Social Push and the Direction of Innovation*

Elias Einio, *VATT Institute for Economic Research*
Josh Feng, *University of Utah – Eccles School of Business*
Xavier Jaravel, *London School of Economics*
[PRELIMINARY, NOT FOR CIRCULATION]

October 2020

Abstract

Innovators' social networks may affect their familiarity with customer needs, and in turn the types of products they bring to market. Consistent with this channel, we document that innovators create products that are more likely to be purchased by customers similar to them along observable dimensions including gender, age, and socio-economic status. With scanner data and a new phone applications database, we find that these homophily patterns hold even within detailed industries. Using quasi-random assignment of individuals to dorms during military service, we provide causal evidence that being exposed to peers from a lower income group increases an entrepreneur's propensity to create necessity products. The finding is similar with an alternative research design leveraging idiosyncratic within-school variation in peer composition across classes and cohorts. Because innovators are predominantly men from privileged backgrounds, the social push channel implies that the gains from innovation are unequally distributed across customer groups, which we quantify in a growth model.

*We are grateful to Corina Boar, Raj Chetty, Louphou Coulibaly, Ben Jones, Kieran Larkin, Katja Mann, Tommaso Porzio, and Heidi Williams for thoughtful comments. The article also benefited from seminar and conference participants at the ASSA meetings, BYU-Utah Joint Research Day, and VMACS. We also thank Gabriel Leite Mariante for excellent research assistance. The permission to use Conscript Registry data is based on the decision number AM4503 of the Finnish Defense Forces. Results reflect researchers' own analyses calculated based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Feng acknowledges support from the National Institute on Aging, Grant Number T32-AG000186 and from the Ewing Marion Kauffman Foundation. The contents of this publication are solely the responsibility of the authors.

I Introduction

What governs the direction of innovation? Much of the economics literature focuses on market size and financial incentives as the main endogenous drivers of the direction of innovation (e.g., Jacob (1966), Acemoglu (2002), Acemoglu (2007)), but other factors may be at play. For example, the discovery of entrepreneurial opportunities depends on the distribution of information in society (e.g., Hayek (1945)) and often requires engagement with specific real-world problems and users (e.g., Von Hippel (1986), Shane (2000)). In particular, an individual's social network may affect the types of problems and customer needs they are familiar with, and thus the kinds of innovations they might bring to market. Despite the apparent plausibility of this "social push" channel, little is known about its relevance for the direction of innovation in practice.

The social push channel could lead to important distributional effects across consumer groups because innovators are not representative of society at large. For example, women, minorities and individuals from low-income backgrounds are under-represented among innovation leaders including startup founders, patent inventors and venture capitalists. Although a growing literature documents such gaps in access to innovation and entrepreneurship (e.g., Aghion et al., 2017, Bell et al., 2017, Hvide and Oyer, 2019, Agarwal and Gaulé, 2020), much less is known about their impacts on the *direction* of innovation, and in turn on the distribution of the gains from innovation in society.

In this paper, we present new stylized facts and quasi-experimental evidence on the relationship between innovators' backgrounds, peer effects, and the direction of innovation. We then assess the quantitative importance of the social push channel for inequality in general equilibrium using a standard growth model. Overall, we find that the under-representation of minorities affects the direction of innovation in a way that exacerbates inequality and the gender gap in purchasing power.

In the first part of the article, we use data from the United States and Finland to present new facts about the direction of innovation and innovators' socio-demographic backgrounds. We build a data set linking consumer characteristics to innovators' parent income and gender. Consumer characteristics are measured in comprehensive consumption surveys, in detailed scanner data for consumer packaged goods, and in a new data set covering phone applications. Innovators and their backgrounds are identified from patent records, start-up and VC databases, registries of firms, and administrative tax records.

We find that innovators from a high-income family are more likely to target high-income consumers. For example, people from high-income families are less likely to get a patent or start a firm within a "necessity" industry like food, but are more likely to do so in a "luxury" industry like finance. Moreover, we find that female consumers are significantly more likely to purchase products from startups that were founded by female entrepreneurs, and that female entrepreneurs are more likely to hire female inventors and obtain funding from a female venture capital partner. These homophily patterns hold across detailed industries as well as across firms within the same industry, in both the United States and Finland. We also find evidence of homophily in terms of age and geography.

In the second part of the article, we study whether innovators' social networks have a causal impact on the direction of innovation. We use the quasi-random assignment of individuals to dorms during military service in Finland to obtain variation in peer groups (similar to Einiö (2019)). We then examine whether exposure to peers from a different income backgrounds has an impact on the direction of entrepreneurship. We find that exposure to a lower-income peer increases the probability of starting a business in a necessities industry (conditional on being an entrepreneur). We confirm this finding with a second quasi-experimental research design using data on study peers at Finnish universities, where peer effects are identified from idiosyncratic within-school variation in peer composition across classes and cohorts (similar to Hoxby (2000)). These results provide direct evidence that peers have a causal effect on the direction of innovation, which stands in contrast with the mechanisms based on economic incentives, such as market size, which are typically emphasized in the economics literature.

In the final part of the article, we present a quantitative assessment of the relevance of these findings for inequality and growth. We first use a static framework to assess the distributional effects of unequal access to innovation. We find that equalizing access to innovation could substantially reduce inequality and increase real income growth at the median of the income distribution. In a CES framework with love-of-variety based on Feenstra (1994), we find that transitioning from the current distribution of innovators' backgrounds to an "equal opportunity" innovator pool would lead to much larger welfare gains for lower-income households. Furthermore, we find that the gender gap in innovation strengthens the gender "real wage" gap through the direction of innovation. In the CES framework, we find that, compared with a male-led startup, a female-led startup generates welfare gains for female consumers that are 27% larger than for male consumers. These results indicate that equalizing access to innovation across groups can help create inclusive innovations targeting a broad consumer base.

To complement the static framework, we also analyze the importance of these channels in a dynamic general equilibrium model building on standard growth models. The model features many sectors, heterogeneity in research productivity and consumer tastes, entry barriers that vary across socio-demographic groups, and interactions between groups that can alleviate these barriers. We derive the balanced-growth path and steady-state inequality in this setting.

The model clarifies the assumptions under which entry barriers matter for long-run growth and steady-state inequality. We first show that under a model with homogeneous preferences and research ability, entry barriers do not matter for long-run growth. However, with heterogeneous ability or heterogeneous consumer preferences, growth rates fall. Furthermore, when household groups have a comparative advantage at innovating in the sector for which they have a stronger preference, access barriers both reduce long-run growth and increase inequality across groups. Finally, motivated by our quasi-experimental results, we model social connectivity as increasing an agent's ability to innovate for other social groups, which generally leads to higher long-term growth and lower inequality.

This article primarily contributes to the literature on unequal access to innovation. Recent research documented that certain groups are underrepresented in the innovation system, particularly in terms of gender, race, parental background, and geography (Cook and Kongcharoen, 2010; Bell et al., 2017; Toole et al., 2019). Several articles have studied potential mechanisms that can explain this underrepresentation, including funding barriers (Brooks et al., 2014; Malmström et al., 2017; Kanze et al., 2018; Guzman and Kacperczyk, 2018; Hvide and Oyer, 2019), preferences (Thébaud, 2010; Bönte and Piegeler, 2013; Caliendo et al., 2015), and intergenerational transmission (Dunn and Holtz-Eakin, 2000; Mishkin, 2017; Hvide and Oyer, 2019). Consistent with our findings, contemporaneous work by Koning et al. (2019) documents that biomedical patents with female first authors are more likely to mention female medical conditions. The idea that women are underserved in a wide variety of settings is the focus in Criado Perez (2019). Our work builds on and extends this literature by providing empirical evidence on the market and consumer welfare impacts of unequal access to innovation.

This article also contribute to the literature on endogenous growth and inequality. Our model builds on the product variety literature, starting with Romer (1990), by adding heterogeneity in research productivity and multiple sectors. The model is also related to Rivera-Batiz and Romer (1991), who analyze the impact of economic integration of two countries, and to Foellmi and Zweimüller (2006), who study endogenous growth when preferences are non-homothetic. By noting the potential importance of social factors in innovation, we contribute to a growing literature on interactions and innovation (Lucas and Moll, 2014; Akcigit et al., 2018). Hsieh et al. (2019) propose an analysis of the impact of misallocation of talent allocation for welfare, which we extend by studying an endogenous growth model. Our results raise the possibility that misallocation in the innovation sector can affect long-run growth rates. Finally, a long-standing literature examines the influence of sorting and social interactions on inequality (Kremer, 1997; Fernandez and Rogerson, 2001). We identify another potential channel through which sorting and peer exposure can affect inequality: the impact of peers on the direction of innovation.

Finally, our results are related to empirical findings in several other fields. In the trade literature, Costinot et al. (2019) find that countries produce and export prescription drugs treating diseases that our prevalent in their own populations. In the development literature, Duflo and Chattopadhydy (2004) find that women leaders are more likely to provide public goods that are relevant to the needs of women. Finally, our results also speak to the industrial organization literature on preference minorities (Waldfogel, 2003; McDevitt and Roberts, 2014; Berry et al., 2016). We show here that inequality across consumers can also result from differential entry costs and innovator specialization.

The rest of the article is organized as follows. Section II presents the data. Section III presents the correlations between the socio-demographic characteristics of innovators and their consumers. Section IV provides quasi-experimental evidence showing that peer background affects entrepreneurs' choice of target markets. Section V examines the implications of these findings for growth and inequality through the lens of standard quantitative models.

II Data, Variable Descriptions and and Summary Statistics

In this section, we describe the data sources, define the sample and key variables used in the analysis, and present summary statistics.

II.A Data Sources

We use two "micro" datasets with respondent-level information on usage of phone applications and consumption of products within consumer packaged goods. We supplement the analysis with industry-level data covering the full consumption basket. We match the consumer information to information on innovators, using patent records (USPTO database) and startup databases (Crunchbase).

Phone applications. The first micro dataset is the Nielsen's Electronic Mobile Measurement (EMM) panel, which to our knowledge has not been used for research purposes. The dataset tracks the mobile application usage of a representative sample of ten thousand US consumers in every month. The data contains detailed information on the gender, race, income, education, and state of residence for each panelist during the period 2017-2019.

The data provides the company name of the developer associated with each application, which we match to information on venture-backed startups in Crunchbase. Crunchbase is a crowdsourced dataset that began tracking information on venture-backed startups and funding events in 2007. For each startup Crunchbase contains data on the name, location, and founders. For each founder, it also records gender and LinkedIn URL, which we use to collect additional information on the founders.[1] We clean names of companies in both the phone app and Crunchbase datasets, and merge based on cleaned name.

Consumer Packaged Goods. To track consumption patterns for consumer packaged goods, we rely on Nielsen's Homescan Consumer Panel, which has been used extensively in prior work. The data tracks between 40k-60k consumers in the period from 2004-2016, and records household-level purchase data, with goods identified by bar code. Each good is classified into 9 departments (e.g. non-food groceries), 118 product groups (e.g. alcoholic beverages), and 1305 product modules (e.g. light beer). For each household, Nielsen also records race, income, education, and family structure, including classifying the household as one of five types (single female, female-led, married, male-led, single male).

To identify the manufacturer associated with each bar code, we use manufacturer prefix data collected by GS1, the organization in charge of allocating bar codes. The data contains the universe of bar code prefixes as of February 2016, with information on the current owner of the prefix. We use the bar code to identify the manufacturer of each purchased good in the Nielsen Consumer Panel. The GS1 data links almost all purchases to a manufacturer (97.5% of total revenue and 98.5% of total quantities).[2]

We match the combined Nielsen-GS1 dataset to information on venture-backed startups in Crunchbase, using the same procedure as for the phone applications dataset. As an additional step, we also manually check pairs of companies that share the same city and first word to look for

[1] Crunchbase uses first name to first guess the gender of each person, and then manually checks for errors. We also verify the data accuracy when we manually collect additional demographic information on each founder.

[2] This excludes bar code entries associated with A.C. Nielsen, which are used to record items without a bar code, and account for 16% of total revenue in the raw data.

additional matches.[3] Finally, we also match all GS1 companies to patent data. This allows us to measure the gender and age composition of the inventors who patent at a given company.

Industry-level data. We leverage several data sets to examine the industry-level patterns in homophily between consumer characteristics and the socio-demographic background of innovation leaders. First, we use data on patent inventors' socio-demographic backgrounds from (Bell et al., 2017). We match inventors' industries to the product categories from the Consumer Expenditure Survey (CEX) to characterize whether inventors from underrepresented groups tend to cater to different populations of consumers. We also use the Panel Survey of Income Dynamics (PSID) to provide additional evidence on family background and the direction of entrepreneurship.

Second, we use Compustat patenting data to assess the relationship between inventor gender or age and consumption patterns from the CEX. Finally, we examine whether similar patterns hold in Finland, using Finnish administrative data covering the full population of Finnish entrepreneurs. Section IV provides a complete description of the Finnish data.

