order to understand my research question

1.1.2 Survey not only PISA but also alternative definitions, even critiques of such definitions

1.1.3 Any practices that are common in maths/literature but uncommon in financial literacy? Meaning? Implies?

1.2 Country-level Financial Knowledge Index

PISA 2018 financial literacy dataset (OECD, 2020) provides rich information about students and schools. For the purpose of cross-country comparison, however, the country-level financial literacy information must be addressed separately by the researchers. Earlier attempts such as Moreno-Herrero et al. (2018) approximated this information using a variable "quality of math and science education" to control for country-level differences since consensus is yet to emerge about the most appropriate measure for countries' financial knowledge. Inspired by the UN's approach to forming Human Development Indices, a recent publication by Oliver-Márquez et al. (2020) proposed a macroeconomic measure for countries' general financial knowledge levels by examining their economic capability, educational training, existing practices in the financial markets as well as incentives to interact with financial products. More specifically, the authors considered a country's economic capability, represented by its GDP per capita, to be a key dimension in bringing about its financial knowledge index (FKI). Secondly, literature converges on the importance of educational training for a country's financial knowledge capability (OECD, 2005). Thirdly, countries with regular engagement with sophisticated financial products and financial markets should possess higher FKI. Lastly, countries with higher aggregate consumption levels and with ageing populations are likely to possess higher FKI due to more frequent exposure and pressure in retirement provision, respetively. Macroeconomic data needed for these computations can be sourced from the World Bank (World Bank, 2020) and the United Nations' *Human Development Reports* (United Nations, 2020).

Combining individual and institutional data cources can be a productive approach in international large-scale assessment (ILSA) research. According to the framework for comparative education analyses (Bray & Thomas, 1995), this project extends education outcome measures to a country level, addresses the aspect of society and labour market, and relates countries' entire populations to ILSA research (Strietholt & Scherer, 2018). By combining education outcome data with countries' economic performance indicators, this project remains most comparable to Hanushek and Woessmann (2012)—while these authors looked into the relationship between countries' education achievement and their GDP growth, the current investigation highlights how countries' GDP, along with other macroeconomic practices, in turn systematically impacts on their youth's educational performance.

Table 1.1

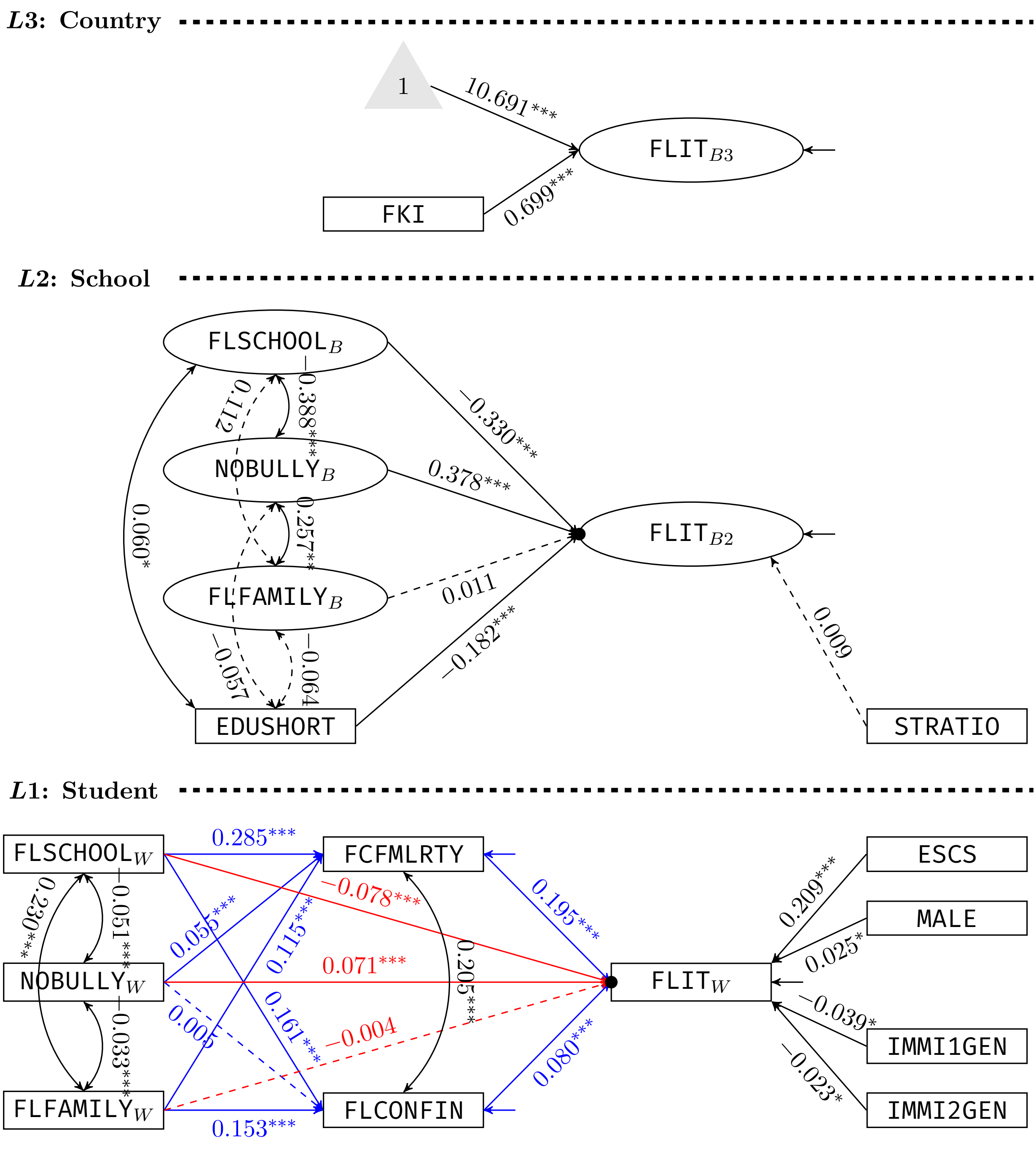
Percentages of Missing Values

CNT	MALE	IMMI1GEN	IMMI2GEN	ESCS	FCFMLRTY	FLCONFIN	PERFEED	TEACHINT	FLSCHOOL	DISCRIM†	BELONG	BULLY	FLFAMILY	CURSUPP†	PASCHPOL†	STRATIO	EDUSHT	STAFFSHT
BGR	0	6	6	3	12	27	10	10	21	28	19	31	22	100	100	8	3	3
BRA	0	5	5	2	12	34	9	8	21	36	23	40	24	17	19	12	6	7
CAN†	0	7	7	5	11	15	100	100	13	100	8	14	14	100	100	100	2	2
CHL	0	4	4	3	10	24	5	4	13	30	15	34	15	9	8	18	9	9
ESP	0	3	3	2	5	21	3	2	7	25	9	29	8	100	100	11	5	6
EST	0	3	3	3	4	8	3	3	6	9	5	11	6	100	100	0	0	0
FIN	0	2	2	2	4	10	3	3	6	100	6	11	7	100	100	2	7	7
GEO	0	5	5	2	9	26	9	9	17	100	15	22	21	4	5	1	2	2
IDN	0	3	3	1	3	6	3	2	5	3	2	5	5	100	100	23	14	14
ITA	0	4	4	3	7	17	4	4	10	23	10	27	12	16	17	9	3	3
LTU	0	3	3	3	4	12	3	3	5	17	8	20	7	100	100	0	0	0
LVA	0	2	2	2	5	9	3	3	6	14	6	15	7	100	100	6	3	4
NLD	0	3	3	2	3	5	3	2	4	100	4	8	4	100	100	11	5	5
PER	0	2	2	1	2	11	5	4	4	56	31	65	5	100	100	2	0	0
POL	0	1	1	1	3	7	2	1	5	9	3	11	5	100	100	0	0	0
PRT	0	6	6	5	8	11	6	6	10	15	8	17	10	10	10	11	1	1
RUS	0	3	3	2	8	13	5	4	11	13	8	15	11	100	100	3	3	3
SRB	0	3	3	1	10	25	8	7	18	25	15	27	19	100	100	8	1	1
SVK	0	2	2	1	4	12	4	3	7	14	6	17	8	100	100	6	6	7
USA	0	3	3	2	3	6	2	1	4	100	4	6	4	100	100	16	10	10

Note. Using shades of red in addition to numbers (measured in %), this table visualises the missing percentages by variable and by country. Variables `DISCRIM`, `CURSUPP` and `PASCHPOL` are no longer pursued in the model because too many countries chose not to respond to these questions. Canada (CAN) is not included due to 100 percent missings on multiple variables. † marks the country and variables that are excluded from subsequent analyses.

