README for: “Since You’re So Rich, You Must Be Really Smart”: Talent, Rent Sharing, and the Finance Wage Premium

Data availability and provenance

The main data used in our paper *“Since You’re So Rich, You Must Be Really Smart”: Talent, Rent Sharing, and the Finance Wage Premium* is from confidential register records provided by Statistics Sweden (SCB). The sources of specific variables are explained in our paper and summarized further below.

There is a standardized procedure to apply for data access with SCB. An application includes a detailed research plan, a list of variables from the different registers, and an approval of the ethical review board. The official instructions can be found at https://www.scb.se/en/services/guidance-for-researchers-and-universities/. The agency’s microdata unit can be contacted at mikrodata@scb.se.

SCB performs a judicial clearance of the order and quotes a price. Once the data has been processed, it can be accessed via Statistics Sweden’s access servers (the system is called Microdata Online Access – MONA). MONA is currently only accessible from EU countries or countries meeting the GDPR requirements of the EU.

The replication scripts, including detailed explanations of data construction, etc. are available at https://doi.org/10.5281/zenodo.6929238.

Source data list

LISA (SCB administrative data): Statistics Sweden’s “Longitudinal Integration Database for Health Insurance and Labour Market Studies” – LISA. We use data from 1990 to 2017. Information about the data and variables can be found at http://www.scb.se/en_/Services/Guidance-for-researchers-and-universities/SCB-Data/Longitudinal-integration-database-for-health-insurance-and-labour-market-studies-LISA-by-Swedish-acronym/.

LINDA (SCB administrative data): Statistics Sweden’s “Longitudinal Individual Data Base” – LINDA. We use data from 1978 to 1990. Information about the data and variables can be found at https://www.scb.se/hitta-statistik/statistik-efter-amne/hushallens-ekonomi/inkomster-och-inkomstfordelning/longitudinell-individdatabas-linda/produktrelaterat/Fordjupad-information/mikrodata-for-longitudinell-individdatabas-linda/.

Krigsarkivet and Rekryteringsmyndigheten (SCB administrative data): Data from Military Archives and Swedish Defence Recruitment Agency. We use this for talent measures of cohorts enlisted from 1969 to 1983 and 1983 to 2010, respectively. Linked to the other data within MONA via the individual’s unique person number. Documentation of this data can be obtained directly from the respective institutions. The data is also described in detail in Lindqvist and Vestman (2011).

High-School Register (SCB administrative data): Information about this data can be found at https://www.scb.se/vara-tjanster/bestall-data-och-statistik/bestalla-mikrodata/vilka-mikrodata-finns/individregister/ (see under “Utbildning”).

Multi-Generation Register (SCB administrative data): We use data on biological parent-child links from Statistics Sweden’s multi-generation register. Linked to the other data within MONA via the individual’s unique person number. Information about this register can be found at https://www.scb.se/vara-tjanster/bestall-data-och-statistik/bestalla-mikrodata/vilka-mikrodata-finns/individregister/flergenerationsregistret/.

Swedish and US Industry Accounts (public data): We use this data from Statistics Sweden (SCB) and US Bureau of Economic Analysis (BEA) to measure relative employment and value added in the Swedish financial sector. We also use BEA data and follow Philippon and Reshef (2012) to extend their US relative wage series in finance to 2007–2018. The series up to 2006 was provided to us by Ariell Reshef directly. The Swedish industry accounts can be accessed at https://www.scb.se/en/finding-statistics/. The US industry accounts can be accessed at https://www.bea.gov/data/.

Serrano and UC (proprietary data): The Serrano dataset provides accounting data for limited liability firms in the nonfinancial private sector (Aktiebolagets (ABs)). ABs are required to file annual financial statements (balance sheet and income statement) with the Swedish Companies Registration Office (Bolagsverket). (Electronic) copies of those statements can be obtained from Bolagsverket. Bisnode AB (now owned by Dun & Bradstreet), a commercial vendor, has collected and processed that data for all ABs from 1998 onward in the Serrano database. The Serrano database also contains information on business groups, allowing the link of different subsidiaries to the corresponding business group. We use Serrano data from 1998 onward. Before 1998, we use corresponding data collected and sold by credit bureau Upplysningscentralen AB (UC). We process the data ourselves following Serrano's procedures. Serrano and UC data are linked to the other data within MONA via the firms' unique organization numbers. More information about Serrano and UC can be obtained directly from the respective institutions.