II.B Summary Statistics

Table 1 provides basic summary statistics for the two micro datasets used in the analysis. We find that companies with at least one woman founder represent 14% of venture-backed startups in the phone app industry and 24% of startups in the consumer packaged goods industry. The rates of female venture capital partner involvement are even lower, at 6% and 4%, respectively.[4] We also find underrepresentation in the industry-level data: only 12% of patent inventors are women, and individuals from families with below median income are significantly less likely to become entrepreneurs.

III Stylized Facts

This section documents new descriptive facts on the homophily relationship between innovators and the consumers they serve. After describing the regression specifications, we present the estimates from cross-industry analysis and the estimates within phone applications and consumer packaged goods.

[3]We also tried to match based on word similarity, but find almost all false positives, similar to the results noted in Guzman and Stern (2016).

[4]This is consistent with the observation that venture capital partners tend to be drawn from the pool of successful entrepreneurs, creating a lag in representation.

III.A Research Design

We document correlations at the product level and at the industry level between consumer characteristics and innovator characteristics. Consumer characteristics are measured as the share of sales to certain consumer groups, for example female consumer, young households, low-income households, or households residing in a certain state. Innovators' characteristics are defined similarly, except that we use parental income instead of own income.

Specifically, at the micro level we run regressions of the form

$$SalesShareConsX_{ijk} = \alpha + \beta \times FractionInnovatorX_j + \mu_k + \epsilon_i, \tag{1}$$

where i indexes the good sold by a startup j, k indexes the product category or subcategory, and X denotes the socio-demographic characteristic of interest. We run the analysis at the product level because some startups are active in multiple categories. Standard errors are clustered at the firm level.

At the macro level, we run a similar regression

$$SalesShareConsX_l = \alpha + \beta \times FractionInnovatorX_l + \mu_m + \varepsilon_l, \tag{2}$$

where l indexes the industry, and m is a higher-level industry fixed effects, for example a 2-digit NAICS code. To map patents and inventors to industry, we use the concordance created by Lybbert and Zolas (2014), which provides a probabilistic mapping from US patent class to 6-digit NAICS industry. As a robustness check, we also run regressions for inventors at Compustat firms, using the main NAICS code associated with the firm.

III.B Cross-Industry Patterns

To document industry-level homophily patterns between consumers and entrepreneurs, we first use the PSID to measure the relationship between parent income and the direction of entrepreneurship.

Panel A of Figure 1 is based on the 2017 sample of the PSID, restricted to all individuals who can be matched to their parents. The figure shows the relationship between parent income and the probability of being self-employed. The relationship is strongly upward sloping: the rate of self-employment is about three times as high for individuals whose parents are near the top of the income distribution, compared with individuals whose parents are close to the bottom of the income distribution. The relationship is similar when focusing on high-income entrepreneurs with annual earnings above $100,000 (not reported).

Panel B of Figure 1 restricts the sample to entrepreneurs and uses the sectoral income elasticity estimates measured in Borusyak and Jaravel (2018). This panels depicts the relationship between the income elasticity of sector in which the entrepreneur is active and the income of this entrepreneur's parents. The relationship is upward sloping, i.e. entrepreneurs from a higher-income background tend to cater to richer consumers. The graph indicates that a 10% increase in parent income leads to a comparable 10% increase in income elasticity of the sector the entrepreneur enters, conditional on being an entrepreneur.

Figure 2 examines cross-industry homophily patterns for gender (panel A) and age (panel B). We find that female inventors are significantly more likely to work in industries catering to women, and similarly older patent inventors are more likely to work in industries with a higher share of sales to older households. Table 2 summarizes the homophily patterns by gender and age. The effects are large and statistically significant, and robust across alternative sample (e.g., all industries versus using only firms in Compustat). The quantitative importance of these patterns is discussion in Section 4 in light of a simple welfare framework.

III.C Within Phone Applications and Consumer Packaged Goods

Next, we turn to the results from our micro datasets. Again, we find strong correlations between innovator characteristics and consumer characteristics.

Table 3 presents the estimates. Panel A presents the estimates for phone application startups, which are depicted graphically in Figure 3. We find that female-founded startups have an 8.2pp higher female market share relative to their male counterparts, on a baseline of about 54%. Startups funded by female VCs are also significantly more likely to have a large population of female users. Finally, we also find a very large home-state effect, even after controlling for state population and category fixed effects.

Panel B of Table 3 reports the estimates for consumer packaged goods, which are also shown in Figure 4. We find that female-founded consumer packaged goods startups are 4.7pp more likely to sell to female-led households, on a baseline rate of 25%.[5] This magnitude is similar to the phone apps setting, at about 20% of the baseline rate.

Next, we study the association between patent inventor gender and consumer gender. We can extend this analysis to the full sample of manufacturers in the GS1-Nielsen Consumer Packaged

[5] We focus on female-led households as the outcome here, but in robustness check we obtain similar patterns when analyzing all households weighted by family member gender composition or when focusing single-person households.

Goods dataset, rather than restricting attention to startups. For each manufacturer, we calculate the percentage of female inventors on their patents. We then compute the correlation between this measure and their sales to female households, again finding a significant effect.

Finally, we also find a significant age homophily in the consumer packaged goods sample. Entrepreneurs that are a year older sell to consumers who are 0.1 years older on average.[6]

III.D Extensions on Environmental and Social Impact

An important area of research on the direction of innovation examines the factor that can encourage the innovation of "green patents" (e.g., Acemoglu et al. (2012), Aghion et al. (2016)). We find that female inventors as well as young inventors are more likely to invent "green" patents, i.e. to have positive environmental externalities. Using the data of Aghion et al. (2016), weo study the differences in characteristics of inventors on "clean" versus "dirty" patents (13.3% of the patents are classified as "clean"). We find that 6.5% of inventors of clean patents are female, as opposed to 2.8% for dirty patents. We also find evidence related to age. In the sample of patents with available information on inventor age, we find that younger individuals are more likely to patent in clean energy technologies (0.1 percentage points less likely to work clean patents for one year's increase in age; p-value 0.083).

More broadly, women entrepreneurs appear to have very different focuses, even within narrow product categories. For example, in the overall sample of venture-backed startups in the Crunchbase dataset, female-led startups are more likely to mention "healthy" and "sustainable" and are also more likely to be certified B Corporations. These results are consistent with recent evidence in the literature that women entrepreneurs are motivated by social impact (Guzman et al., 2019).

III.E Discussion

In this section, we have established new empirical facts relating innovator characteristics and consumer characteristics. We discuss here the potential interpretations of the results.

Our data comes from actual market transactions, meaning that many factors during the idea generation, development, and commercialization process can create the correlations documented here. If all individuals pursued the same types of ideas on average and all ideas were developed in the same way, then there would be no correlations in our data. Our results suggest otherwise. First,

[6] We do not find evidence of correlations along racial dimensions, although this may be due to the lack of statistical power. After manually collecting founder race for the set of consumer packaged goods startups, we find that 11% of the founders in are data are Asian, 5.2% are Hispanic, and none are African-American. There is no robust correlation with consumer characteristics, likely due to the small number of minority entrepreneurs.

innovators may differ in terms of the ideas they generate and pursue. For example, this may be a result of them encountering an unmet need, hearing about it through their social interactions, or just being able to better understand the pain points of similar people.[7] Second, project funders may consider innovator characteristics when deciding whether or not to fund a given idea. For example, venture capitalists may only fund an idea targeted at female consumers if the firm has a female founder, which would generate the empirical patterns we find. Funders may also influence how the idea is developed. Finally, there may also be differences in marketing. Companies may directly market the innovator's identity or the innovator may have a role in the design and packaging of the product.

Related to the previous points, we have focused on traditional innovator groups such as venture-backed entrepreneurs and patent inventors, neglecting the role of other potentially important individuals. Entrepreneurs and patent inventors offer a measurable way to relate innovator characteristics to products. However, this ignores the roles of non-patenting innovators, product managers, marketing personnel, and executives, who may play an equally important role in shaping the nature of the final product. Our results may partly reflect systematic correlations in the characteristics of measurable innovators and people in these other roles.

IV Peer Background and the Direction of Innovation: Causal Estimates from Finland

The empirical evidence on homophily between innovators and entrepreneurs raises the possibility that a person's background affects the direction of his or her inventive activity. In this section, we present direct evidence supporting this conjecture by using peers as a source of variation in a person's background. We first show that homophily patterns between entrepreneurs and consumers in Finland are similar to those documented in Section III. We then describe a first quasi-experimental research design, which leverages random assignment to dorms during military service. The results indicate that peer effects have a causal effect on the direction of innovation, which stands in contrast with the mechanisms based on economic incentives (such as market size) that are typically emphasized in the economics literature. Finally, we present a second research design exploiting idiosyncratic within-school variation in peer composition across classes.

[7] This mechanism is related to the line of research on consumer-driven innovation started by Von Hippel (1986).

IV.A Homophily Patterns between Entrepreneurs and Consumers in Finland

In this section, we examine homophily patterns in the Finnish population panel, which covers the whole working aged population in Finland.

Data. The data set is based on administrative registers and it is compiled by Statistics Finland. It provides individual-level information on income, occupation, entrepreneurship, and industry of employment. The data set also includes information on family links, which allows merging parent income to children.

A key variable in the population panel is entrepreneurship status, which is based on pension contribution and income tax records. We use the status for the last week of the year which allows for temporal consistency across variables.[8] An individual is defined as an entrepreneur if she/he has received only entrepreneurial income, and no employee salary income, during the year and is associated with a private business in the entrepreneur pension insurance system in the last week of the year. She/he is also identified as an entrepreneur in the last week of the year if she/he has made entrepreneur pension contributions in that week. If an individual has both entrepreneur and employee pension contributions in the last week of the year, she/he will be defined as an entrepreneur if the entrepreneurial income associated with the contributions is larger.

A second key variable is the unique company identifier that provides information on the company a worker or entrepreneur worked in. This information is based on work spells reported in the national pension systems for entrepreneurs and employees. We use the code for the company an employee/entrepreneur is associated with in the last week of the year.

We link industry income elasticities to the population panel by the industry code of the company an individual is associated with in the last week of the year. The match rate for the sample of individuals with industry codes in the population panel is 80%.[9]

Results. Figure 5 shows industry income elasticities by parent income deciles for individuals who become entrepreneurs (panel A) and for all individuals (panel B). The figure indicates that entrepreneurs who have high-income parents are working in industries with higher average income elasticity. In panel B, the gradient for all individuals is also positive but considerably smaller.

[8] The data includes also codes for the company which is associated with the longest employment spell during the year.

[9] There are 27,292,828 individual-year observations in the population panel over the years 2010-2016. Industry code for the company an individual worked in in the last week of the year is available for 15,930,880 observations. We link industry income elasticities for 6,638,780 individual-year observations by 4-digit industry code. For the remaining unmatched industry codes, we match 1,943,906 at the 3-digit and 4,171,209 at the 2-digit industry level. This result in 12,753,895 matches.

We calculate company-level average parent income of employees from the population panel and link it to entrepreneurs by the company code. Figure A5 shows the average parent income of employees of an entrepreneur by entrepreneur's own parent income decile. The figure shows that the average employee income is higher, on average, among entrepreneurs who come from high-income families.

IV.B Quasi-Experimental Peer Assignment in Finnish Military Service

Research design. To examine whether exposure to social interaction with peers from high-income families could be one mechanism that explains our findings, we follow the approach in Einiö (2019), who estimates the effects of peer parent income among dormmates in the Finnish military conscription, finding positive impacts of peers from high-income families on earnings and wages.

The basic approach is as follows. Within military squadrons of about eighty people, conscripts are assigned to dorms of about eight people. A common way to assign dorms within a squadron is alphabetically. Even within a dorm, individuals tend to bunk with those next to them in alphabetical order, for logistical reasons such as roll call. Therefore, variation in the backgrounds of the two closest alphabetical peers will often translate to variation in peer exposure.

Formally, Einiö (2019) uses the following IV procedures based on the alphabetic rule in assigning conscripts to dorms within squadrons:

$$y_{idst} = \gamma \overline{X}^{(2)}_{(i)ds} + \beta_1 X_{ids} + \alpha_s + \alpha_t + \epsilon_{idst} \tag{3}$$

$$\overline{X}^{(2)}_{(i)ds} = \rho \overline{Z}^{(2)}_{(i)ds} + \theta_1 X_{ids} + \eta_s + \eta_t + \nu_{idst}. \tag{4}$$

where $\overline{X}^{(2)}_{(i)ds}$ is the parent income of two alphabetically nearest dormmates of conscript i in dorm d and squadron s; X_{ids} is the conscript's own parent income; and y_{idst} is the outcome of interest measured in year t. The instrumental variable is the parent income of two alphabetically nearest squadmates of the conscript $\overline{Z}^{(2)}_{(i)ds}$. Conditioning on squad fixed effects, α_s and η_s, means that the model identifies γ from variation in the parent income of a conscript's two alphabetically nearest squadmates across the squad's alphabetic ordering. Einiö, 2019 shows that this instrument is uncorrelated with pre-service characteristics of a conscript and his parents, which indicates that the instrument generates within-squad variation in dormmate parent income that is as good as random.