Figure 1.1

Three-level Structural Equation Model Predicting Youth's Financial Literacy Outcomes

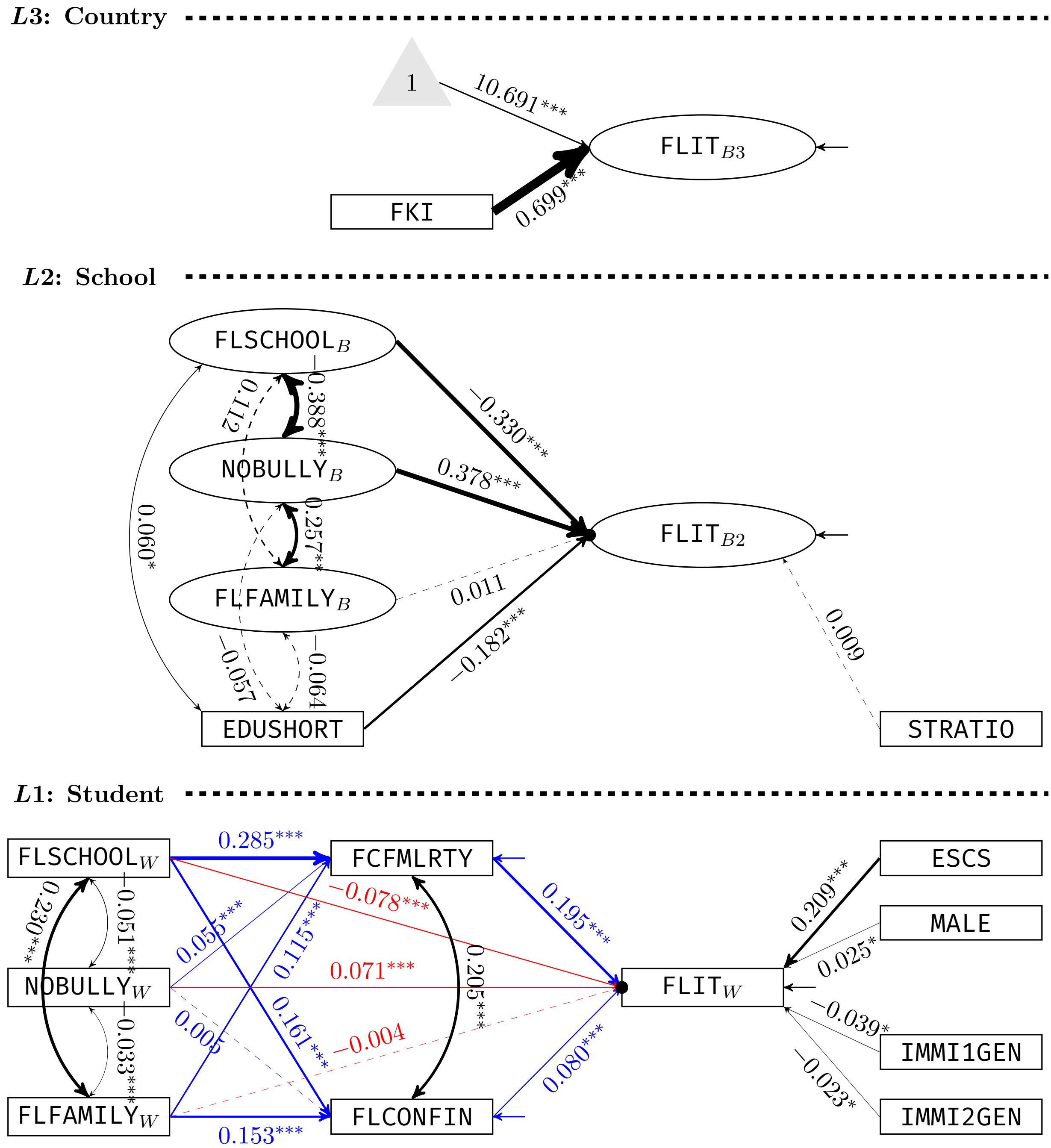


Note. This multilevel structural equation model predicts youth's financial literacy outcomes using PISA 2018 data, with mediating effects of familiarity with, and confidence in, financial matters at student-level (direct and indirect pathways). Statistics are standardised regression coefficients. Dashed lines represent nonsignificant relations at $\alpha = 0.05$ level. FKI = financial knowledge indices, FLIT = financial literacy. Subscript $_W$ = within, $_B$ = between.

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Figure 1.2

Three-level Structural Equation Model Predicting Youth's Financial Literacy Outcomes



Note. This multilevel structural equation model predicts youth's financial literacy outcomes using PISA 2018 data, with mediating effects of familiarity with, and confidence in, financial matters at student-level (direct and indirect pathways). Statistics are standardised regression coefficients. Dashed lines represent nonsignificant relations at $\alpha = 0.05$ level. FKI = financial knowledge indices, FLIT = financial literacy. Subscript $_W$ = within, $_B$ = between.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Chapter 2 Methods

2.1 Data / Sample / Participants

This study drew its primary data soruce from PISA 2018 database (OECD, 2020) containing 107,174 observations spanning 20 countries, in which students were asked about their demographic background, family lives and school experiences. For the financial literacy section, in particular, students responded to qustions about their confidence about financial matters, familiarity with concepts of finance, their parental involvement in matters of fianncial literacy. Ten plausible values were subsequently generated by PISA organisers as measures of students' financial literacy outcomes and were used as the dependent variable.

Student-level independent variables are

School-level independent variables are

Country-level independent variables are

Missing data are handled using Mplus's multiple imputation procedure with ten imputations generated and pooled subsequently following Rubin's Rule (Rubin, 1976).

A three-level multigroup structural equation model was employed to account for the hiearchical structure of the PISA design, with private versus public school as the grouping variable.

2.2 Measurement of financial literacy

2.2.1 Background questions

2.2.2 Students' motivation of spending money

2.2.3 Four-point Likert scale

2.2.4 Averages

2.3 Country-level Financial Knowledge Index

This project closely follows Oliver-Márquez et al.'s (2020) procedure in developing country-level financial kowledge indices using four sub-indices: economic capability ($\mathtt{EC}$), educational training ($\mathtt{ET}$), existing practices in financial market ($\mathtt{Use}$), and incentives ($\mathtt{Need}$) to engage with financial products. The first sub-index $\mathtt{EC}$ is calculated using the logarithm of a country's GDP per capita in current international dollars (purchasing power parity adjusted). For the $\mathtt{ET}$ sub-index, a country's highly skilled workforce is represented by its postgraduate to total tertiary graduation ratio, and the mean years of schooling is used to measure its general education level. For the $\mathtt{Use}$ sub-index, gross portfolio equity assets (GPEA) and insurance company assets (ICA) are considered sophisticated financial products a country engages in. Additionally, in order to capture the central role of technology in amplifying the proliferation and use of financial assets, the proportion of a country's Internet users (IUS) enters the definition via

$$\mathtt{Use} = (\mathrm{GPEA} + \mathrm{ICA})^{\mathrm{IUS}}.$$

The final sub-index $\mathtt{Need}$ is compiled as

$$\mathtt{Need} = (\mathrm{PFA} + \mathrm{AC})^{\mathrm{AGEING}},$$

where PFA is the pension fund assets to GDP ratio. Aggregate consumption is defined as:

$$\mathrm{AC} = \frac{2\% \times \text{household final consumption expenditure}}{\mathrm{GDP}},$$

with the "2% rule" being drawn from Caliendo and Findley's (2013) derivation, and the proportion of ageing population is computed as

$$\mathrm{AGEING} = \frac{\left[\frac{\text{population}(>65)}{\text{population}(20\sim64)}\right]_{2018} - \left[\frac{\text{population}(>65)}{\text{population}(20\sim64)}\right]_{2009}}{\left[\frac{\text{population}(>65)}{\text{population}(20\sim64)}\right]_{2009}}.$$

2.3.1 Data Collection and Missing Data Treatment

The data sources for FKI computation are documented in Table 2.1 and its associated notes. Sub-indices $\mathtt{ET}$ and $\mathtt{Use}$ both contain missing observations for the year 2018. Majority of such missing data appear to be the result of administrative delay, with historic observations available until 2017. It is therefore feasible to conduct time-series forecasts using prior year observations to best approximate 2018 values.

Table 2.1
Data Sources for FKI Computation

Database[a]	Country[b]	Series	Time
		Economic Capacity	
WB-dev	20	GDP per capita, PPP (current international $)	2018
		Educational Training	
WB-ed	20 \ Russia	Graduates from ISCED 7 programmes in tertiary education, both sexes (number)	2013–**2018**
		Graduates from ISCED 8 programmes in tertiary education, both sexes (number)	2013–**2018**
		Graduates from tertiary education, both sexes (number)	2013–**2018**
RS	Russia	PhD (Type 1)[c], PhD (Type 2)[d]	2018
RE	Russia	Master (Type 1)[e], Master (Type 2)[f], total tertiary *excluding* PhD[g]	2018
HDR	20	Dimension = Education; Education = Mean years of schooling (years)	2018
		Use	
WB-fin	20	Gross portfolio equity assets to GDP (%)	2011–**2018**
		Insurance company assets to GDP (%)	2011–**2018**
WB-dev	20	Individuals using the Internet (% of population)	2009–**2018**
		Need	
WB-fin	20 \ Georgia	Pension fund assets to GDP (%)	2008–**2018**
GP	Georgia	Minutes of the meeting of the investment board of the Pension Agency[h]	2019*
GS	Georgia	GDP at current prices, billion GEL[i]	2018
WB-dev	20	Household and NPISHs final consumption expenditure, PPP (current international $)	2018
		GDP, PPP (current international $)	2018
		Population ages 0–14, male	2009, 2018
		Population ages 0–14, female	2009, 2018
		Population ages 15–64, male	2009, 2018
		Population ages 15–64, female	2009, 2018
		Population ages 65 and above, male	2009, 2018
		Population ages 65 and above, female	2009, 2018
		Population ages 15–19, male (% of male population)	2009, 2018
		Population ages 15–19, female (% of female population)	2009, 2018

Note. Sub-indices are shaded in gray. Bold font signifies this year contains missing data.