Finansinspektionen and Finansiella Företag (proprietary data): Data on financial sector firms' annual accounts from the Swedish financial supervisory authority – Finansinspektionen – and available from 1996 onward. Before 1996, we hand-collect data from financial companies' yearbooks – Finansiella Företag – published in print by Statistics Sweden. Finansinspektionen and Finansiella Företag are linked to the other data within MONA via the firm's unique organization number. Information about these data can be obtained directly from the respective institutions.

Description of intermediate and analysis datasets

workdata_LISAnew.dta

Population definition:

- All individuals in LISA (i.e., all residents of Sweden from age 16 onward) who work as non-selfemployed in the non-primary private sector and earn above the basic taxation value (prisbasbelopp) in the respective calendar year.

Panel structure:

- Annual panel data during 1990–2017.

Construction:

- Start with LISA and link with military test scores from Krigsarkivet / Rekryteringsmyndigheten and high school grades from High-School Register. Also link with business group information from Serrano.
- Drop observations with missings in any of: sex, potential experience, broad industry sector. Also drop observations when declared labor income (deklon) below the prisbasbelopp, when self-employed, or when industry of employment is primary or public sector.

network_prep.dta

Population definition:

- All individuals in LISA (i.e., all residents of Sweden from age 16 onward).

Panel structure:

- Annual panel data during 1990–2017.

Construction:

- Start with LISA and link with multi-generational register.
- For each municipality and each year, calculate share of workers employed in different sectors (broad sector level).
- For each municipality and each year, calculate share of individuals moving to a different municipality.
- Obtain sector of employment (broad sector level) of biological father in every year.
- Obtain municipality (and state) information of residence at age 16.

linda_prep.dta

Population definition:

- All individuals in LINDA in the private sector.

Panel structure:

- Annual panel data during 1978–1991.

Construction:

- Processed similar to main dataset LISA.

relwage_US.dta

Population definition:

- U.S. relative wages in the financial sector.

Panel structure:

- Annual time series during 1978–2018.

Construction:

- Period 1978–2006 directly obtained from Ariell Reshef.
- For the remainder of the period, processed industry data from the Bureau of Economic Analysis (BEA) following Philippon and Reshef (2012).

bankperform3_workdata.dta (intermediate dataset)

Population definition:

- Financial firms' accounts where those are available.

Panel structure:

- Annual panel data during 1991–2017.

Construction:

- Append Finansinspektionen raw data with Finansiella Företag for the years before 1996.
- Apply variable definitions for value added, total assets, operating profits, and equity. Interpolate missing values when non-missing values available in a prior and a subsequent year.

ucSer_workdataNew.dta (intermediate dataset)

Population definition:

- Non-financial firms' accounts where those are available.

Panel structure:

- Annual panel data during 1991–2017.

Construction:

- Append Serrano data with UC for the years before 1998.
- Apply variable definitions for value added, total assets, operating profits, sales, equity, depreciation, and personnel costs.

LISA_firmaccounts.dta

Population definition:

- LISA data matched with financial and non-financial firms' accounts.

Panel structure:

- Annual panel data during 1991–2017.

Construction:

- Start with **workdata_LISAnew.dta** and link with **ucSer_workdataNew.dta** and **bankperform3_workdata.dta** according to the individual's employer's unique organization number.

Computational details

The code for the analysis of the administrative data was run on Statistics Sweden's secure server MONA, which is a virtual server in VMware with 40 threads, 500 GB of RAM, 500 GB of fast storage.

The industry accounts data for Sweden and the US were last processed on an Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz Laptop with 16GB of ram and Windows 11.

The software Stata 17MP was used in both cases.

References

Bisnode AB. " *The Serrano Database for Analysis and Register-Based Statistics, 1998-2018 [database].*" accessed 2020.

Finansinspektionen. 2022. "*Annual Accounts of Financial Firms, 1996-2018 [database].*" accessed 2020.