We merge industry income elasticities with the conscript data by unique person identifiers and estimate the IV model in Equations (3) and (4) for individuals who are entrepreneurs. Sample size is a challenge here, because dorm data is only available for 24% of conscripts.[10] To increase the sample size, we also include individuals who are in managerial occupations.[11]

Results. Table 4 shows the estimates and Figure 6 the corresponding reduced-form plots. In panel A of table 4, the first-stage coefficient on the instrument is large and statistically significant. The reduced-form and IV estimates are all positive, but significant only for the income elasticity outcome. Panel B shows that these effects are concentrated in the sample of individuals who are in the highest own parent income tercile.[12] When the average parent income of two alphabetically nearest dormmates is increased by ten thousand euro, which is equivalent to around one standard deviation, the income elasticity of the industry is increased by around 0.1, or is 8.5% higher compared to the sample mean of 1.22. The results are similar when we use the shares of industry consumption by rich households as the outcomes.

IV.C Quasi-experimental Within-school Variation in Peer Composition across Classes

Research design. To address the sample size issues and confirm the results above, we construct a design based on variation in university peers in Finland. We use data drawn from the student register maintained by Statistics Finland, covering the years 1999-2013. The data provide annual individual-level information on the institution and study program a student is enrolled in.

We define study peers as individuals who start in the same institution and study program in the same year. For individuals who are observed in several programs, we use the last program entered. Our baseline model is a standard linear-in-means peer regression for individual i who starts in program j of school k in year s, controlling for school-by-program fixed effects, school-by-start-year fixed effects, and own parent income:

$$y_{ijkt} = \gamma \overline{X}_{(i)jks} + \beta_1 X_{ijks} + \beta_2 W_{ijkt} + \alpha_{js} + \lambda_{ks} + \epsilon_{ijks}. \tag{5}$$

Here X_{ijks} is own parent income; $\overline{X}_{(i)jks}$ is the leave-own-out mean of parent income of study peers; and y_{idst} is the outcome of interest measured in year t. The terms λ and α are school-by-program and school-by-start-year fixed effects. The parameter of interest is γ, which is the coefficient on

[10] As reported in Einiö (2019), squadron data is available for all conscripts. The individuals in the sample with dorm data available do not appear to be different to the general population.

[11] Information on managerial occupation is based on the occupation code in the population panel.

[12] Individuals from high-income backgrounds appear to be more sensitive to peer exposure, potentially because individuals from low-income backgrounds have already been exposed to ventures targeting high-income consumers through other channels, including the media.

average parent income of study peers.[13] We also include a vector of control variables denoted by W including own and average parent wage earnings and completed years of education, measured one year before the student enters the study program, and dummies for gender, age, and year of outcome measurement. Conditioning on school-by-program fixed effects means that the peer effect is identified from idiosyncratic within-school variation in peer composition across classes. This approach follows several previous studies that have estimated peer effects in education in settings where randomization of students to peer groups is not available (e.g., Hoxby, 2000; Hanushek et al., 2003; Carrell et al., 2018). We follow this approach and assume that, while there is selection into schools and programs, the variation in peer parent income across entering classes of students in the same study program and school is uncorrelated with an individual's own characteristics.

In Table 5, we test for the validity of the empirical design by estimating the impacts on a predicted residual outcome and on pre-peer-exposure characteristics. We construct the predicted residual outcome by first calculating the residuals from a regression of expenditure elasticity on school-by-program fixed effects, school-by-start-year fixed effects, and dummies for age, gender, and year of outcome measurement. We then run a regression of these residuals on own and average parent wage earnings and years of schooling, and construct the predicted outcome as the linear prediction from this model. This variable captures the variation in expenditure elasticity within the same program and school associated with a linear combination of pre-peer-exposure characteristics, where weights are chosen to best predict expenditure elasticity of future industry of activity (see e.g., Carrell et al., 2018). If idiosyncratic variation in peer composition across classes within the same program and school is as good as random, this variable should not be correlated with characteristics of peers to whom an individual is exposed to in later years. Reassuringly, the coefficients for the predicted residual outcome and pre-peer-exposure characteristics are all statistically insignificant.

Results. Table 6 shows the estimates of peer effects on industry expenditure elasticity when an individual is at age 28 or older for a sample of entrepreneurs and for a sample including both entrepreneurs and employees.[14] Column 1 shows a significant positive peer effect for entrepreneurs. Figure 7 shows a graphical presentation of this result. It plots binned averages of the residuals from separate regressions of industry expenditure elasticity and peer parent income on own parent income, school-by-program fixed effects, program-by-start-year fixed effects, and dummies for age,

[13] We note that controlling for own parent income eliminates the potential mechanical correlation between it and the peer mean, which may arise in a peer regression where an individual is allowed to be both the subject of peer effects and the peer (Angrist, 2014).

[14] For employees, the industry elasticity is based on industry of employment.

gender, and year of outcome measurement. A ten thousand euro increase in peer parent income leads to an increase in the industry expenditure elasticity of around 1.8 points. In Columns 2 and 3, we estimate the impact separately for individuals who have high and low parent income. For those who have high-income parents, the estimate is larger and statistically significant at the 10% level, similar to the asymmetric effects found in the military service design. Finally, Column 4 reports the results for both entrepreneurs and employees, the peer effect is small and insignificant. This result suggests that peer effects are small for labor supply decisions.

V Implications for Growth and Inequality

In this section, we assess the implications of the empirical findings for growth and inequality through the lens of standard quantitative models. We first use a static framework to assess the distributional effects of unequal access to innovation, and then turn to a dynamic general equilibrium model.

V.A Static Model

We first develop a simple framework that can be used to quantify the distributional effects that may result from a more equal distribution of socio-economic backgrounds and gender among innovators. Throughout this section, we assume that marginal innovators exhibit the same average patterns found in Section III.

V.A.1 Quantitative Framework

We present calibrations that use the observed homophily patterns to assess the distributional effects of unequal access to innovation.

Consumer preferences and welfare effects of innovation. Assume consumers have CES preferences over a set of goods index by $k \in \Omega_t$, within a product module. The set of available goods Ω_t may vary over time, for instance as startups introduce new goods in the market. Utility of agent i is CES:

$$U_i = \left(\sum_{k \in \Omega_t} \omega_{k,i} q_{k,i}^{1-\sigma} \right)^{1/(1-\sigma)},$$

where σ is the elasticity of substitution between products within the product category, $q_{k,i}$ is the quantity of good k consumed by agent i, and $\omega_{k,i}$ is a taste parameter reflecting the intensity of i's preference for k.

Feenstra (1994) showed that with CES preferences, the welfare gain from the introduction of new goods (i.e., the set of available products Ω_t increases in the product category) can be expressed as a percentage of i's current income (equivalent variation for household i) as follows:

$$\pi_i = \frac{1}{1-\sigma} log \left(\frac{1 + GrowthSpendingContinuedGoods_i}{1 + GrowthTotalSpending_i} \right).$$

Assuming inelastic labor supply and taking the wage as the numeraire,[15] the growth of total spending is effectively normalized to zero, and the growth in spending on continued products is mechanically related to the share of spending on new goods S_i^N, with

$$GrowthSpendingContinuedGoods_i = -S_i^N.$$

With a first-order Taylor expansion around $S_i^N = 0$, the formula becomes:

$$\pi_i \approx \frac{S_i^N}{\sigma - 1}$$

For example, with $\sigma = 6$, a spending share on new goods of 2% is equivalent to a fall in inflation of $\frac{10}{5} = 2\%$ in welfare terms.

Innovators' backgrounds. We now consider the welfare impact of two startups that cater to different types of consumers. We consider a startup drawn from the baseline distribution of entrepreneur background ("Baseline"), which is skewed toward rich parents and male innovators, compared with a hypothetical equalized distribution ("Equal"), which could match the population gender ratio and the population distribution of parental income.

Distributional effects across consumer groups. Next, we consider two representative households, denoted "Type 1" and "Type 2." We derive the welfare comparison between these two households when transitioning from the "Baseline" to the "Equal" distribution of innovator background. We then bring the formulas to the data, computing the distributional effects between high- and low-income households, as well as between male and female customers.

As discussed in Deaton and Muellbauer (1980), CES preferences for a representative agent can be interpreted as the aggregation of discrete-choice logit preferences from a population of underlying agents. We assume that the startups drawn from different distributions of innovator backgrounds have similar elasticities of substitution σ, but differ in their preferences $\omega_{i,k}$, such that they may have different spending shares on the new goods introduced by different startups. S_1^N is the spending share of the "Type 1" representative agent from the bottom income decile on the

[15] Formally, we assume that there is only one wage rate in the economy. But different households can have different income and spending levels because they are endowed with different efficiency units of labor.

startup's products, while S_2^N corresponds to the spending share of the "Type 2" representative agent. Y_1 and Y_2 denote the total spending of the two household types.

Next, consider the entry of a new startup in the market. Each representative household buys products from this startup depending on its preferences, and the relative welfare gains are given by the following formula:

$$\frac{\pi_1}{\pi_2} \approx \frac{S_1^N}{S_2^N} = \frac{S_1^N \cdot Y_1}{S_2^N \cdot Y_2} \cdot \frac{Y_2}{Y_1} = \frac{R_1/R_2}{Y_1/Y_2}$$

where R_i denotes the total sales of the startup to representative household i. The ratio of sales to each of the representative agents is thus a sufficient statistic for the relative welfare effect, when appropriately normalized by the ratio of total spending of each of the agent. This result is intuitive: when agents have CES preferences with similar elasticities of substitution σ's, welfare differences can be reduced to differences in spending shares, and in turn to differences in the firm's revenue share from each agent, with a normalization for total purchasing power.

We wish to examine whether one of the household types benefits more from transitioning to a new distribution of innovator background. The unequal welfare effect across household types from the actual distribution of entrepreneur background ("Baseline") vs. a counterfactual distribution ("Equal") can be expressed as:

$$\Delta W \equiv \frac{\pi_1^{Equal}/\pi_1^{Baseline}}{\pi_2^{Equal}/\pi_2^{Baseline}} = \frac{R_1^{Equal}/R_2^{Equal}}{R_1^{Baseline}/R_2^{Baseline}}.$$

The relative welfare effect is thus governed by the share of sales to household groups of different types.

Linking the formula to regression specifications. We can link the formula for ΔW to our regression estimates from Tables 2 and 3, which document the revenue share of different types of startups for different types of households. Denoting by λ the share of sales to "Type 1" households, we can write $R_1/R_2 = \lambda/(1-\lambda)$. Tables 2 and 3 are directly informative about λ for startups with different innovator backgrounds.

V.A.2 Results

Using the formula from the previous section, we find that equalizing access to innovation could substantially reduce inequality and increase real income growth at the median of the income distribution. We also find that increase the share of female innovators would benefit female consumer substantially more. The results are reported in Table 3.

Denoting by λ the share of revenue from female-led households (which is the outcome in Table 2), we can write $\lambda = R_F/(R_M + R_F) \implies R_F/R_M = \lambda/(1-\lambda)$. Our previous results are directly informative about the expected λ for a female-founded startup ($\lambda^F = 0.2974$) compared to a male-founded startup ($\lambda^M = 0.25$). Therefore our estimates imply:

$$\Delta W = \frac{\lambda^F/(1-\lambda^F)}{\lambda^M/(1-\lambda^M)} = \frac{0.2974/(1-0.2974)}{0.25/(1-0.25)} \approx 1.2698$$

Thus, in the consumer packaged goods sample our preferred specification indicates that the welfare gains from female-founded startups are 27% larger for the representative female household, relative to the representative male household.[16] This number increases to 46.4% in the sample of phone applications.[17]

Taken together, these results indicate that the distributional effects from unequal access to innovation can be large. Next, we study these questions in a dynamic setting.

V.B Product Variety Growth Model

We now study the impact of unequal access to innovation for long-run growth and inequality using a simple product variety framework based on Romer (1990). We clarify the assumptions under which unequal access to innovation leads to both greater inequality and slower growth. While entry barriers do not affect long-term outcome with homogeneous consumers and innovators of equal research productivity, they do with heterogeneous research productivity and hetergeneous consumer tastes.

V.B.1 Single-Sector Model

To start, we lay out a simple product variety model of growth. We then derive results on long-run growth with homogeneous and heterogeneous research productivity.

Baseline model The basic model follows Romer (1990) and Acemoglu (2009). There is a representative household with preference[18]

[16] Note that this difference is substantially larger than the difference in revenue shares from female households that arises between female-founded and male founded startup. As shown in Column (4) of Table 2, the revenue share from female-led households is 18.96% larger for female-founded startups compared with male-founded startups (29.74% vs 25%). The welfare calculation from CES utility indicates that the comparison of revenue shares if biased downward. Intuitively, a downward bias arises because the revenue from female consumers appears in both the numerator and the denominator in the revenue share approach, while it appears only in the numerator of the welfare-relevant formula.

[17] To be applied to the setting of free phone applications, the quantitative framework presented above can be re-cast using a time constraint instead of a budget constraint.