[a] WB-dev = World Bank – World development indicators
WB-ed = World Bank – Education statistics – All indicators
WB-fin = World Bank – Global financial development
HDR = Human Development Reports – Data
RS = Russian Federal State Statistic Service
RE = Russian Ministry of Education and Science
GP = Pension Agency of Georgia
GS = National Statistics Office of Georgia

[b] "20" = the 20 participating countries in 2018 PISA financial literacy test: Brazil, Bulgaria, Canada, Chile, Estonia, Finland, Georgia, Indonesia, Italy, Latvia, Lithuania, the Netherlands, Peru, Poland, Portugal, Russian Federation, Serbia, Slovak Republic, Spain, and the USA. "\" = exluding or except

[c] https://rosstat.gov.ru/storage/mediabank/asp-2(1).xls, Sheet "по направлениям подготовки", Cell C7 = number of PhD graduates (Type 1)

[d] https://rosstat.gov.ru/storage/mediabank/asp-3.xls, Sheet "по научным специальностям", Cell B7 = number of PhD graduates (Type 2)

[e–g] https://minobrnauki.gov.ru/common/upload/download/VPO_1_2018.rar contains a spreadsheet СВОД_ВПО1_ВСЕГО.xls, Sheet "Р2_1_3(1)", Cell E198 = number of master graduates (Type 1)[e], Cell E410 = number of master graduates (Type 2)[f], Cell E592 = total tertiary graduates *excluding* PhD[g]

[h] Minutes of the meeting of the investment board of the Pension Agency, p. 4, no. 3

[i] Gross domestic product (GDP), row = GDP at current prices, billion GEL, column = 2018

[*] Georgia started a new pension system on 1 January 2019. Since 2018 was a transitional period with scarce data, 2019 is used as the best approximation for Georgia's pension system for 2018.

Sub-index ET

The 2018 archive for the number of master (ISCED 7), PhD (ISCED 8), and total tertiary graduates are incomplete for all participating countries except Georgia, Indonesia and Serbia. Figure 2.1 presents a time series plot of

$$\text{SKILLED} = \frac{\text{number of masters} + \text{number of PhDs}}{\text{total number of tertiary graduates}}$$

Figure 2.1

Proportion of Postgraduates to Total Tertiary Graduations

Note. "Postgraduate" is defined as master (ISCED 7) and PhD (ISCED 8) graduates. Countries not shown: GEO, IDN and SRB (2018 data available) and RUS (consult other sources)

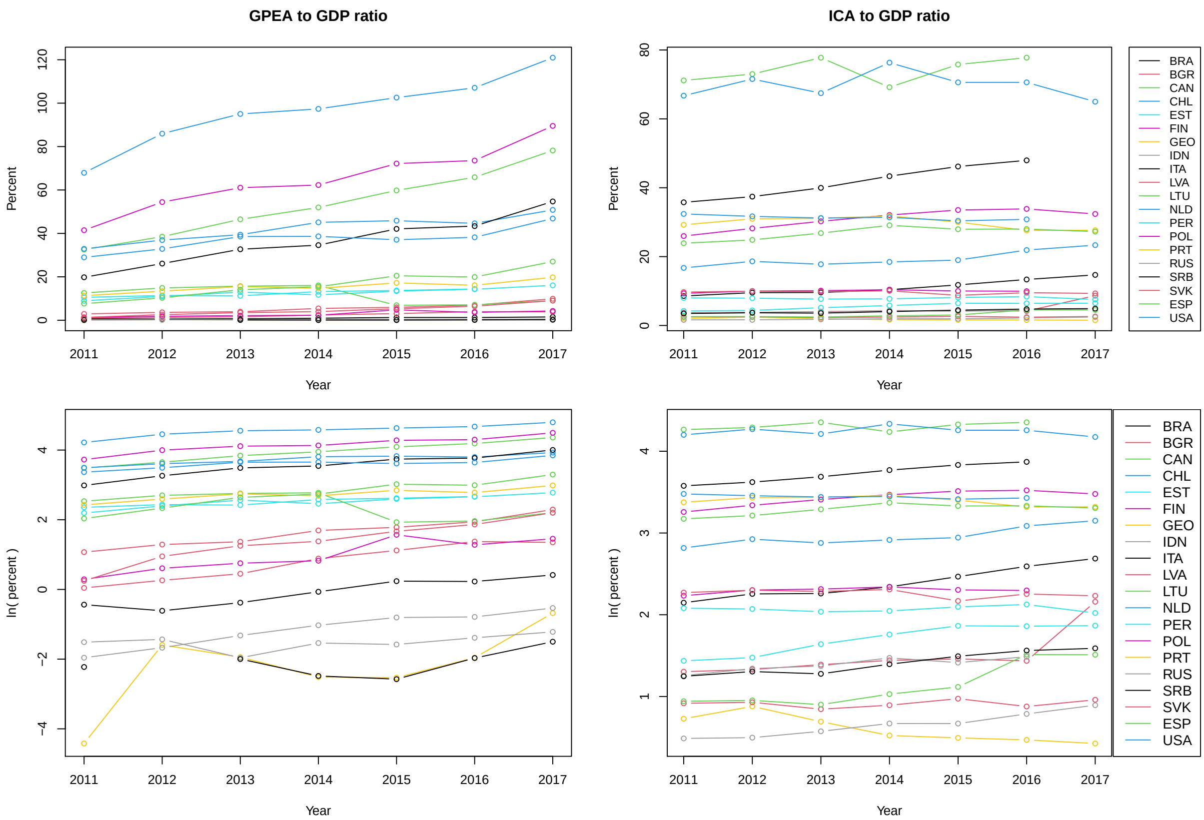
and suggests that this ratio is likely to be stable over time, especially between adjacent years. A "naive forecast", where the nearest available year's data are to be duplicated for 2018, is applied for SKILLED.

Sub-index **Use**

All series involved in calculating this sub-index, GPEA, ICA and IUS, contain missing data. When time series data contain only exponential growth but no underlying trend, a simple exponential smoothing would suffice (Garder, 1985); if trend is present, Holt-Winters method is superior (Chatfield, 1978). Figure 2.2 facilitates this decision making by plotting both the original and log-transformed versions of GPEA and ICA series. Since curves after log-transformations have slopes, it is prudent to apply the Holt-Winters forecasting method in order to account for possible trends contained in the original series.

Figure 2.2

Time Series Trend Test



Note. The time series plots after natural logarithm transformations (bottom panels) are not flat, suggesting the original series (top panels) contain trends. Holt-Winters method therefore is preferred over simple exponential smoothing for 2018 forecasts.

The IUS series contains missing data for Canada, Chile and the United States. Similar Holt-Winters procedure is applied to recover 2018 IUS data.

Table 2.2

Data Utilised for Computing FKI

	Economic Capacity	Educational Training		Use			Need		
	GDP per capita	Skilled	Schooling	GPEA	ICA	IUS	PFA	AC	AGEING
BRA	9.612	6.484	7.8	1.683	16.259	70.434	11.827	1.210	0.288
BGR	10.026	45.294	11.8	4.114	7.044	64.782	13.577	1.091	0.234
CAN	10.821	15.832	13.3	84.010	77.728	93.588	96.205	1.068	0.271
CHL	10.117	16.371	10.4	51.755	25.591	89.531	73.225	1.073	0.214
EST	10.501	36.765	13.0	16.399	7.681	89.357	18.012	0.876	0.163
FIN	10.807	35.024	12.4	93.626	31.481	88.890	52.024	0.974	0.370
GEO	9.588	24.039	12.8	0.784	1.469	62.718	0.834	1.227	0.042
IDN	9.362	7.771	8.0	0.636	4.612	39.905	1.826	1.059	0.145
ITA	10.665	44.771	10.2	57.434	51.260	74.387	10.589	1.075	0.155
LVA	10.330	29.554	12.8	8.598	2.538	83.577	14.732	1.027	0.142
LTU	10.487	28.749	13.0	9.008	5.500	79.723	7.457	1.107	0.149
NLD	10.961	32.590	12.2	124.171	64.956	94.712	207.938	0.805	0.326
PER	9.479	13.577	9.2	16.027	6.505	52.540	22.530	1.187	0.227
POL	10.368	36.725	12.3	4.853	9.535	77.542	9.838	1.085	0.355
PRT	10.444	34.454	9.2	19.353	25.579	74.661	8.761	1.133	0.237
RUS	10.267	30.349	12.0	0.302	2.614	80.865	4.415	0.941	0.155
SRB	9.774	26.946	11.2	0.306	5.111	73.361	0.845	1.171	0.280
SVK	10.391	54.417	12.6	10.644	8.873	80.660	12.497	0.962	0.300
ESP	10.609	33.929	9.8	27.681	28.230	86.107	10.235	1.044	0.186
USA	11.048	24.825	13.4	55.505	30.183	84.881	150.040	1.364	0.252

Note. Full variable names: Skilled = Postgraduate to total tertiary ratio; Schooling = Mean year of schooling; GPEA = Gross portfolio to GDP ratio; ICA = Insurance company assets to GDP ratio; IUS = Number of Internet users per 100 population; PFA = Pension fund assets to GDP ratio; AC = 2% of household final consumption expenditure to GDP ratio; AGEING = Aged-to-productive-population ratio (% change between 2009 and 2018)

Other Items with Data Concerns

Russia reported 67.96% and 61.01% of its total university degree receipients to be postgraduates for the year 2013 and 2015 respectively (2014 missing). This figure rapidly declines to 41.6% in 2016 and further down to 25.69% in 2017. Such volatility goes against the stable patterns shared by most countries in Figure 2.1, casting doubt on data reliability. Separate investigation is therefore conducted using Russian government archive (Notes c to g in Table 2.1).