Lindqvist, Erik, and Roine Vestman. 2011. "*The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment.*" American Economic Journal: Applied Economics, 3 (1): 101-28.

Philippon, Thomas, and Ariell Reshef. "*Wages and human capital in the US finance industry: 1909–2006.*" The Quarterly Journal of Economics 127.4 (2012): 1551-1609.

Statistics Sweden (SCB). "*Financial Enterprises 1988–1998 (Finansiella företag).*" ISSN 1103-5234; Published: Stockholm : SCB, 1993-2004; Produced: Örebro : SCB-tr.

Statistics Sweden (SCB). 2022. "*Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA), 1990-2017 [database].*" accessed 2020.

Statistics Sweden (SCB). 2022. "*Longitudinal Individual Data Base (LINDA), 1978-1990 [database].*" accessed 2020.

Statistics Sweden (SCB). 2022. "*Military Archives (Krigsarkivet), 1969-1983 [database].*" accessed 2020.

Statistics Sweden (SCB). 2022. "*Recruitment Agency (Rekryteringsmyndigheten), 1983-2010 [database].*" accessed 2020.

Statistics Sweden (SCB). 2022. "*High-School Register (Registret över slutbetyg från gymnasieskolan) [database].*" accessed 2020.

Statistics Sweden (SCB). 2022. "*Multigenerational Register (Flergenerationsregistret) [database].*" accessed 2020.

Statistics Sweden (SCB). " *GDP Accounts (Current Prices) and National Employment Statistics.*" https://www.scb.se/en/finding-statistics/ (accessed 23 November 2020).

UC (2022). "*Annual Reports 1989-1998 [database].*" Upplysningscentralen (UC), accessed (2022).

U.S. Bureau of Economic Analysis. "*National Income and Product Accounts.*" https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2&isuri=1&1921=survey (accessed 23 November 2020).

Do-files	Input	Output
Figure_1.do	Figure 1 - Finance share of GDP and labor - replication file.xlsx sheets “Sweden” and “US”, Figure_1.do	Figure 1 -...xlsx sheet “Graphs” Fig1A_swe_va_empl.pdf, Fig1B_usa_va_empl.pdf
Figure_2A.do	linda_prep.dta workdata_LISAnew.dta relwage_US.dta	Fig2A_relwage_LisaLindaUS.pdf
Figure_2BC.do	workdata_LISAnew.dta	Fig2B_relwage_Qreg.pdf Fig2C_relwage_Qreg99.pdf
Figure_3.do	workdata_LISAnew.dta	Fig3_reltalent.pdf
Figure_4.do	workdata_LISAnew.dta	Fig4A_distr_ratio_cog.pdf Fig4B_distr_ratio_noncog.pdf
Figure_5.do	workdata_LISAnew.dta	Fig5A_talentcontrol.pdf Fig5B_FEcontrol.pdf Fig5C_educcontrol.pdf Fig5D_interactcontrol.pdf
Figure_6.do	workdata_LISAnew.dta	Fig6A_deklon_ratio_cog.pdf Fig6B_deklon_ratio_noncog.pdf
Figure_7.do	LISA_firmaccounts.dta	Fig7_rel_VA_wage_perworker.pdf Figure_7.log
Figure_8A.do	workdata_LISAnew.dta	Fig8A_entry_young_old.pdf
Figure_8B.do	workdata_LISAnew.dta	Fig8B_wagegrowth_3045.pdf
Table_1.do	workdata_LISAnew.dta	Table_1.log (lines 84-100, 121-137, 160-176, 199-257)
Table_2.do	workdata_LISAnew.dta	Table_2.log (lines 77-88, 114-125, 151-162)
Table_3.do	workdata_LISAnew.dta	Table_3.log (lines 415-630), Table_3.xlsx
Table_4.do	LISA_firmaccounts.dta	Table_4.log (lines 205-217)
N/A	Values copied into excel from Table_4.log (lines 95-199) and Figure_7.log (lines 118-140)	Table_5.xlsx
Table_6.do	workdata_LISAnew.dta network_prep.dta	Table_6.log (lines 160-192)