[18] The log form can be generalized to a CRRA function. This just alters the Euler equation by the inverse of the risk aversion parameter: $\frac{\dot{C}(t)}{C(t)} = \frac{1}{\theta}(r(t) - \rho)$.

$$\int_0^\infty e^{-\rho t} log(C(t))dt,$$

where

$$C(t) = [\int_0^{N(t)} c(\nu, t)^{\frac{\epsilon-1}{\epsilon}} d\nu]^{\frac{\epsilon}{\epsilon-1}}.$$

The representative household maximizes lifetime discounted utility subject to the interest rate r. Entrepreneurs invent new varieties, receive perpetual patents, and produce the variety using a linear production function:

$$y(\nu, t) = l(\nu, t).$$

The research production function takes the form

$$\dot{N}(t) = \eta N(t) L_R(t),$$

where η is the research productivity of all individuals and L_R is the amount of labor allocated to research rather than production. All labor must be allocated to production or research, such that the labor market clears:

$$\int_0^{N(t)} l(\nu, t) d\nu + L_R(t) \leq L.$$

Throughout, we assume that there is a single market-clearing wage $w(t)$. This will be especially important when heterogeneity is added in the model.

In this model, there exists a unique balanced growth path where the growth rate, the wage, the interest rate, and the labor allocated to research are all constant.Along the balanced growth path, the equilibrium growth rate is $g^* = \frac{1}{2}(\frac{\eta}{\epsilon-1}L - \rho)$ and the population working in research is $L_R^* = \frac{L}{2} - \frac{\epsilon-1}{2\eta}\rho$. The derivations are presented in Appendix A.

Adding Barriers to Innovation Now, we add barriers to innovation for a subset of agents. This can be modeled in two ways. One way is to have a "real" barrier that directly impacts research productivity. Individuals facing barriers will be $1 - \tau$ times as productive, which could reflect lack of opportunity to develop skills or lack of access to adequate funding. A second way is to have a "preference" barrier, which only distorts the decision to enter into research, but does not impact research productivity conditional on entry. This could capture factors such as discrimination and preferences.

In both cases, the indifference equation becomes

$$(1-\tau)\eta N(t)V(t) = w(t)$$

and the only difference between the two cases is the law of motion governing product variety growth.

Proposition 1. *In an expanding product variety model with access barriers, if the size of the unrestricted group is greater than L_R^*, there exists a BGP where aggregate consumption expenditure grows at $g^* = \frac{1}{2}(\frac{\eta}{\epsilon-1}L - \rho)$.*

Adding the wedge does not violate any equilibrium conditions in the baseline model. Only people from the unrestricted group participate in research production, because the equilibrium wage rate is higher than the effective returns to entrepreneurship for the restricted group. As long as the unrestricted group is larger than L_R^*, there is no impact on growth.[19] Intuitively, unrestricted individuals can fill in for individuals who are restricted and still produce varieties at the same rate. Thus, barriers to innovation do not matter for growth if there is homogeneity in preferences and research productivity.

Heterogeneous Research Productivity and Barriers Next, we show that entry barriers do matter for long-run growth if there is heterogeneous research productivity. To obtain closed-form solution, we consider a uniform distribution of ability.[20]

In the basic case, we can solve for growth rates in terms of distributional parameters.[21] We assume that research productivity is equal to $\eta\kappa$, where $\kappa \sim U[0,1]$. In equilibrium there is a cutoff $\bar{\kappa}$ below which the agents work in production. This leads to the following varieties growth equation:

$$\frac{\dot{N}(t)}{N(t)} = \eta L \int_{\bar{\kappa}}^{1} \kappa f(\kappa) d\kappa = \frac{1-\bar{\kappa}^2}{2}\eta L$$

The marginal agent is indifferent between production wages and the returns to entrepreneurship:

[19] Note that the assumption that there is a single labor market clearing wage is important for this result. This prevents firms from offering lower wages to the restricted group, knowing that they do not want to pursue entrepreneurship.

[20] The solution can be easily generalized to other distributions, including a Pareto distribution. However, in the case of the Pareto distribution, the model doesn't generally have a closed form solution. In that case, the equation governing the number of entrepreneurs is a polynomial with order that depends on the Pareto parameter. We present calibrations in Appendix B.

[21] Details are shown in Appendix A.B

$$\bar{\kappa}\eta N(t)V(t) = w(t),$$

and the inframarginal entrepreneurs obtain rents above market wage. Solving the model, we obtain $\bar{\kappa} = \sqrt{\frac{1}{3} + \frac{2\rho(\epsilon-1)}{3\eta L}}$ and $g^* = \frac{1}{3}\left(\frac{\eta L}{\epsilon-1} - \rho\right)$.[22]

Next, we add entry barriers. We focus here on the case of "preference" barriers.[23] Assume that there are two groups of size $\frac{L}{2}$, and group 2, the "restricted" group, faces barriers to entrepreneurship. Both groups have the same underlying ability distributions. Now, there will be two cutoffs $\bar{\kappa}_1, \bar{\kappa}_2$ that are governed by the following indifference equations:

$$\bar{\kappa}_1\eta N(t)V(t) = w(t),$$

$$(1-\tau)\bar{\kappa}_2\eta N(t)V(t) = w(t),$$

which implies that $\bar{\kappa}_1 = (1-\tau)\bar{\kappa}_2$. In other words, the marginal agent in the restricted group has higher research productivity. This leads to our second proposition.

Proposition 2. *In an expanding product variety model with uniformly distributed research productivity, $\eta\kappa, \kappa \sim U[0,1]$, and a random half of the population facing barriers to entrepreneurship, the aggregate consumption expenditure along the unique BGP grows at*

$$g^* = \frac{1}{\epsilon-1}(2-\tau)(1-\tau)\cdot\frac{\eta L + 2\rho(\epsilon-1)}{(3\tau^2 - 8\tau + 6)} - \rho$$

Greater τ reduces growth rate, increases the number of entrepreneurs from the unrestricted group, decreases the number of entrepreneurs from the restricted group, and increases relative wages across the groups.

The derivations are presented in Appendix A.B. g^* is equal to the solution without access barriers at $\tau = 0$ and is decreasing in τ on the interval from 0 to 1. Therefore, the entry barriers for one group reduce long-run growth. In addition, it also creates differences in average income across the two groups, hence inequality, because there are fewer inframarginal entrepreneurs in the restricted group.

[22]Note that this is lower than the growth rate in the baseline model with homogeneous research productivity, but this mechanically results from the fact that now the average productivity in the population is $\frac{\eta}{2}$. The productivity parameter can be normalized such that the growth rate remains unchanged.

[23]We can also solve the case where the barrier affects productivity. The solution is similar to the solution to the basic heterogeneous research productivity model above, but with a step function representing the PDF of the ability distribution.

V.B.2 Two-Sector Model

Next, we add a second sector to the model in order to incorporate differences in tastes across consumer groups.[24] Using the model, we investigate the impact of unequal access and exposure on long-run growth and inequality in consumer welfare.

Baseline model with homogeneous consumers We first derive the BGP with two sectors and homogeneous consumers. The representative consumer now has Cobb-Douglas utility over two sectors:

$$C(t) = C_1(t)^{\alpha} C_2(t)^{1-\alpha}$$

$$C_i(t) = [\int_0^{N_i(t)} c_i(\nu, t)^{\frac{\epsilon-1}{\epsilon}} d\nu]^{\frac{\epsilon}{\epsilon-1}}$$

Given the Cobb-Douglas form, $C_1(t) = \alpha C(t), P_2 C_2(t) = (1-\alpha)C(t)$. Sector 1 prices are the numeraire and the implied relative price $P_2 = \left(\frac{N_2}{N_1}\right)^{\frac{1}{1-\epsilon}}$. Let $N(t) = N_1(t) + N_2(t)$. The innovation productivity in each sector i is:[25]

$$\dot{N}_i(t) = \eta_i N(t) L_{iR}(t)$$

Let L_{iM}, L_{iR} be the production and research labor allocated to sector i, respectively. The labor market clearing is now

$$L = L_{1R} + L_{2R} + L_{1M} + L_{2M}$$

Along a BGP, the growth rate of varieties in both sectors will be constant, along with the amount of labor allocated to each sector, and to research and production within each sector. We find that, along a balanced growth path, $\frac{L_{1M}}{L_{2M}} = \frac{L_{1R}}{L_{2R}} = \frac{\alpha}{1-\alpha}$. The research allocation and growth rate are given by:

$$L_{1R}^* = \frac{\alpha}{2}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)\eta_2)}\right), L_{2R}^* = \frac{1-\alpha}{2}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)\eta_2)}\right),$$

[24] This can easily be generalized to multiple sectors, which can help us incorporate issues surrounding mass market goods vs. specialized goods.

[25] If sectors had different productivities and varieties production only depended on varieties in the same sector, then there would be explosive growth in one sector and the equilibrium would not admit the existence of a balanced growth path.

$$g^* = \frac{1}{2}\left(\frac{1}{\epsilon - 1}\left(\alpha\eta_1 + (1-\alpha)\eta_2\right) L - \rho\right). \tag{6}$$

The formula reduces to the growth rate in the baseline model when $\alpha = 0$ or $\alpha = 1$.

Taste heterogeneity We now examine the impact of taste heterogeneity across consumers. Let there be two groups of consumers. Group 1 has size $1-\delta$ and preference parameter α. Group 2 has size δ and preference parameter α'. Let $\alpha > \alpha'$, i.e. Group 2 has a stronger preference for sector 2 compared with Group 1. Let $\delta < \frac{1}{2}$, i.e. Group 2 is the minority group.

In this model, the equilibrium growth rate is given by (6), using $\tilde{\alpha} = (1-\delta)\alpha + \delta\alpha'$ as the effective taste parameter. However, in this case, the model features welfare inequality across consumers. Welfare inequality persists at a constant ratio along the BGP. Starting from the primitives, we can derive the ratio of consumption across the two groups:

$$\begin{aligned}
\frac{C^1}{C^2} &= \frac{[L_{1M}N_1(t)^{\frac{1}{\epsilon-1}}]^{\alpha}[L_{2M}N_2(t)^{\frac{1}{\epsilon-1}}]^{1-\alpha}}{[L_{1M}N_1(t)^{\frac{1}{\epsilon-1}}]^{\alpha'}[L_{2M}N_2(t)^{\frac{1}{\epsilon-1}}]^{1-\alpha'}} \\
&= \left(\frac{L_{1M}}{L_{2M}}\right)^{\alpha-\alpha'}\left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}} \\
&= \left(\frac{(1-\delta)\alpha + \delta\alpha'}{(1-\delta)(1-\alpha) + \delta(1-\alpha')}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta_1}{\eta_2}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}} \qquad (7)
\end{aligned}$$

This expression shows that welfare inequality is determined by two terms, corresponding to exogenous determinants of technology and endogenous market size effects. The second term depends on the ratio of technological parameters, $\frac{\eta_1}{\eta_2}$, i.e. consumer groups who prefer the sector with higher research productivity have relatively higher welfare along the BGP. The ratio in the first term corresponds to the relative sales of the two sectors, $\frac{\tilde{\alpha}}{1-\tilde{\alpha}}$; it shows that welfare is higher for consumer groups who prefer the sector with a higher market size. Intuitively, the larger sector endogenously gets more varieties along the BGP. The market size imbalance is weaker when the minority consumer group is larger.

These results are summarized in our third proposition:

Proposition 3. *In a two-sector model with heterogeneous consumer preferences and heterogeneous sectoral research productivites, inequality increases in taste heterogeneity ($\alpha - \alpha'$) and decreases in the size of the minority consumer group δ.*

Motivated by the evidence in previous sections that innovators are more likely to innovate for consumers like themselves, we add innovator specialization to the model.

Innovator specialization and entry barriers Assume that agents in Group 1 have research productivity η when working in sector 1 and $\eta' < \eta$ when working in sector 2. The opposite holds for agents in Group 2. For tractability, we assume that there are enough agents in each group relative to the equilibrium allocations in the baseline two-sector model, i.e, $(1-\delta)L > L^*_{1R}$ and $\delta L > L^*_{2R}$, where L^*_{1R}, L^*_{2R} are defined in Equation 6. In this case, innovators from each group specialize and enter a single sector. Along the BGP, the equilibrium growth rate is $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}\eta L - \rho\right)$, and inequality across consumer groups is $\frac{C^1}{C^2} = \left(\frac{(1-\delta)\alpha+\delta\alpha'}{(1-\delta)(1-\alpha)+\delta(1-\alpha')}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}}$.[26]

Next, we add entry barriers for agents from Group 2, i.e. there is a wedge $(1-\tau)$ governing their choice between entrepreneurship and production work. Again, we present the case of "preference" barriers, but results also hold for "real" barriers. Using this setup, we derive our fourth proposition:

Proposition 4. *In a two-sector model with preference heterogeneity and innovator specialization, the presence of barriers to innovation creates greater inequality and lowers long-run growth relative to the case with no barriers under either of the following conditions:*

1. $\eta' > (1-\tau)\eta$ and $\delta L < \frac{\rho(\epsilon-1)}{(\alpha\eta+(1-\alpha)\eta')}$
2. $\eta' < (1-\tau)\eta$ and $(1-\delta)L > \frac{\alpha}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1+(1-\alpha)(1-\tau)\eta_2)}\right)$ and $\delta L > \frac{(1-\alpha)(1-\tau)}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1+(1-\alpha)(1-\tau)\eta_2)}\right)$

To derive Proposition 4, two cases must be consider. In the first case, $\eta' > (1-\tau)\eta$. Under this scenario, individuals from Group 1 will perform all research in the economy, as long as there are enough of them. Individuals from Group 2 do not receive enough returns from entrepreneurship to enter. Under this scenario, $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}\left(\tilde{\alpha}\eta + (1-\tilde{\alpha})\eta'\right)L - \rho\right)$ and $\frac{C^1}{C^2} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta}{\eta'}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$. The growth rate is smaller than the one without entry barriers, and the inequality is higher.