Georgia underwent pension reform in 2018 with fund balance gradually transitioning to State Pension Agency for its official resumption of duty on 1 January 2019. Resultantly, 2018 pension balance for this country is unavailable but to be best appoximated using 2019 official data (Notes h, i and * of Table 2.1).

Table 2.2 documents the results of the abovementioned data recovery process.

2.3.2 Standardisation, Weights and FKI

Following Oliver-Márquez et al. (2020)'s procedure, all series in Table 2.2 undergo min-max normalisation such that the smallest entry receives a new score of 0.01 and the biggest number is re-coded to 0.99. This slight deviation from the original paper (where the min-max normalisation yields 0 to 1) is to avoid multiplying a series by zero or raising a base to the power of zero.

Variable weights are calculated following Oliver-Márquez et al. (2020)'s recipe to be the inverses of each series' standard deviations. Whereas a sub-index combines more than one series, each weight is further divided by the sum of the constituent weights so that total weights add to one.

FKI is finally computed by taking the geometric mean of all four sub-indices, subject to sub-index-weights similar to variable weights above, as presented in Table 2.3.

Table 2.3

FKI and Sub-indices

	FKI	EC	ET	Use	Need
NLD	0.940	0.939	0.640	1.805	1.000
USA	0.937	0.990	0.589	0.856	1.406
CAN	0.784	0.858	0.409	1.637	0.953
ITA	0.762	0.767	0.602	1.069	0.807
FIN	0.724	0.850	0.685	1.127	0.562
ESP	0.627	0.735	0.464	0.635	0.726
LTU	0.613	0.664	0.632	0.243	0.836
PRT	0.591	0.639	0.401	0.630	0.762
BGR	0.583	0.396	0.760	0.384	0.729
EST	0.577	0.672	0.746	0.266	0.575
SVK	0.559	0.608	0.924	0.301	0.441
POL	0.555	0.595	0.699	0.286	0.572
LVA	0.550	0.573	0.633	0.161	0.795
CHL	0.544	0.449	0.302	0.761	0.908
RUS	0.450	0.536	0.597	0.083	0.639
GEO	0.424	0.141	0.547	0.210	0.997
SRB	0.423	0.249	0.500	0.193	0.742
PER	0.309	0.078	0.194	0.691	0.877
BRA	0.141	0.155	0.010	0.432	0.833
IDN	0.122	0.010	0.040	0.973	0.787

Note. Table sorted in descending order by countries' FKI. FKI = financial knowledge index, EC = Economic Capability, ET = Educational Training.

2.4 What exactly I was using to address my research question

2.4.1 Sum score? Averages? One particular question?

2.4.2 Factor loading? Latent variables?

2.4.3 Motivation for choosing these measures

2.5 Software and version

2.6 My models

2.6.1 Motivation for choosing this particular model

2.6.2 Refer to my research question

2.7 Estimators I obtained

2.7.1 Motivation why these estimators rather than others

2.8 Weights? Plausible values?

2.9 Missing data and how I treated missing data

2.10 Model comparison

2.11 Guidelines and indices

References

Bray, M. & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard Educational Review*, *65*(3), 472–491. https://doi.org/10.17763/haer.65.3.g3228437224v4877

Caliendo, F. N. & Findley, T. S. (2013). Time inconsistency and retirement planning. *Economics Letters*, *121*(1), 30–34. https://doi.org/10.1016/j.econlet.2013.06.041

Chatfield, C. (1978). The Holt-Winters forecasting procedure. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *27*(3), 264–279. https://www.jstor.org/stable/pdf/2347162.pdf

Garder, E., Jr. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1–28. https://www.bauer.uh.edu/gardner/Exp-Sm-1985.pdf

Hanushek, E. A. & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321. https://doi.org/10.1007/s10887-012-9081-x

Moreno-Herrero, D., Salas-Velasco, M. & Sánchez-Campillo, J. (2018). Factors that influence the level of financial literacy among young people: The role of parental engagement and students' experiences with money matters. *Children and Youth Services Review*, *95*, 334–351. https://doi.org/10.1016/j.childyouth.2018.10.042

OECD. (2005). *Improving financial literacy: Analysis of issues and policies*. OECD Publishing. https://www.oecd-ilibrary.org/docserver/9789264012578-en.pdf

OECD. (2020). *Financial literacy data file* (Data set). OECD Publishing. https://webfs.oecd.org/pisa2018/SPSS_STU_FLT.zip

Oliver-Márquez, F. J., Guarnido-Rueda, A. & Amate-Fortes, I. (2020). Measuring financial knowledge: A macroeconomic perspective. *International Economics and Economic Policy*, 1–46. https://doi.org/10.1007/s10368-020-00482-2

Strietholt, R. & Scherer, R. (2018). The contribution of international large-scale assessments to educational research: Combining individual and institutional data sources. *Scandinavian Journal of Educational Research*, *62*(3), 368–385. https://doi.org/10.1080/00313831.2016.1258729

United Nations. (2020). *Human development reports* (Data set). The United Nations. http://hdr.undp.org/en/data

World Bank. (2020). *World Bank DataBank* (Data set). The World Bank. https://databank.worldbank.org/home.aspx

Appendices

Appendix A GDPR Documentation and Ethical Approval

This research project discharges its duty imposed by the EEA's general data protection regulation (GDPR) by following Norwegian Centre for Research Data (NSD)'s notification test on Friday, 11 September 2020. Both PISA 2018 Database and the World Bank Open Data contain only aggregated and de-personalised datasets with no possibility of back-tracing to any particular participant. Resultantly, no identifiable personal data were collected or used at any stage of this research. The NSD's assessment letter outlines the agency's decision of not subjecting this project to the GDPR notification. The NSD decision letter also satisfies University of Oslo's ethical approval requirement and concludes the approval process.

About us (/personvernombud/en/about_us.html)
Norwegian (/personvernombud/meld_prosjekt/meldeplikttest.html)

NSD (/) > Personverntjenester (/personvernombud/) > Data Protection Services (/personvernombud/en/) > Notify project (/personvernombud/en/notify/) > Notification Test

Denne siden på norsk (/personvernombud/meld_prosjekt/meldeplikttest.html)

Will you be processing personal data?

Are you unsure whether your project is subject to notification? Feel free to try our informal Notification test. Note that the test is intended as a guidance and is not a formal assessment.

Will you be collecting/processing directly identifiable personal data? ○ **Yes** ◉ **No**

A person will be directly identifiable through name, social security number, or other uniquely personal characteristics.

Read more about personal data (/personvernombud/en/help/vocabulary.html?id=8) and notification (/personvernombud/en/notify/index.html).

NB! Even though information is to be anonymized in the final thesis/report, check the box if identifying personal data is to be collected/processed in connection with the project.

Will directly identifiable personal information be linked to the data (e.g. through a reference number which refers to a separate list of names/scrambling key)? ○ **Yes** ◉ **No**

Note that the project will be subject to notification even if you cannot access the scrambling key (/personvernombud/en/help/vocabulary.html?id=11), as the procedure often is when using a data processor (/personvernombud/en/help/vocabulary.html?id=3), or in register-based studies (/personvernombud/en/help/research_methods/register_studies.html).

Will you be collecting/processing background information that may identify individuals (indirectly identifiable personal data)? ○ **Yes** ◉ **No**

A person will be indirectly identifiable if it is possible to identify him/her through a combination of background information (such as place of residence or workplace/school, combined with information such as age, gender, occupation, etc.).

Will there be registered personal data (directly/indirectly/via IP or email address, etc.) using online surveys? ○ **Yes** ◉ **No**

Please note that the project will be subject to notification even if you as a student/researcher cannot access the link to the IP or email address, as the procedure often is when using a data processor.

Read more about online surveys (/personvernombud/en/help/research_methods/online_surveys.html).

Will there be registered personal data using digital photo or video files? ○ Yes ◉ No

Photo/video recordings of faces will be regarded as identifiable personal data. In order for a voice to be considered as identifiable, it must be registered in combination with other background information, in such a way that people can be recognized.

Show results

Notify project

Do I have to notify my project? (/personvernombud/en/notify/index.html)

Notification Form (/personvernombud/en/notify/meldeskjema_link)

Notifying changes (/personvernombud/en/notify/notifying_changes.html)

Get help notifying your project

Processing the notification (/personvernombud/en/help/index.html)

Frequently asked questions (/personvernombud/en/help/faq.html)

Vocabulary (/personvernombud/en/help/vocabulary.html)

Research topics (/personvernombud/en/help/research_topics/)

Research methods (/personvernombud/en/help/research_methods/)

Information and consent (/personvernombud/en/help/information_consent/)

Other approvals (/personvernombud/en/help/other_approvals/)

Result of Notification Test: Not Subject to Notification

You have indicated that neither directly or indirectly identifiable personal data will be registered in the project.