The second case to consider features $\eta' < (1-\tau)\eta$. Under this scenario, Group 2 individuals do enter entrepreneurship, but fewer do so than in a world without entry barriers. Intuitively, the outcome is in-between a world where their research productivity is $(1-\tau)\eta$ and a world where there

[26] Note that the expression for inequality is identical to the first term of the expression for inequality discussed in Proposition 3. Intuitively, because we assumed that research productivities were symmetric and that inventors specialized fully, steady-state inequality now depends purely on market size effects.

are no taxes, because they are still mechanically productive, but behave according to an indifference equation with a wedge.

Appendix A.D provides a derivation of a general two-sector model with a "preference" barrier affecting entry into research in one sector and equal research productivities for people working in the two sectors. The equilibrium growth rate and inequality under this scenario is

$$g^* = \frac{1}{2-(1-\tilde{\alpha})\tau}\left(\frac{1}{\epsilon-1}[1-(1-\tilde{\alpha})\tau]\eta L - \rho\right) \tag{8}$$

$$\frac{C^1}{C^2} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{1}{(1-\tau)}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$$

In this case, an increase in entry barriers leads to a lower equilibrium growth rate and an increase in inequality along the balanced growth path. The effect on inequality is greater when taste differences are larger.

Finally, we note that Proposition 4 also holds for a "real" barrier. If $\eta' > (1-\tau)\eta$, then individual from Group 1 will perform all research again. If $\eta' < (1-\tau)\eta$, then Group 2 will enter entrepreneurship in their own sector. This case will be equivalent to the two-sector model where research productivity in sector 1 is η and research productivity in sector 2 is $\tilde{\eta} = (1-\tau)\eta$. The growth rate, based on Equation 6, will be decreasing in τ, and inequality, based on Equation 7, will be increasing.

The role of interactions with peers Motivated by the quasi-experimental evidence on the impact of peers on the direction of entrepreneurship, we introduce a new parameter to parsimoniously model sociological interactions. Starting from the same natural research productivities η and η' defined in the model above, we introduce the parameter ψ, which governs the intensity of interaction with individuals from the other group. We assume that when an agent interacts with another agent with a higher productivity, they will learn from them and increase their own productivity in innovating for the other group. Through these interactions, research productivity in the sector for which agents have a comparative disadvantage is given by $\eta'' = \psi\eta + (1-\psi)\eta'$. Thus η'' depends on the degree of interactions and gives the effective productivity in the sector where an agent has a comparative disadvantage.

Proposition 5. *In a two-sector model with preference heterogeneity, innovator specialization, and real barriers, higher ψ leads to a weakly higher growth rate and weakly lower inequality along the* *balanced growth path.*

To derive Proposition 5, several cases must be considered. Start with the case where $\eta' < (1-\tau)\eta$. Absent exposure, individuals specialize in research. As we increase the exposure parameter, η'' increases, which at first has no effect on outcomes. For a high enough ψ, $\eta'' > (1-\tau)\eta$. At this point, Group 1 individuals will enter entrepreneurship in sector 2, while Group 2 individuals drop out of entrepreneurship. As exposure grows further, the growth rate $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}\left(\tilde{\alpha}\eta + (1-\tilde{\alpha})\eta''\right)L - \rho\right)$ will increase and the inequality across the two groups will decrease $\frac{C^{1*}}{C^{2*}} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta}{\eta''}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$.

In the case of a "preference" barrier, Proposition 5 holds in the region where $\eta'' > (1-\tau)\eta$. At the point where $\eta'' = (1-\tau)\eta$, Group 1 individuals replace Group 2 individuals in innovating, causing a drop in long-run growth rate, because Group 1 individuals are less productive at innovating in Sector 2. Eventually, as ψ rises further, the growth rate will rise above the growth rates under specialization.

V.B.3 Discussion

Building on a standard endogenous growth model, we have shown formally that greater ability to innovate for one's own group implies that entry barriers lead to lower long-term growth and greater inequality across consumer groups. The models presented in this section have been simplified to maintain analytical tractability. In particular, the two-sector model does not address potential heterogeneity in research productivity within population groups. However, the model still yields a series of insights.

First, the models presented here can be applied to a variety of settings. In this article, both our micro and macro empirical findings are based on consumer-facing industries, where there is likely to be significant taste heterogeneity. The impact of access barriers on consumer inequality will matter less in industries facing less taste heterogeneity, such as those producing basic inputs sold to other firms. However, as highlighted in Proposition 2, even in a case without taste heterogeneity entry barriers can still lead to income inequality across groups and lower growth rates.

Second, the pool of innovators and entrepreneurs has a very low fractions of women, underrepresented minorities, and individuals from low-income families. This suggests that the barriers to entrepreneurship, τ, is likely to be substantial. Marginal changes in τ may not have significant impacts on growth rates, as people from these groups would have to be much more productive than their unrestricted peers.

Third, the models can rationalize the asymmetric exposure results discussed in Section IV. As noted in Proposition 5, increases in exposure can lead the unrestricted group to enter research in their less-preferred sector (while the restricted group would drop out). Even with heterogeneous ability within groups, if individuals from the restricted group become more able to innovate in the other sector, they are unlikely to do so in equilibrium, because they still face an additional wedge.

One element missing from the two-sector model is the potential for access barriers to create income inequality. In a heterogeneous ability model, inframarginal entrepreneurs will earn incomes above the production wage. In this case, the wedge will also distort the income distribution across the groups. Under a two-sector model, this would then give greater weight to the preferences of the unrestricted group, creating a feedback loop that further increases inequality across the groups.

VI Conclusion

In this article, we have documented novel empirical facts on the relationship between innovators and the consumers they serve. We find significant evidence that innovators are more likely to work on goods used by consumers similar to them along observable dimensions. A simple quantitative framework suggests that unequal access to innovation leads to large distributional effects across consumer groups. In addition, we provide causal evidence that being exposed to peers from a different socio-economic background has an impact of the type of venture an entrepreneur starts, in particular regarding which customers to target. These findings show that personal and social backgrounds have a causal impact on the direction of innovation.

In addition, we investigate the implications of our empirical findings for long-run growth and inequality. In a two-sector model with greater ability to innovate for one's own group, unequal access leads to both slower long-run growth and greater welfare differences across groups. The model illustrates that greater social interaction across groups could act as a second-best solution to increasing growth and reducing inequality in the presence of barriers to entrepreneurship.

These findings speak to policies and initiatives aiming at promoting access to entrepreneurship for women and minorities. For example, the Small Business Act for the Small Business Innovation Research (SBIR) proposes to "foster and encourage" the participation of women, minorities, and businesses in underrepresented areas. Another recent piece of legislation, the SUCCESS Act, encourages the USPTO to measure and encourage the participation of women, minorities, and veterans in invention. Beyond explicit policies, there has also been increased hiring of female partners in venture capital and initiatives in the tech industry to encourage more women and underrepresented minorities to participate in engineering. Our results suggest that these initiatives are likely to lead to a more diverse set of new goods and services, and simultaneously increase growth and reduce inequality.

References

Acemoglu, Daron, "Directed technical change," *The Review of Economic Studies*, 2002, *69* (4), 781–809.

_ , "Equilibrium bias of technology," *Econometrica*, 2007, *75* (5), 1371–1409.

_ , *Introduction to Modern Economic Growth* 2009.

_ , Philippe Aghion, Leonardo Bursztyn, and David Hemous, "The environment and directed technical change," *American economic review*, 2012, *102* (1), 131–66.

Agarwal, Ruchir and Patrick Gaulé, "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?," *American Economic Review: Insights*, 2020.

Aghion, Philippe, Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John van Reenen, "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry," *Journal of Political Economy*, 2016, *124* (1), 1–51.

_ , Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen, "The Social Origins of Inventors," 2017.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi, "Dancing with the Stars: Innovation Through Interactions," 2018.

Angrist, Joshua D., "The perils of peer effects," *Labour Economics*, 2014, *30*, 98–108.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," 2017.

Berry, Steven, Alon Eizenberg, and Joel Waldfogel, "Fixed Costs and the Product Market Treatment of Preference Minorities," *Journal of Industrial Economics*, 2016, *64* (3), 466–493.

Bönte, Werner and Monika Piegeler, "Gender gap in latent and nascent entrepreneurship: Driven by competitiveness," *Small Business Economics*, 2013, *41* (4), 961–987.

Borusyak, Kirill and Xavier Jaravel, "The Distributional Effects of Trade: Theory and Evidence from the United States," *SSRN Electronic Journal*, 2018.

Brooks, Alison Wood, Laura Huang, Sarah Wood Kearney, and Fiona E Murray, "Investors prefer entrepreneurial ventures pitched by attractive men," *PNAS*, 2014, *111* (12), 4427–4431.

Caliendo, Marco, Frank M. Fossen, Alexander Kritikos, and Miriam Wetter, "The gender gap in entrepreneurship: Not just a matter of personality," *CESifo Economic Studies*, 2015, *61* (1), 202–238.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka, "The long-run effects of disruptive peers," *American Economic Review*, 2018, *108* (11), 3377–3415.

Cook, Lisa D and Chaleampong Kongcharoen, "The Idea Gap in Pink and Black," 2010.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams, "The more we die, the more we sell? A simple test of the home-market effect," *Quarterly Journal of Economics*, 2019, *134* (2), 843–894.

Criado Perez, Caroline, *Invisible Women: Data Bias in a World Designed for Men* 2019.

Deaton, Angus and John Muellbauer, "An Almost Ideal Demand System," *The American Economic Review*, 1980, *70* (3), 312–326.

Duflo, Esther and Raghabendra Chattopadhydy, "Women as Policy Makers : Evidence from a Randomized Policy Experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

Dunn, Thomas and Douglas Holtz-Eakin, "Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links," *Journal of Labor Economics*, 2000, *18* (2), 282–305.

Einiö, Elias, "Mixing the Rich and Poor: The Impact of Peers on Education and Earnings," 2019.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157–177.

Fernandez, Raquel and Richard Rogerson, "Sorting and long-run inequality," *Quarterly Journal of Economics*, 2001, *116* (4), 1305–1341.

Foellmi, Reto and Josef Zweimüller, "Income Distribution and Demand-induced Innovations Income Distribution and Demand-induced Innovations," *Review of Economic Studies*, 2006, *73* (212), 941–960.

Guzman, Jorge and Aleksandra Kacperczyk, "Gender Gap in Entrepreneurship," 2018.

_ **and Scott Stern**, "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014," 2016.

_ **, Jean Joohyun Oh, and Ananya Sen**, "What Motivates Innovative Entrepreneurs ? Evidence from Three Field Experiments," 2019.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin, "Does peer ability affect student achievement?," *Journal of Applied Econometrics*, 2003, *18* (5), 527–544.

Hayek, Friedrich August, "The use of knowledge in society," *The American economic review*, 1945, *35* (4), 519–530.

Hoxby, Caroline M., "Peer Effects in the Classroom: Learning From Gender and Race Variation," 2000.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," 2019.

Hvide, Hans K. and Paul Oyer, "Dinner Table Human Capital and Entrepreneurship," 2019.

Jacob, Schmookler, "Invention and economic growth," *J. Schmookler.–Cambridge Mass: Harvard University Press*, 1966.

Kanze, Dana, Laura Huang, Mark A. Conley, and E. Tory Higgins, "We ask men to win and women not to lose: Closing the gender gap in startup funding," *Academy of Management Journal*, 2018, *61* (2), 586–614.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Female Inventors and Inventions," *SSRN Electronic Journal*, 2019, (799330), 1–43.

Kremer, Michael, "How much does sorting increase inequality?," *Quarterly Journal of Economics*, 1997, *112* (1), 115–139.

Lucas, Robert E. and Benjamin Moll, "Knowledge growth and the allocation of time," *Journal of Political Economy*, 2014, *122* (1), 1–51.

Lybbert, Travis J. and Nikolas J. Zolas, "Getting patents and economic data to speak to each other: An 'Algorithmic Links with Probabilities' approach for joint analyses of patenting and economic activity," *Research Policy*, 2014, *43* (3), 530–542.

Malmström, M, J Johansson, and J Wincent, "We Recorded VCs' Conversations and Analyzed How DifferentlyThey Talk About Female Entrepreneurs," *Harvard Business Review*, 2017, pp. 1–6.

McDevitt, Ryan C. and James W. Roberts, "Market structure and gender disparity in health care: Preferences, competition, and quality of care," *RAND Journal of Economics*, 2014, *45* (1), 116–139.

Mishkin, Elizabeth, "Gender and Sibling Dynamics in the Intergenerational Transmission of Entrepreneurship," *Working Paper*, 2017.

Rivera-Batiz, Luis A and Paul M Romer, "Economic Integration and Endogenous Growth," 1991, *106* (2), 531–555.

Romer, Paul M., "Endogenous technological change," *Journal of Political Economy*, 1990, *98* (5), S71–S102.

Shane, Scott, "Prior knowledge and the discovery of entrepreneurial opportunities," *Organization science*, 2000, *11* (4), 448–469.

Thébaud, Sarah, "Gender and entrepreneurship as a career choice: Do self-assessments of ability matter?," *Social Psychology Quarterly*, 2010, *73* (3), 288–304.

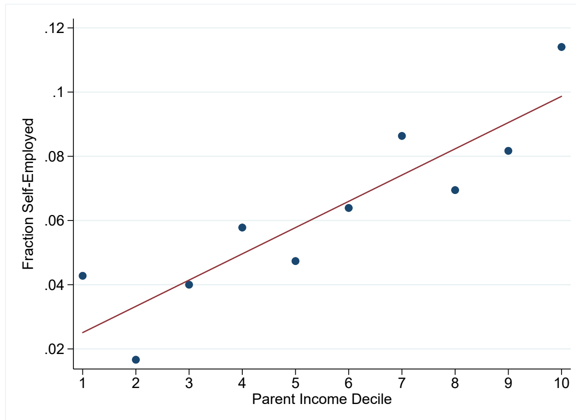
Toole, Andrew A., Stefano Breschi, Ernest Miguelez, Amanda Myers, Edoardo Fer-rucci, Valerio Sterzi, Charles A.W. DeGrazia, Francesco Lissoni, and Gianluca Taras-coni, "Progress and Potential: A Profile of Women Inventors on U.S. Patents," 2019.

Von Hippel, Eric, "Lead users - a source of novel product concepts," *Management Science*, 1986, *32* (7), 791–806.

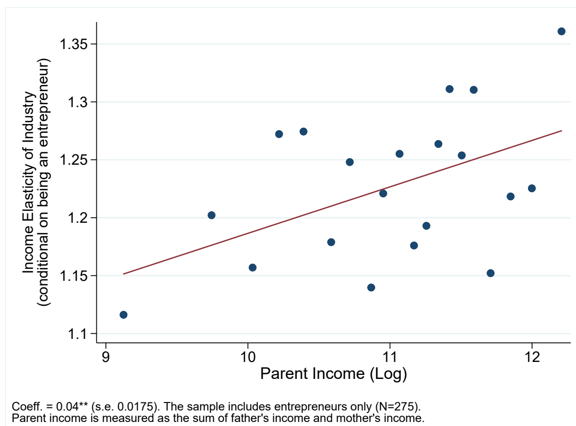
Waldfogel, Joel, "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *The RAND Journal of Economics*, 2003, *34* (3), 557.

Figure 1: Parent Income and the Direction of Entrepreneurship in the U.S.

Panel A: Probability of Self-Employment by Parent Income



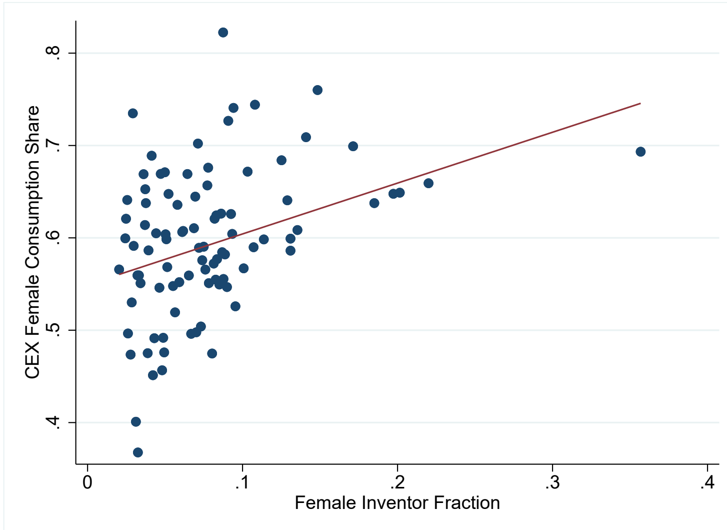
Panel B: Income Elasticities and Parent Income



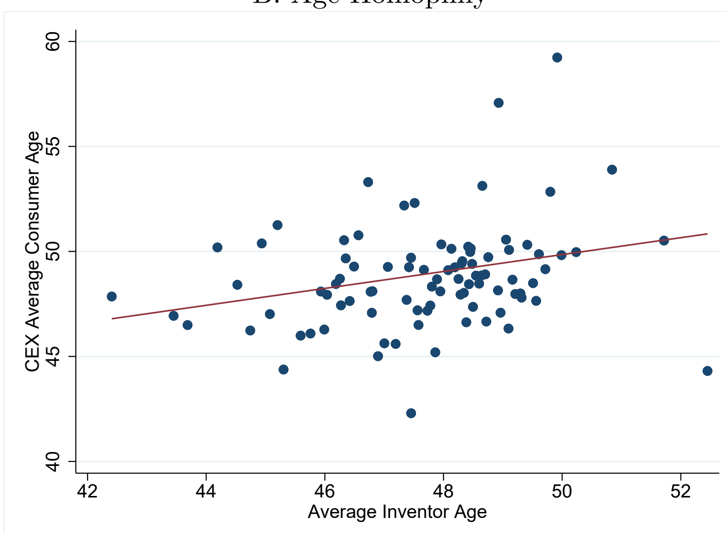
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 2: Gender and Age Homophily between Patent Inventors and Consumers across Industries

A. Gender Homophily



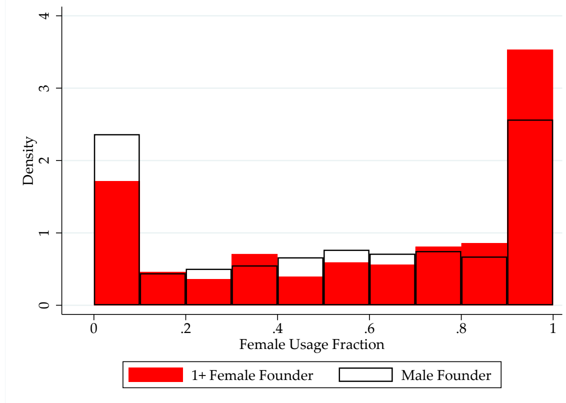
B. Age Homophily



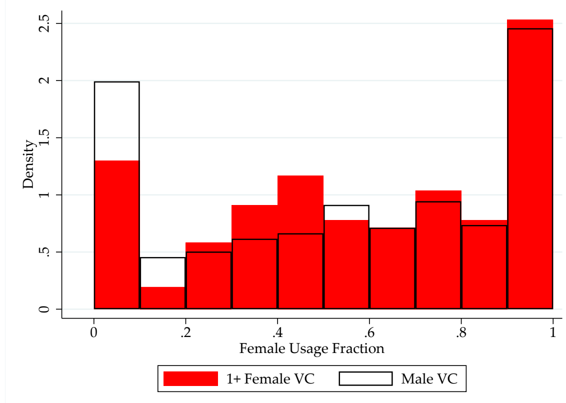
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 3: User-Innovator Homophily within Phone Applications

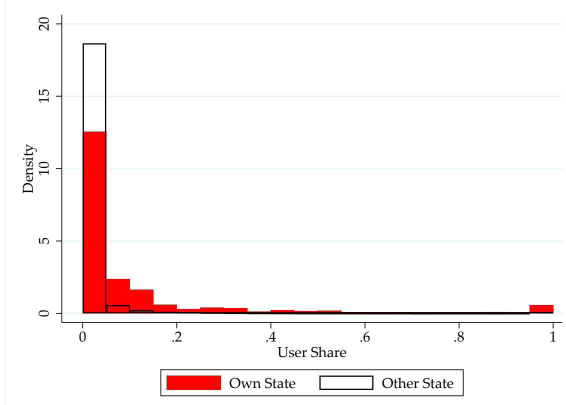
A. Female User Fraction vs. Founder Gender



B. Female User Fraction vs. VC Gender



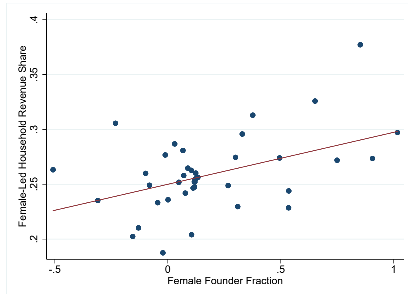
C. Home State Sales Share



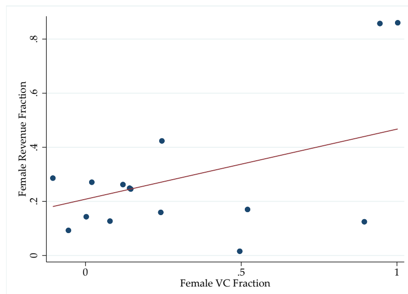
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 4: Consumer-Innovator Homophily within Consumer Packaged Goods

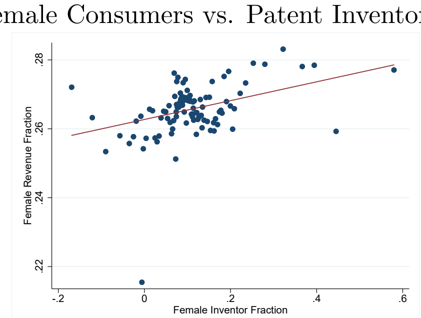
A. Sales Share to Female Consumers vs. Founder Gender (Startups)



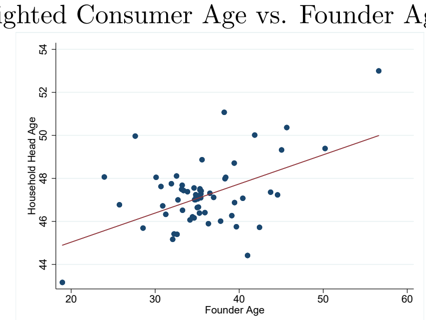
B. Sales Share to Female Consumers vs. VC Partner Gender (Startups)



C. Sales Share to Female Consumers vs. Patent Inventor Gender (All Firms)



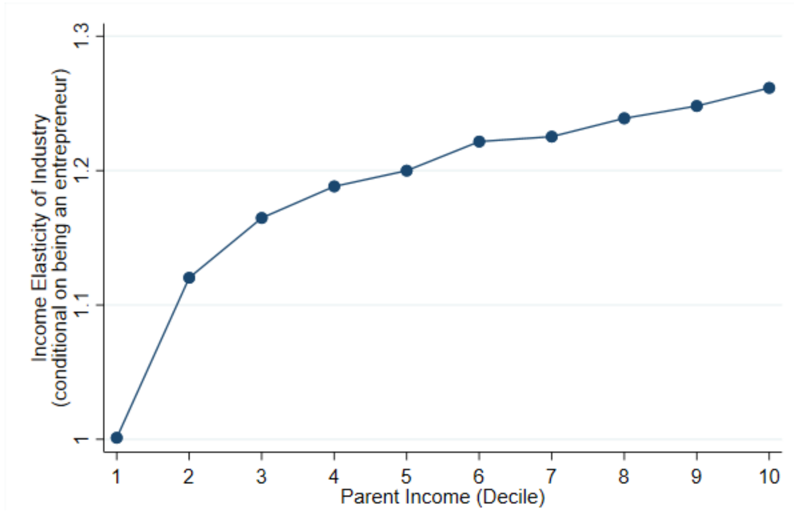
D. Sales-weighted Consumer Age vs. Founder Age (Startups)



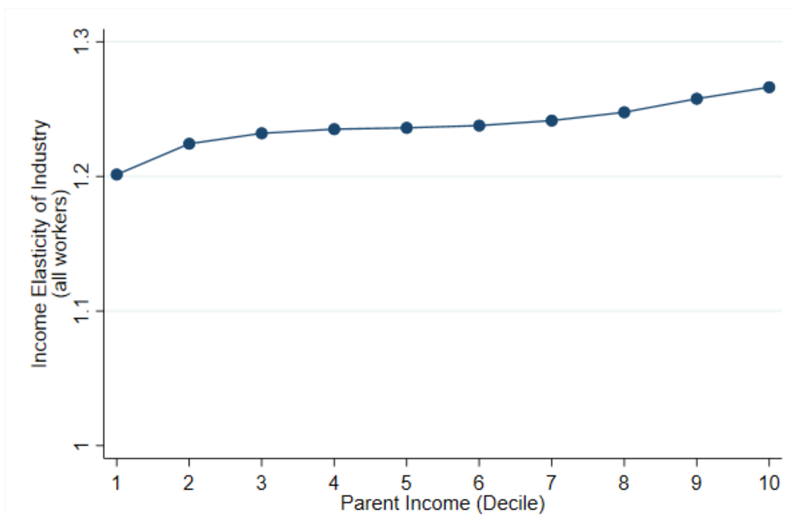
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 5: Homophily Patterns for Entrepreneurship in Finland