If no personal data is to be registered, the project will not be subject to notification, and you will not have to submit a notification form.

Please note that this is a guidance based on information that you have given in the notification test and not a formal confirmation.

For your information: *In order for a project not to be subject to notification, we presuppose that all information processed using electronic equipment in the project remains anonymous.*

Anonymous information is defined as information that cannot identify individuals in the data set in any of the following ways:

- *directly, through uniquely identifiable characteristic (such as name, social security number, email address, etc.)*
- *indirectly, through a combination of background variables (such as residence/institution, gender, age, etc.)*
- *through a list of names referring to an encryption formula or code, or*
- *through recognizable faces on photographs or video recordings.*

Furthermore, we presuppose that names/consent forms are not linked to sensitive personal data.

Kind regards,
NSD Data Protection

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Appendix B Analysis Code

B.1 Chapter 1

There is no analysis code in ??.

B.2 Chapter 2

B.2.1 Data Import

```
# Housekeeping
library(Orcs)
setwdOS(lin = "~/", win = Sys.getenv("USERPROFILE"))

# Import SPSS file into R
library(intsvy)
finlit <- pisa.select.merge(
    student.file = "CY07_MSU_FLT_QQQ.SAV", # file ext in capital
    school.file = "CY07_MSU_SCH_QQQ.sav", # file ext in lower case
    student = c(
        # Control variables
        "ST004D01T", # Student (Standardized) Gender
        "IMMIG", # Index Immigration status
        "ESCS", # Index of economic, social and cultural status
        # Mediators
        "FCFMLRTY", # Familiarity with concepts of finance (Sum)
        "FLCONFIN", # Confidence about financial matters (WLE)
        # Academic
        "PERFEED", # Perceived feedback (WLE)
        "TEACHINT", # Perceived teacher's interest (WLE)
        "FLSCHOOL", # Financial education in school lessons (WLE)
        # Safety
        "DISCRIM", # Discriminating school climate (WLE)
        "BELONG", # Sense of belonging to school (WLE)
        "BEINGBULLIED", # Student's experience of being bullied (WLE)
        # Community
        "FLFAMILY", # Parental involvement in matters of Financial Literacy (WLE)
        "CURSUPP", # Current parental support for learning at home (WLE)
        "PASCHPOL" # School policies for parental involvement (WLE)
    ),
    school = c(
        "STRATIO", # Student-teacher ratio
        "EDUSHORT", # Shortage of educational material (WLE)
        "STAFFSHORT" # Shortage of educational staff (WLE)
    ),
    countries = c(
        "BGR", "BRA", "CAN", "CHL", "ESP",
        "EST", "FIN", "GEO", "IDN", "ITA",
        "LTU", "LVA", "NLD", "PER", "POL",
        "PRT", "QMR", "QRT", "RUS", "SRB",
        "SVK", "USA"
    )
)
```

```r

# Inspect table header
names(finlit)

# Remove columns that I do not need
finlit <- finlit[, -c(5, 7:86)] # 5 = BOOKID; 7:86 = resampling weights

# Some var need recording
library(car)

# Re-code Russian territories to RUS
finlit$CNT <- recode(finlit$CNT, "
    'QMR' = 'RUS';
    'QRT' = 'RUS'
")

# Recode ST004D01T from Sex to Male
MALE <- finlit$ST004D01T - 1

# Recode IMMIG to 1st and 2nd generation
IMMI1GEN <- recode(finlit$IMMIG, "
    1 = 0;
    2 = 0;
    3 = 1
")

IMMI2GEN <- recode(finlit$IMMIG, "
    1 = 0;
    2 = 1;
    3 = 0
")

# Revert coding direction: bigger number => safer school
NOBULLY <- finlit$BEINGBULLIED * (-1)

# Stitch spreadsheets together
finlit <- cbind(
    finlit[, c(1:35)],
    MALE, IMMI1GEN, IMMI2GEN,
    finlit[, c(38:45)],
    NOBULLY,
    finlit[, c(47:53)]
)

# Use data.table for better RAM management
library(data.table); setDTthreads(0) # 0 means all the available cores
# Export data into a CSV file for faster import next time
fwrite(finlit,
    file = "finlit.csv",
    na = "NA", row.names = F, col.names = T
)
```

B.2.2 Missing Pattern Inspection

```r
library(Orcs)
setwdOS(lin = "~/", win = Sys.getenv("USERPROFILE"))

library(data.table); setDTthreads(0)
finlit <- fread("finlit.csv", nThread = getDTthreads())

library(dplyr)
```

```
# Record how many missings each country has for each var
missings <- finlit %>%
    select(everything()) %>%
    group_by(CNT) %>%
    summarise_all(funs(sum(is.na(.))))
# Give me the headcount for each country
headcount <- finlit %>%
    group_by(CNT) %>%
    summarize(n())
# Stitch these two tables together
missing_table <- tibble(headcount, missings[, -1])
# Save this file
fwrite(missing_table, "missing_table.csv", row.names = F, col.names = T)

# Inspect the missing table using Excel
# Throw away the following countries
#    CAN: 100% missing on too many var
#    BRA, FIN, LVA, NLD, RUS, SRB: private/public info missing
# Throw away these var
#    DISCRIM, CURSUPP, PASCHPOL: Too many countries have 100% missing
```

B.2.3 Financial Knowledge Index

```
# Section 0: Housekeeping
library(Orcs) # Set working directory depending on operating system
setwdOS(
    lin = "~/uio/", win = "M:/",
    ext = "pc/Dokumenter/MSc/Thesis/Data/L3/"
)

# Set up a "bookshelf" to hold variables nessary to compute FKI
fki_raw <- matrix(NA,
    nrow = 20, ncol = 10, dimnames = list(
        c( # row names
            "BRA", "BGR", "CAN", "CHL", "EST",
            "FIN", "GEO", "IDN", "ITA", "LVA",
            "LTU", "NLD", "PER", "POL", "PRT",
            "RUS", "SRB", "SVK", "ESP", "USA"
        ),
        c( # column names
            "gdp_per_capita", # economic capability (sub_ind_ec)
            "highly_skilled", "mean_year_of_schooling", # (sub_ind_et)
            "gpea", "ica", "ius", # use (sub_ind_use)
            "pfa", "ac", "gdp", "ageing" # need (sub_ind_need)
        )
    ) # End list()
) # End matrix()


# Section 1: Economic Capacity (EC)

gdp_per_capita <- read.csv("gdp_per_capita.csv", header = T, sep = "\t")

fki_raw[, 1] <- log(gdp_per_capita[, 2])

rm(gdp_per_capita)


# Section 2: Educational Training (ET)
    # Subsection 2.1: Highly skilled
        # Masters
```

```
isced_7 ← read.csv("isced_7.csv", header = T, sep = "\t")
        # PhDs
isced_8 ← read.csv("isced_8.csv", header = T, sep = "\t")
        # Total tertiary
total_tertiary ← read.csv("total_tertiary.csv", header = T, sep = "\t")

# Compute highly skilled (master + PhD) to total tertiary ratio
highly_skilled ← ts(
    (isced_7 + isced_8) / total_tertiary,
    start = 2013, end = 2018, frequency = 1
)

# # Visualise highly_skilled. Turn off GEO (#7), IDN (#8), SRB (#17) and RUS (#16)
# pdf("../../Figures/skilled.pdf")
# ts.plot(100 * highly_skilled[, -c(7, 8, 16, 17)],
#     type = "b", col = 1:15,
#     xlab = "Year", ylab = "Percent"
# )
# legend("topright", colnames(highly_skilled[, -c(6, 7, 15, 16)]),
#     col = 1:15, lty = 1, cex = 0.65
# )
# dev.off()

# Decision: naive forcasts, i.e., copy-paste nearest available year
library(forecast)
# Create a placeholder matrix
placeholder ← matrix(NA, nrow = 20, ncol = 1)

# Run a loop to foreecast all 20 countries, using naive method
for (.i in 1:20) {
    m_naive_i ← naive(highly_skilled[, .i], h = 1)
    placeholder[.i] ← data.frame(unlist(m_naive_i[5])[6])[1, 1]
}
# [5] = fitted values; [6] = 2018; [1,1] = only the numeric value

# GEO and IDN have 2018 data, plug actual numbers back
placeholder[c(7, 8)] ← highly_skilled[6, c(7, 8)]

# RUS needs separate calculation
# ISCED 7 = 101766 (Type 1) + 170437 (Type 2) = 272203 (total masters)
# ISCED 8 = 15465 (Type 1) + 330 (Type 2) = 15795 (total PhDs)
# Total tertiary WITHOUT PhD = 933153
# ⇒ Total tertiary = 933153 + 15795 = 948948
# highly_skilled (RUS) = (272203 + 15795) / 948948 = 0.30349187
placeholder[16] ← 0.30349187

# Save results to "bookshelf"
fki_raw[, 2] ← placeholder * 100

rm(
    isced_7, isced_8, total_tertiary,
    highly_skilled, placeholder, m_naive_i
)

# Sub-section 2.2: Mean year of schooling
mean_year_of_schooling ← read.csv("mean_year_of_schooling.csv",
    header = F, sep = "\t"
)
fki_raw[, 3] ← mean_year_of_schooling[, 2]

rm(mean_year_of_schooling)
```

```r


# Section 3: Use

gpea ← read.csv("gpea.csv", header = T, sep = "\t")
gpea ← ts(gpea, start = 2011, end = 2017, frequency = 1)

# # Visualise data in both original and ln forms. Contain trend?
# pdf("../../Figures/use.pdf", width = 12.94, height = 9.15)

# # Re-set canvas layout to 2x2
# par(mfcol = c(2, 2))

# # Add extra space to the right of plot area
# par(mar = c(5.1, 4.1, 4.1, 2.1), xpd = TRUE)

# # Plot GPEA in original form
# ts.plot(gpea,
#     type = "b", col = 1:20,
#     xlab = "Year", ylab = "Percent", main = "GPEA to GDP ratio"
# )

# # Remove extra gap between the two graphs
# par(mar = c(5.1, 4.1, 0, 2.1), xpd = TRUE)

# # Repeat GPEA, but for the ln() version
# ts.plot(log(gpea),
#     type = "b", col = 1:20,
#     xlab = "Year", ylab = "ln( percent )"
# )

# # Plot ICA in original form
# par(mar = c(5.1, 4.1, 4.1, 6.1), xpd = TRUE)
# ts.plot(ica,
#     type = "b", col = 1:20,
#     xlab = "Year", ylab = "Percent", main = "ICA to GDP ratio"
# )
# # Add the legend
# legend("topright",
#     inset = c(-0.2, 0), colnames(ica),
#     col = 1:20, lty = 1, cex = 0.875
# )

# # Remove extra gap between the two graphs
# par(mar = c(5.1, 4.1, 0, 6.1), xpd = TRUE)

# # Repeat, but for the ln()
# ts.plot(log(ica),
#     type = "b", col = 1:20,
#     xlab = "Year", ylab = "ln( percent )"
# )
# # Add the legend
# legend("topright",
#     inset = c(-0.2, 0), colnames(ica),
#     col = 1:20, lty = 1, cex = 1.07
# )
# dev.off()

# Decision: since the ln() version is not flat, original time series
# contain trend. Use Holt method rather than simple exponential smoothing.
```

```r

# Run a time series forecast using Holt method

# Create a placeholder matrix
placeholder ← matrix(NA, nrow = 20, ncol = 1)

# Run a loop to forecast all 13 countries, using Holt method
for (.i in 1:20) {
    m_holt_i ← holt(gpea[, .i], h = 1)
    placeholder[.i] ← m_holt_i[2]
} # Ignore warnings

# Only keep the 2018 forecasts
placeholder ← unlist(placeholder)

# Run PER (#13) separately because it misses both 2017 and 2018 data
m_holt_PER ← holt(gpea[, 13], h = 2); summary(m_holt_PER)
placeholder[13] ← 16.02698

# Push placeholder to fki_raw
fki_raw[, 4] ← placeholder

rm(gpea, placeholder, m_holt_i, m_holt_PER)

# Sub-section 3.2: Insurance company assets (ica)

ica ← read.csv("ica.csv", header = T, sep = "\t")
ica ← ts(ica, start = 2011, end = 2017, frequency = 1)

placeholder ← matrix(NA, nrow = 20, ncol = 1)

for (.i in 1:20) {
    m_holt_i ← holt(ica[, .i], h = 1)
    placeholder[.i] ← m_holt_i[2]
} # Ignore warnings

placeholder ← unlist(placeholder)

m_holt_CAN ← holt(ica[, 3], h = 2); summary(m_holt_CAN)
m_holt_IND ← holt(ica[, 8], h = 2); summary(m_holt_IND)
m_holt_ITA ← holt(ica[, 9], h = 2); summary(m_holt_ITA)
m_holt_POL ← holt(ica[, 14], h = 2); summary(m_holt_POL)
m_holt_USA ← holt(ica[, 20], h = 2); summary(m_holt_USA)

placeholder[c(3, 8, 9, 14, 20)] ← c(
    77.72768, 4.611597, 51.2596, 9.534750, 30.18295
)

fki_raw[, 5] ← placeholder

rm(ica, placeholder, list = ls(pattern = "^m.holt"))

# Sub-section 3.3: Individuals using the Internet (ius)

ius ← read.csv("ius.csv", header = T, sep = "\t")
ius ← ts(ius, start = 2009, end = 2018, frequency = 1)

m_holt_CAN ← holt(ius[1:9, 3], h = 1); summary(m_holt_CAN)
m_holt_CHL ← holt(ius[1:9, 4], h = 1); summary(m_holt_CHL)
m_holt_USA ← holt(ius[1:9, 20], h = 1); summary(m_holt_USA)
```

```r

ius_2018 <- ius[10, ] # Only want 2018 data
ius_2018[3] <- 93.58751 # CAN
ius_2018[4] <- 89.5309 # CHL
ius_2018[20] <- 84.88108 # USA

fki_raw[, 6] <- ius_2018

rm(list = ls(pattern = "^ius"))
rm(list = ls(pattern = "^m_holt_"))


# Section 4: Need

# Subsection 4.1: Pension fund assets (pfa)
pfa <- read.csv("pfa.csv", header = T, sep = "\t")
# Delete GEO (#7) due to all missing. Will come back to it later.
pfa <- ts(pfa[, -7], start = 2008, end = 2017, frequency = 1)

placeholder <- matrix(NA, nrow = 19, ncol = 1)

for (.i in 1:19) {
    m_holt_i <- holt(pfa[, .i], h = 1)
    placeholder[.i] <- m_holt_i[2]
}

placeholder <- unlist(placeholder)

# Calculate GEO
# From Georgia Pension Agency:
#   2019 mesub_ind_eting minute: 372,113,933 GEL
# From GeoStat website:
#   2018 gdp = 44.6 billion GEL

fki_raw[, 7] <- c(
    placeholder[1:6],
    372113934 / 44600000000 * 100, # Insert GEO figure
    placeholder[7:19]
)

rm(pfa, placeholder, m_holt_i)

# Subsection 4.2: Aggregate consumption (ac)

ac <- read.csv("ac.csv", header = F, row.names = 1, sep = "\t")
gdp <- read.csv("gdp.csv", header = F, row.names = 1, sep = "\t")

fki_raw[, 8] <- unlist(ac * 0.02 / gdp * 100)

fki_raw[, 9] <- unlist(gdp)

rm(ac, gdp)

# Subsection 4.3: Ageing

ageing <- read.csv("ageing.csv", header = T, sep = "\t")
attach(ageing)
names(ageing)

# Calculate total population
poptotal_f <- pop0to14_f + pop15to64_f + pop65plus_f
```

```r
poptotal_m ← pop0to14_m + pop15to64_m + pop65plus_m
# Calculate population between 15 and 19
# Need to divide by 100 to get decimals
pop15to19_f ← poptotal_f * per15to19_f / 100
pop15to19_m ← poptotal_m * per15to19_m / 100
# Calculate population between 0 and 19
pop0to19_f ← pop0to14_f + pop15to19_f
pop0to19_m ← pop0to14_m + pop15to19_m
# Calculate population between 20 and 64
pop20to64_f ← poptotal_f - pop0to19_f - pop65plus_f
pop20to64_m ← poptotal_m - pop0to19_m - pop65plus_m
# Calculate 64+ / 20-to-64 ratio    'GEO' = 0.419;
ageing_ratio ← I(
    (pop65plus_f + pop65plus_m) / (pop20to64_f + pop20to64_m)
)
# Split data into 2018 [ , 1] and 2009 [ , 2] portions
ageing ← cbind(ageing_ratio[1:20], ageing_ratio[21:40])
fki_raw[, 10] ← (ageing[, 1] - ageing[, 2]) / ageing[, 2]

rm(ageing, ageing_ratio, list = ls(pattern = "^pop"))


# Section 5: FKI

fki_raw ← fki_raw[, -9] # Throw away gdp (already in ac)
round(fki_raw, digits = 3) # Inspect data

# Save data to an external file
library(data.table); setDTthreads(0)
fwrite(round(fki_raw, digits = 3), file = "fki_raw.csv", row.names = T)

# Subection 5.0: Standardise each variable to [0.01,0.99] range
fki_stand ← matrix(NA, nrow = dim(fki_raw)[1], ncol = dim(fki_raw)[2])
dimnames(fki_stand) ← dimnames(fki_raw)

library(scales) # I wish this function could have "by.col = T". Oh well.
for (.j in 1:dim(fki_raw)[2]) {
    fki_stand[, .j] ← rescale(fki_raw[, .j], to = c(0.01, 0.99))
}

rm(fki_raw)

fki_stand ← data.frame(fki_stand)
attach(fki_stand)

# Subsection 5.1: Economic capacity (sub_ind_ec)

sub_ind_ec ← gdp_per_capita

# Subsection 5.2: Education and training (sub_ind_et)

wt_highly_skilled ← 1 / sd(highly_skilled)
wt_mean_year_of_schooling ← 1 / sd(mean_year_of_schooling)

sub_ind_et ← (highly_skilled^wt_highly_skilled *
    mean_year_of_schooling^wt_mean_year_of_schooling)^
    (1 / (wt_highly_skilled + wt_mean_year_of_schooling))

rm(list = ls(pattern = "^wt"))

# Subsection 5.3: Use (sub_ind_use)
```

```
sub_ind_u <- (gpea + ica)^ius

# Subsection 5.4: Need (sub_ind_need)

sub_ind_n <- (pfa + ac)^ageing

## Subsection 5.5: FKI

wt_ec <- 1 / sd(sub_ind_ec)
wt_et <- 1 / sd(sub_ind_et)
wt_u <- 1 / sd(sub_ind_u)
wt_n <- 1 / sd(sub_ind_n)

fki <- (
    sub_ind_ec^wt_ec *
        sub_ind_et^wt_et *
        sub_ind_u^wt_u *
        sub_ind_n^wt_n
) ^ (
    1 / (wt_ec + wt_et + wt_u + wt_n)
)

rm(list = ls(pattern = "^wt"))

l3 <- data.frame(
    round(
        cbind(fki, sub_ind_ec, sub_ind_et, sub_ind_u, sub_ind_n),
        digits = 3
    )
)
rownames(l3) <- rownames(fki_stand)
attach(l3)

rm(fki_stand, fki, sub_ind_ec, sub_ind_et, sub_ind_u, sub_ind_n)

# Display country-level FKI, default by country code
l3

# Sort FKI by country (highest to lowest)
l3_ordered <- l3[order(-fki), ]
l3_ordered
fwrite(l3_ordered, file = "fki.csv", row.names = T)

pdf("../../Figures/FKI.pdf")
    barplot(l3_ordered$fki,
        names.arg = rownames(l3_ordered),
        xlab = "Country", las = 2, ylab = "Financial Knowledge Index (FKI)",
        ylim = c(0, 1), main = "FKI of 20 participating countries"
    )
dev.off()
```