A. Entrepreneurs



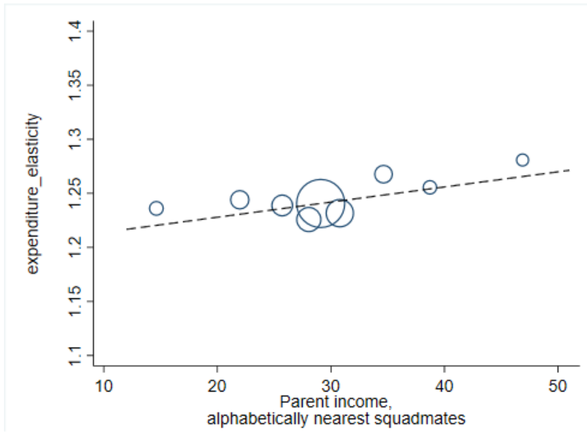
B. All



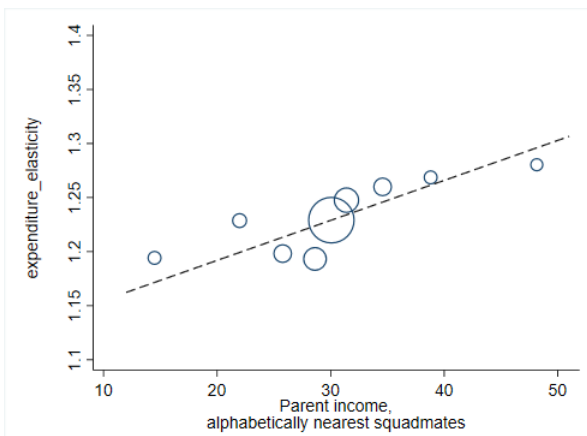
Notes: In panel A, the coefficient (standard error) from an individual-level regression of industry income elasticity on log of parent income is 0.1316 (0.0037). In panel B, the corresponding coefficient (standard error) is 0.0287 (0.0008). The number of observations is 299,810 in panel A and 4,762,964 in panel B. Data sources: Finnish FLEED and FOLK datasets, years 2010-2016.

Figure 6: The Causal Effect of Peers on the Direction of Entrepreneurship: Peer Parent Income vs. Entrepreneur's Industry Income Elasticity

A. Entrepreneurs and managers

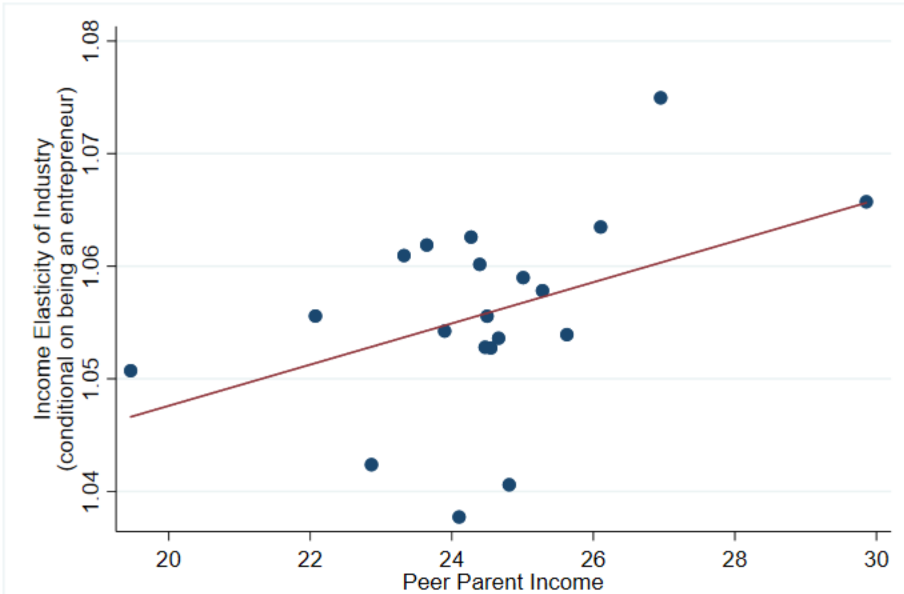


B. Entrepreneurs and managers with own parent income in the highest tercile



Notes: The figure displays the reduced-form relationship between parent income of two alphabetically nearest squadmates and industry expenditure elasticity in a sample including entrepreneurs and managers (panel A) and a sample restricted to entrepreneurs and managers in the highest own parent income tercile (panel B). The figure plots the residuals from separate regressions of industry income elasticity and squadmate parent income on own parent income, squadron fixed effects, and year dummies. The size of the circle represents the number of conscripts within each bin. Data sources: Finnish FLEED and FOLK datasets and FDF conscript registry.

Figure 7: Study peer parent income and industry expenditure elasticity



Notes: The figure displays the relationship between parent income of study peers and industry expenditure elasticity in a sample including entrepreneurs. The figure plots the residuals from separate regressions of industry income elasticity and study peer parent income on own parent income, school-by-program fixed effects, school-by-start-year fixed effects, and dummies for gender, age, and year of outcome measurement. Data sources: Statistics Finland FLEED and FOLK datasets.

Table 1: Summary Statistics

	Phone Applications	Consumer Packaged Goods
# VC-Backed Startups	1,679	158
Female Founders ≥ 1	0.14	0.24
Female VC Partner ≥ 1	0.06	0.04
# New Products	3,380	4,058
# Categories/Subcategories	14/54	90/294
# Panelists	50,725	168,775
Timeframe	2017-2019	2004-2016

Notes: See Section 1 for a description of the datasets.

Table 2: Innovator-Consumer Homophily by Gender and Age across Industries

	Specification		OLS Estimates		
#	Patent Inventor Char. (LHS)	Consumer-based Outcome (RHS)	Coeff.	Constant	Sample
1.	Fraction of Female Inventors	Share of Sales to Households with Female Head	0.551** (0.105)	0.549** (0.012)	All industries, $N = 325$
2.	Fraction of Female Inventors	Share of Sales to Households with Female Head	0.100** (0.048)	0.549*** (0.012)	Compustat, $N = 206$
3.	Average Age of Patent Inventors	Average Consumer Age, weighted by Sales	0.402*** (0.147)	29.73** (7.002)	All industries, $N = 325$
4.	Average Age of Patent Inventors	Average Consumer Age, weighted by Sales	0.007 (0.014)	47.41*** (0.816)	Compustat, $N = 406$

Notes: See Section 2 for a description of the data and Section 3 for the methodology. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Innovator-Consumer Homophily within Industries

A. Within Phone Applications

Specification			OLS Estimates			
#	Founder Char. (LHS)	User-based Outcome (RHS)	Baseline	With Category F.E.	Mean of Dep. Var.	Sample
1.	Female	Share of Time Use by Female Users	0.092*** (0.034)	0.082** (0.032)	0.542	All apps, $N = 3,380$
2.	Home State	Share of Time Use in Founder Home State	0.086*** (0.007)	0.041*** (0.008)	0.02	All apps, $N = 3,380$

B. Within Consumer Packaged Goods

Specification			OLS Estimates			
#	Innovator Char. (LHS)	Consumer-based Outcome (RHS)	Baseline	With Category F.E.	Mean of Dep. Var.	Sample
1.	Female Founder	Share of Sales to HHs with Female Head	0.030 (0.024)	0.047** (0.021)	0.25	New Goods, $N = 4,058$
2.	Age of Founder	Average Consumer Age, weighted by Sales	0.151*** (0.047)	0.135** (0.052)	47.2	New Goods, $N = 4,058$
3.	Fraction of Female Patent Inventors	Share of Sales to HHs with Female Head	0.039** (0.019)	0.027* (0.015)	0.265	All goods, $N = 1,094,229$

Notes: See Section 2 for a description of the data and Section 3 for the methodology. For the consumer packaged goods sample, the analysis is conducted at the product level, and standard errors are clustered by firm. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: The Effects of Peer Parent Income on Industry Expenditure Elasticity and Sales Share to High-Income Consumers

Dependent variable	First stage	Reduced form	IV	Dependent mean	N
	A. Entrepreneurs and managers				
Expenditure elasticity	0.3358***	0.0014***	0.0043***	1.24	9755
	(0.0377)	(0.0004)	(0.0011)		
Rich share (60k)	0.3358***	(0.0003)	(0.0008)	0.63	9755
	(0.0377)	(0.0002)	(0.0008)		
Rich share (100k-30k)	0.3358***	0.0004	0.0011	0.70	9755
	(0.0377)	(0.0003)	(0.0010)		
	B. Entrepreneurs and managers with own parent income in the highest tercile				
Expenditure elasticity	0.3518***	0.0037**	0.0105**	1.23	4319
	(0.0661)	(0.0018)	(0.0051)		
Rich share (60k)	0.3518***	0.0008**	0.0023**	0.62	4319
	(0.0661)	(0.0004)	(0.0010)		
Rich share (100k-30k)	0.3518***	0.0009*	0.0025*	0.70	4319
	(0.0661)	(0.0005)	(0.0014)		

Notes: The figure displays the estimates of the impact of parent income of two alphabetically nearest dormmates on the outcome at age 28-42, using parent income of two alphabetically nearest squadmates as the instrument. Parent income is in thousand euro. All regressions include linear terms for parent income, dummies for calendar year, and fully interacted service start year and squadron dummies. Standard errors allowing for clustering at the level of management unit are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.

Table 5: Falsification tests for the study peer design

		Pre-peer-exposure characteristics			
	Predicted residual industry expenditure elasticity	Own wage earnings	Average parent wage earnings	Own years of education	Average parent years of education
Peer parent income (leave-own-out mean)	-0.00012	-0.0184	-0.0492	0.0030	-0.0008
	(0.00012)	(0.0281)	(0.0787)	(0.0033)	(0.0049)
Observations	155,542		36,884		

Notes: The table displays the estimates of the impact of parent income of study peers at university on the predicted residual industry expenditure elasticity and pre-peer-exposure characteristics. Pre-peer-exposure characteristics are measured one year before a student enters the study program. Parent income is in thousand euro. All regressions include dummies for gender, year of outcome measurement, age of outcome measurement, school-by-program fixed effects, and school-by-start-year fixed effects. Standard errors robust for clustering by school-program-start-year are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.

Table 6: The Impact of Study Peer Parent income on Industry Expenditure Elasticity

	Entrepreneurs			
	Baseline	Low own parent income	High own parent income	Entrepreneurs and employees
Peer parent income (leave-own-out mean)	0.00181**	0.00083	0.00295*	0.00013
	(0.00092)	(0.00140)	(0.00153)	(0.00023)
Observations	155,542	97,210	57,656	2,073,402

Notes: The table displays the estimates of the impact of parent income of study peers at university on the industry expenditure elasticity of entrepreneurship/employment. Outcome is measured from age 28 onward, conditioning that the student has entered the program. Parent income is in thousand euro. All regressions include dummies for gender, year of outcome measurement, age of outcome measurement, school-by-program fixed effects, and school-by-start-year fixed effects, and the following control variables measured one year before a student enters the study program: years of education, wage earnings, average parent years of education, and average parent wage earnings. Standard errors robust for clustering by school-program-start-year are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.

Table 7: Distributional Effects from “Equal Opportunity Innovator Pools”

Sample	Counterfactual Innovator Pool	Distributional Effects rel. to Baseline Pool	Uses estimate from
Within phone apps	Female founder instead of male founder	**46.4%** larger gains for female users	Table 3.A spec. 1
Within consumer packaged goods	Female founder instead of male founder	**27%** larger gains for HHs with female head	Table 3.B spec. 1

Notes: See Section 4 for a description of the methodology..

Appendix to "What Drives Inclusive Innovation? The Importance of Innovators' Backgrounds"

Elias Einio, *VATT Institute for Economic Research*

Josh Feng, *University of Utah – Eccles School of Business*

Xavier Jaravel, *London School of Economics*

June 2020

A Model Solutions

A.A *Solving the Basic Model*

The intuition behind solving all variants of the model is that along a balance growth path, we need to allocate labor to research and production in a way that creates

1. Balance, in the form of the Euler equation. Allocating too much to research doesn't make sense if consumers are impatient.

2. Correct incentives, in the form of entrepreneur indifference equations

These equations will pin down the labor allocation along a balanced growth path. The quickest way to arrive at the balanced growth path solution is to take the following steps:

1. Along a balanced growth path, there will be some equilibrium interest rate r^*

2. People are indifferent between production work and entrepreneurship: $\eta N(t)V(t) = w(t)$, where $V(t)$ is the NPV of profits per variety

3. Per-period profits from entrepreneurship is governed by consumer preferences (pins down the markup) and amount of production labor used (divided equally between the varieties): $\pi(t) = \frac{1}{\epsilon-1}\frac{L-L_R(t)}{N(t)}w(t) = \frac{1}{\epsilon-1}(L-L_R(t))\eta V(t)$. The second expression comes from plugging in the indifference condition.