B.2.4 Data Reimport

```
# Housekeeping
library(Orcs)
setwdOS(lin = "~/", win = Sys.getenv("USERPROFILE"))

# Import SPSS file into R
library(intsvy)
```

```r
finlit <- pisa.select.merge(
    student.file = "CY07_MSU_FLT_QQQ.SAV", # file ext in capital
    school.file = "CY07_MSU_SCH_QQQ.sav", # file ext in lower case
    student = c(
    # Control variables
        "ST004D01T", # Student (Standardized) Gender
        "IMMIG", # Index Immigration status
        "ESCS", # Index of economic, social and cultural status
    # Mediators
        "FCFMLRTY", # Familiarity with concepts of finance (Sum)
        "FLCONFIN", # Confidence about financial matters (WLE)
    # Academic
        "FLSCHOOL", # Financial education in school lessons (WLE)
    # Safety
        "BEINGBULLIED", # Student's experience of being bullied (WLE)
    # Community
        "FLFAMILY" # Parental involvement in matters of Financial Literacy (WLE)
    ),
    school = c(
        "STRATIO", # Student-teacher ratio
        "EDUSHORT" # Shortage of educational material (WLE)
    ),
    countries = c(
        "BRA", "BGR", "CAN", "CHL", "EST",
        "FIN", "GEO", "IDN", "ITA", "LVA",
        "LTU", "NLD", "PER", "POL", "PRT",
        "RUS", "SRB", "SVK", "ESP", "USA"
    )
)

names(finlit)
# Throw away columns that I do not need
finlit <- finlit[, -c(5,7:86)] # 5 = BOOKID; 7:86 = resampling weights

# Some var need recording
library(car)

# Re-code Russian territories to RUS
finlit$CNT <- recode(finlit$CNT, "
    'QMR' = 'RUS';
    'QRT' = 'RUS'
")

# Input country-level FKI
FKI <- recode(finlit$CNT, "
    'NLD' = 0.940;
    'USA' = 0.937;
    'CAN' = 0.784;
    'ITA' = 0.762;
    'FIN' = 0.724;
    'ESP' = 0.627;
    'LTU' = 0.613;
    'PRT' = 0.591;
    'BGR' = 0.583;
    'EST' = 0.577;
    'SVK' = 0.559;
    'POL' = 0.555;
    'LVA' = 0.550;
    'CHL' = 0.544;
    'RUS' = 0.450;
    'GEO' = 0.424;
```

```r
    'SRB' = 0.423;
    'PER' = 0.309;
    'BRA' = 0.141;
    'IDN' = 0.122
")

# Recode ST004D01T from Sex to Male
MALE <- finlit$ST004D01T - 1

# Recode IMMIG to 1st and 2nd generation
IMMI1GEN <- recode(finlit$IMMIG, "
    1 = 0;
    2 = 0;
    3 = 1
")

IMMI2GEN <- recode(finlit$IMMIG, "
    1 = 0;
    2 = 1;
    3 = 0
")

# Revert coding direction: bigger number => safer school
NOBULLY <- finlit$BEINGBULLIED * (-1)

# Stitch spreadsheet together
names(finlit)
finlit <- cbind(FKI, finlit[, c(2:35)], MALE, IMMI1GEN, IMMI2GEN, finlit[, c(38:41)], NO
head(finlit)
names(finlit)

# Remove cases whose school weights (col #45) are NA
obs0 <- dim(finlit)[1]
finlit <- finlit[complete.cases(finlit[, 45]), ]
obs1 <- dim(finlit)[1]
obs0 - obs1 # 12 cases contained missing school weights and have been dropped
rm(obs0, obs1)

# Use data.table for better RAM management
library(data.table); setDTthreads(0) # 0 means all the available cores
# Export data into a CSV file for faster import next time
fwrite(finlit, file = "finlit.csv", na = "NA", row.names = F, col.names = T)
```

Appendix C Derivation of Moderated Mediation Effect

C.1 Models with Mediators Only

Consider a SEM model shown in Figure C.1 (excluding any paths in green), where

$$\begin{cases} Y = \mu_0 + b_1 M_1 + b_2 M_2 + c_1 X_1 + c_2 X_2 + c_3 X_3 \\ M_1 = \mu_1 + a_{11} X_1 + a_{21} X_2 + a_{31} X_3 \\ M_2 = \mu_2 + a_{12} X_1 + a_{22} X_2 + a_{32} X_3 \end{cases}$$

or, in matrix form

$$\begin{cases} Y = \mu_0 + \boldsymbol{b}^\intercal \boldsymbol{m} + \boldsymbol{c}^\intercal \boldsymbol{x} \\ \boldsymbol{m} = \boldsymbol{\mu} + \boldsymbol{A}^\intercal \boldsymbol{x} \end{cases} \tag{C.1}$$

where

$$\underset{3\times 1}{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \ \underset{2\times 1}{\boldsymbol{m}} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \ \underset{2\times 1}{\boldsymbol{b}} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \ \underset{3\times 1}{\boldsymbol{c}} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, \ \underset{2\times 1}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \underset{3\times 2}{\boldsymbol{A}} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

Equation (C.1) can be written as a total equation:

$$Y = \mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{b}^\intercal \boldsymbol{A}^\intercal \boldsymbol{x} + \boldsymbol{c}^\intercal \boldsymbol{x} = \left(\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu}\right) + \boldsymbol{x}^\intercal (\boldsymbol{A}\boldsymbol{b} + \boldsymbol{c}) \tag{C.2}$$

where $\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu}$ is the intercept, $\boldsymbol{Ab}$ is the indirect effect and $\boldsymbol{c}$ is the direct effect.

C.2 Models with Moderated Mediators

Now introduce two moderators D_1 and D_2 (green paths in Figure C.1).

In scalar notation:

$$
\begin{aligned}
Y_{\text{mod}} = \mu_0 &+ b_1 M_1 + b_2 M_2 + c_1 X_1 + c_2 X_2 + c_3 X_3 \\
&+ f_1 D_1 + f_2 D_2 \\
&+ g_{11} X_1 D_1 + g_{12} X_1 D_2 \\
&+ g_{21} X_2 D_1 + g_{22} X_2 D_2 \\
&+ g_{31} X_3 D_1 + g_{32} X_1 D_2 \\
&+ h_{11} M_1 D_1 + h_{12} M_1 D_2 \\
&+ h_{21} M_2 D_1 + h_{22} M_2 D_2
\end{aligned}
$$

and in matrix notation:

$$
Y_{\text{mod}} = \mu_0 + \boldsymbol{b}^\intercal \boldsymbol{m} + \boldsymbol{c}^\intercal \boldsymbol{x} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{G}^\intercal \boldsymbol{x}\boldsymbol{d}^\intercal\right) + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{m}\boldsymbol{d}^\intercal\right) \tag{C.3}
$$

where,

$$
\underset{2\times 1}{\boldsymbol{f}} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \quad \underset{2\times 1}{\boldsymbol{d}} = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}, \quad \underset{3\times 2}{\boldsymbol{G}} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \\ g_{31} & g_{32} \end{pmatrix}, \quad \underset{2\times 2}{\boldsymbol{H}} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix},
$$

and $\operatorname{tr}(\cdot)$ is the trace operator.

Since $\boldsymbol{m} = \boldsymbol{\mu} + \boldsymbol{A}^\intercal \boldsymbol{x}$, Equation (C.3) can be expanded into:

$$
\begin{aligned}
Y_{\text{mod}} &= \mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{b}^\intercal \boldsymbol{A}^\intercal \boldsymbol{x} + \boldsymbol{c}^\intercal \boldsymbol{x} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{G}^\intercal \boldsymbol{x}\boldsymbol{d}^\intercal\right) + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right) + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{A}^\intercal \boldsymbol{x}\boldsymbol{d}^\intercal\right) \\
&= \left[\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right)\right] + \left[\left(\boldsymbol{b}^\intercal \boldsymbol{A}^\intercal + \boldsymbol{c}^\intercal\right)\boldsymbol{x} + \operatorname{tr}\left(\boldsymbol{d}^\intercal \left(\boldsymbol{G}^\intercal + \boldsymbol{H}^\intercal \boldsymbol{A}^\intercal\right)\boldsymbol{x}\right)\right] \\
&= \left[\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right)\right] + \left[\left(\boldsymbol{b}^\intercal \boldsymbol{A}^\intercal + \boldsymbol{c}^\intercal\right)\boldsymbol{x} + \boldsymbol{d}^\intercal \left(\boldsymbol{G}^\intercal + \boldsymbol{H}^\intercal \boldsymbol{A}^\intercal\right)\boldsymbol{x}\right] \\
&= \left[\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right)\right] + \boldsymbol{x}^\intercal \left[\boldsymbol{A}\boldsymbol{b} + \boldsymbol{c} + \boldsymbol{G}\boldsymbol{d} + \boldsymbol{A}\boldsymbol{H}\boldsymbol{d}\right] \\
&= \left[\mu_0 + \boldsymbol{b}^\intercal \boldsymbol{\mu} + \boldsymbol{f}^\intercal \boldsymbol{d} + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right)\right] + \boldsymbol{x}^\intercal \left[\boldsymbol{A}\left(\boldsymbol{b} + \boldsymbol{H}\boldsymbol{d}\right) + \left(\boldsymbol{c} + \boldsymbol{G}\boldsymbol{d}\right)\right]
\end{aligned} \tag{C.4}
$$

Equation (C.4) differs from Equation (C.2) by one extra term $\boldsymbol{f}\boldsymbol{d}^\intercal + \operatorname{tr}\left(\boldsymbol{H}^\intercal \boldsymbol{\mu}\boldsymbol{d}^\intercal\right)$ in the intercept. The indirect effect $\boldsymbol{A}\boldsymbol{b}$ expanded to $\boldsymbol{A}\left(\boldsymbol{b} + \boldsymbol{H}\boldsymbol{d}\right)$ as a result of introducing the moderators and the direct effect grows from $\boldsymbol{c}$ to $\boldsymbol{c} + \boldsymbol{G}\boldsymbol{d}$.

Expand the indirect and direct effects back to their scalar forms:

$$
\begin{aligned}
&\text{indirect effects} \\
=&\boldsymbol{A}\left(\boldsymbol{b}+\boldsymbol{H}\boldsymbol{d}\right) \\
=&\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \left[\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}\right] \\
=&\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} b_1 + h_{11}D_1 + h_{12}D_2 \\ b_2 + h_{21}D_1 + h_{22}D_2 \end{pmatrix} \\
=&\begin{pmatrix} a_{11}b_1 + a_{11}h_{11}D_1 + a_{11}h_{12}D_2 + a_{12}b_2 + a_{12}h_{21}D_1 + a_{12}h_{22}D_2 \\ a_{21}b_1 + a_{21}h_{11}D_1 + a_{21}h_{12}D_2 + a_{22}b_2 + a_{22}h_{21}D_1 + a_{22}h_{22}D_2 \\ a_{31}b_1 + a_{31}h_{11}D_1 + a_{31}h_{12}D_2 + a_{32}b_2 + a_{32}h_{21}D_1 + a_{32}h_{22}D_2 \end{pmatrix}; \\
&\text{direct effects} \\
=&\boldsymbol{c}+\boldsymbol{G}\boldsymbol{d} \\
=&\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \\ g_{31} & g_{32} \end{pmatrix} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} \\
=&\begin{pmatrix} c_1 + g_{11}D_1 + g_{12}D_2 \\ c_2 + g_{21}D_1 + g_{22}D_2 \\ c_3 + g_{31}D_1 + g_{32}D_2 \end{pmatrix}.
\end{aligned}
$$

C.3 Mplus Execution

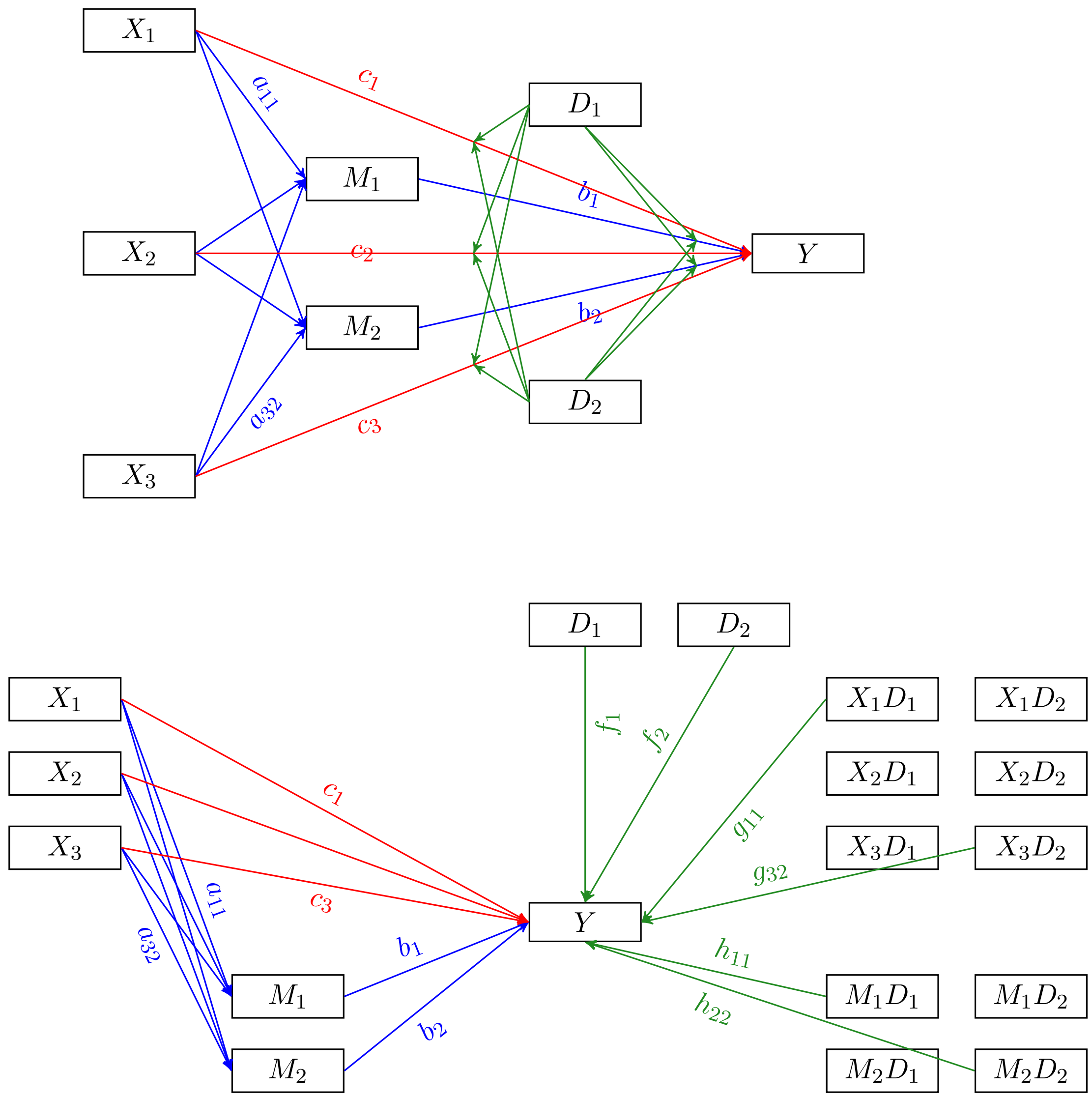
The `DEFINE:` and `MODEL:` sections of the Mplus code is given as following:

```
DEFINE:

    ! G matrix
    X1D1 = X1 * D1;
    X2D1 = X2 * D1;
    X3D1 = X3 * D1;
    X1D2 = X1 * D2;
    X2D2 = X2 * D2;
    X3D2 = X3 * D2;
```

```
    ! H matrix
    M1D1 = M1 * D1;
    M2D1 = M2 * D1;
    M1D2 = M1 * D2;
    M2D2 = M2 * D2;

MODEL:

    [Y] (mu0);
    Y on M1 (b1);
    Y on M2 (b2);
    ! ---
    Y on M1D1 (h11);
    Y on M2D1 (h21);
    Y on M1D1 (h12);
    Y on M2D1 (h22);
    ! ---
    Y on X1 (c1);
    Y on X2 (c2);
    Y on X3 (c3);
    ! ---
    Y on D1 (f1);
    Y on D2 (f2);
    ! ---
    Y on X1D1 (g11);
    Y on X2D1 (g21);
    Y on X3D1 (g31);
    Y on X1D2 (g12);
    Y on X2D2 (g22);
    Y on X3D2 (g32);

    [M1] (mu1);
    M1 on X1 (a11);
    M1 on X2 (a21);
    M1 on X3 (a31);

    [M2] (mu2);
    M2 on X1 (a12);
    M2 on X2 (a22);
    M2 on X3 (a32);

```

Figure C.1

Moderated Mediation Model



Note. A moderated mediation is shown in both model diagram (upper panel) and statistical diagram (lower panel). Direct paths, indirect paths and moderations are differentiated by colour.