4. $V(t) = \frac{\pi(t)}{r^*}$ based on balanced growth path interest rates. Plugging in allows us to express $r^* = \frac{\eta}{\epsilon-1}(L-L_R(t))$

5. Finally, we can plug everything into the Euler equation, which ensures that the entrepreneurial labor allocation and growth in varieties accords with time preferences. The main thing to note here is that we can use household preferences to express the relationship between the growth rate of consumption and varieties $C(t) \sim N(t)^{\frac{1}{\epsilon-1}}$

$$
\begin{aligned}
r(t) - \rho &= \frac{1}{\epsilon - 1}\frac{\dot{N}(t)}{N(t)} \\
\frac{\eta}{\epsilon - 1}(L - L_R(t)) - \rho &= \frac{1}{\epsilon - 1}\eta L_R^* \\
\frac{\eta}{\epsilon - 1}L - \rho &= \frac{2}{\epsilon - 1}\eta L_R^* \\
L_R^* &= \frac{L}{2} - \frac{\epsilon - 1}{2\eta}\rho
\end{aligned}
$$

A.B Heterogeneous Research Productivity

We start off with the equations discussed in the main text:

$$
\frac{\dot{N}(t)}{N(t)} = \eta L \int_{\bar{\kappa}}^{1} \kappa f(\kappa) d\kappa = \frac{1 - \bar{\kappa}^2}{2}\eta L
$$

$$
\bar{\kappa}\eta N(t) V(t) = w(t)
$$

The monopolist profits per line per period is governed by the number of production workers:

$$
\pi(\nu, t) = \frac{1}{\epsilon - 1}\frac{\bar{\kappa}L}{N(t)}w(t) = \frac{1}{\epsilon - 1}\bar{\kappa}^2 \eta L V(t)
$$

We can then solve for the balanced growth path interest rate

$$
\begin{aligned}
V(t) &= \frac{\pi(t)}{r^*} \\
r^* &= \frac{\bar{\kappa}^2 \eta L}{\epsilon - 1}
\end{aligned}
$$

We can then plug everything into the Euler equation

$$
\begin{aligned}
\frac{\dot{C}(t)}{C(t)} = r(t) - \rho &= \frac{1}{\epsilon - 1}\frac{\dot{N}(t)}{N(t)} \\
\frac{\bar{\kappa}^2 \eta L}{\epsilon - 1} - \rho &= \frac{1}{\epsilon - 1} \cdot \frac{1 - \bar{\kappa}^2}{2}\eta L
\end{aligned}
$$

We can then solve this out to arrive at the results in the text.

Turning to the case with a wedge for one of the groups, the equations in the main text show the indifference conditions and the relationship between the two cutoffs. This means that the overall innovation rate can be expressed as:

$$\begin{aligned}\frac{\dot{N}}{N} &= \frac{1-\bar{\kappa}_1^2}{4}\eta L + \frac{1-\bar{\kappa}_2^2}{4}\eta L \\ &= \frac{\eta L}{2}\left(1 - \frac{1+(1-\tau)^2}{2}\bar{\kappa}_2^2\right)\end{aligned}$$

We can solve out the model, expressing everything in terms of $\bar{\kappa}_2$:

$$\begin{aligned}c(\nu,t) &= l(\nu,t) = \frac{(\bar{\kappa}_1+\bar{\kappa}_2)L}{2N(t)} = \frac{(2-\tau)\bar{\kappa}_2 L}{2N(t)} \\ C(t) &= \frac{2-\tau}{2}\bar{\kappa}_2 L N(t)^{\frac{1}{\epsilon-1}} \\ \pi(\nu,t) &= \frac{1}{\epsilon-1}\frac{(2-\tau)\bar{\kappa}_2 L}{2N(t)}w(t) \\ &= \frac{1}{\epsilon-1}\cdot\frac{(2-\tau)(1-\tau)}{2}\bar{\kappa}_2^2\eta L V(t) \\ r^* &= \frac{1}{\epsilon-1}\cdot\frac{(2-\tau)(1-\tau)}{2}\bar{\kappa}_2^2\eta L\end{aligned}$$

We can then plug this back into the Euler equation to solve for $\bar{\kappa}_2$:

$$\frac{1}{\epsilon-1}\frac{(2-\tau)(1-\tau)}{2}\bar{\kappa}_2^2\eta L - \rho = \frac{1}{\epsilon-1}\cdot\frac{\eta L}{2}\left(1 - \frac{1+(1-\tau)^2}{2}\bar{\kappa}_2^2\right)$$

$$\bar{\kappa}_2^2\left((2-\tau)(1-\tau) + \frac{1+(1-\tau)^2}{2}\right) = \frac{2\rho(\epsilon-1)}{\eta L} + 1$$

$$\bar{\kappa}_2^2 = \frac{1+\frac{2\rho(\epsilon-1)}{\eta L}}{(3\tau^2-8\tau+6)}$$

which leads to the growth rate presented in the main text. The denominator is decreasing on the interval from zero to one, which means that $\frac{\partial\bar{\kappa}_2}{\partial\tau} > 0$. Greater barriers mean that fewer people from the restricted group enter entrepreneurship. Using the fact that $\bar{\kappa}_1 = (1-\tau)\bar{\kappa}_2$, we can show that $\frac{\partial\bar{\kappa}_1}{\partial\tau} < 0$.

In terms of inequality in earnings, people working in production will earn the prevailing market wage w. So do the marginal entrepreneurs in each group. Inframarginal entrepreneurs earn at some multiple relative to the marginal entrepreneur $\frac{\kappa}{\bar{\kappa}}w$, where $\bar{\kappa}$ is the equilibrium cutoff. The total wages for a group, conditional on the equilibrium cutoff $\bar{\kappa}$, will be:

$$
\begin{aligned}
W &= \int_0^{\bar{\kappa}} wf(\kappa)d\kappa + \int_{\bar{\kappa}}^1 \frac{\kappa}{\bar{\kappa}} wf(\kappa)d\kappa \\
&= w\left(\kappa + \frac{1}{2\bar{\kappa}} - \frac{\bar{\kappa}}{2}\right) \\
&= \frac{w}{2}\left(\bar{\kappa} + \frac{1}{\bar{\kappa}}\right)
\end{aligned}
$$

Because $0 \leq \bar{\kappa} \leq 1$, the total wage will be decreasing in $\bar{\kappa}$:

$$\frac{\partial W}{\partial \bar{\kappa}} = \frac{w}{2}\left(1 - \frac{1}{\bar{\kappa}^2}\right) < 0$$

Given that $\bar{\kappa}_1 = (1-\tau)\bar{\kappa}_2 < \bar{\kappa}_2$, the wedge will decrease the relative wages of the restricted group.

A.C Two-Sector Model

Starting with the basic setup in the main text, we can derive the relationships between research and manufacturing across sectors, and then plug into the Euler equation to pin down the absolute levels.

On a balanced growth path,

$$r^* = \frac{1}{\epsilon - 1} \frac{N(t)}{N_i(t)} L_{iM}(t)\eta_i, \forall i$$

This implies that $\eta_1 L_{1M}/N_1(t) = \eta_2 L_{2M}/N_2(t)$.

The amount of labor needed to produce varieties to accommodate Cobb-Douglas tastes will be $\frac{L_{1M}}{L_{2M}} = \frac{\alpha}{1-\alpha}$. Put together, this means that along a BGP, varieties will satisfy

$$\frac{N_1(t)}{N_2(t)} = \frac{\alpha\eta_1}{(1-\alpha)\eta_2}$$

The BGP interest rate can then be written in terms of one of the production labor variables:

$$r^* = \frac{1}{\epsilon - 1}(\eta_1 + \frac{1-\alpha}{\alpha}\eta_2)L_{1M}$$

Because varieties grow at the same rate over time, this pins down the research allocation ratio across sectors $\frac{L_{1R}(t)}{L_{2R}(t)} = \frac{\frac{\alpha}{1-\alpha}\eta_1 + \eta_2}{\eta_1 + \frac{1-\alpha}{\alpha}\eta_2} = \frac{\alpha}{1-\alpha}$

We can then plug everything into the Euler equation to establish a relationship between manufacturing and research:

$$L_{1M} = \frac{\alpha\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)\eta_2)} + L_{1R}$$

And then we can solve everything out to arrive at allocations and equilibrium growth rate.

A.D Two-Sector Model with Access Barriers

Here, we derive the solution to a two-sector model where one sector has a wedge. This can be applied specifically to the case with innovator specialization and access barriers, where the group facing barriers still ends up entering into entrepreneurship. We highlight the differences in the equations that pin down a balanced growth path solution.

Let there be a wedge $1-\tau$ in sector 2. The labor market indifference equation becomes

$$w(t) = \eta_1 N(t) V_1(t) = (1-\tau)\eta_2 N(t) V_2(t)$$

The production labor allocation remains the same, because tastes have not changed. On a balanced growth path, varieties will satisfy

$$\frac{N_1(t)}{N_2(t)} = \frac{\alpha\eta_1}{(1-\alpha)(1-\tau)\eta_2}$$

The equilibrium interest rate will be $r^* = \frac{1}{\epsilon-1}(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2)L_{1M}$. The ratio of research labor devoted to each sector is now $\frac{L_{1R}(t)}{L_{2R}(t)} = \frac{\frac{\alpha}{(1-\alpha)(1-\tau)}\eta_1+\eta_2}{\eta_1+\frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2} = \frac{\alpha}{(1-\alpha)(1-\tau)}$, based on equal growth rates in varieties and labor market indifference. Using the Euler equation, we can pin down the

$$\begin{aligned}
\frac{\dot{C}(t)}{C(t)} = r(t) - \rho &= \frac{\alpha}{\epsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\alpha}{\epsilon-1}\frac{\dot{N}_2(t)}{N_2(t)} \\
\frac{1}{\epsilon-1}(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2)L_{1M} - \rho &= \frac{1}{\epsilon-1}\left(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2\right)L_{1R}
\end{aligned}$$

We can then plug it into the labor budget constraint to solve out for research:

$$L_{1R} = \frac{\alpha}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2)}\right)$$

The equilibrium growth rate will be

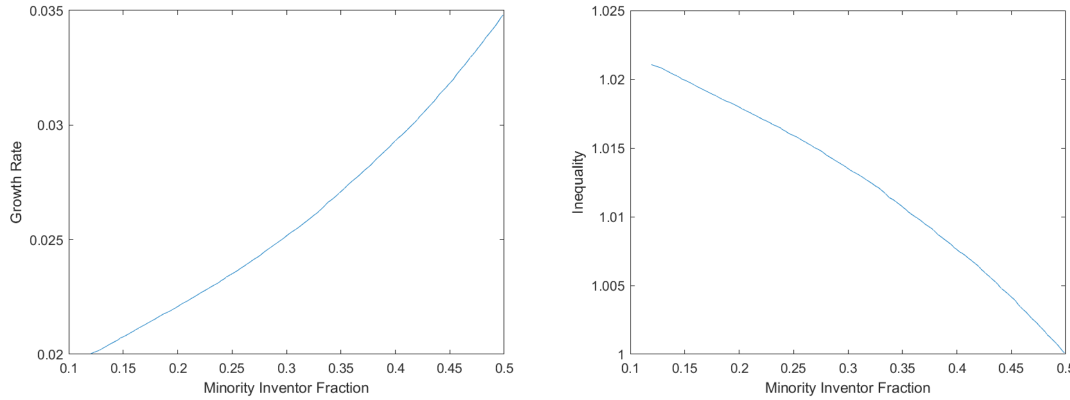
$$
\begin{aligned}
g^* &= \frac{\dot{C}(t)}{C(t)} = \frac{1}{\epsilon - 1}\left(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2\right) L_{1R} \\
&= \frac{1}{\epsilon - 1}\left(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2\right) \cdot \frac{1}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2)}\right) \\
&= \frac{1}{1+\alpha+(1-\alpha)(1-\tau)}\left(\frac{1}{\epsilon-1}\left(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2\right) L - \rho\right)
\end{aligned}
$$

This formula reduces back to the simple two-sector model when $\tau = 0$. An interesting aspect of the solution is that there are cases where the wedge is actually good for long-run growth rates, because the wedge skews allocation towards the more productive research sector. People move from innovating in sector 2 to innovating and producing in sector 1. However, in the case where research productivities are equal across sectors, the expression is clearly decreasing in the size of the wedge.

A.E Graphical Illustration

We can graphically illustrate the impact of barriers and ability differences on inequality and long-run growth in the two-sector model.

Figure A1: Simulations of Heterogeneous Research Productivity Model



B Calibrating the Growth Model

We calibrate the growth models presented in Section V.

B.A *One-Sector, Pareto Productivity*

First, we calibrate the model with heterogeneous research productivity, using a Pareto productivity distribution and "real" barriers for 50 percent of the population ("minority" group). The approach is as follows:

1. Fix the following parameters based on estimates in the literature: $\epsilon = 10, \rho = 0.95$. We also set a Pareto distribution in research productivity with $x_m = 1, \alpha = 1.5$.

2. Calibrate τ, η so that $g^* = 0.02$ and the fraction of "minority" group in the pool of inventors is 0.12, recovering $\tau = 0.7423, \eta = 0.225$

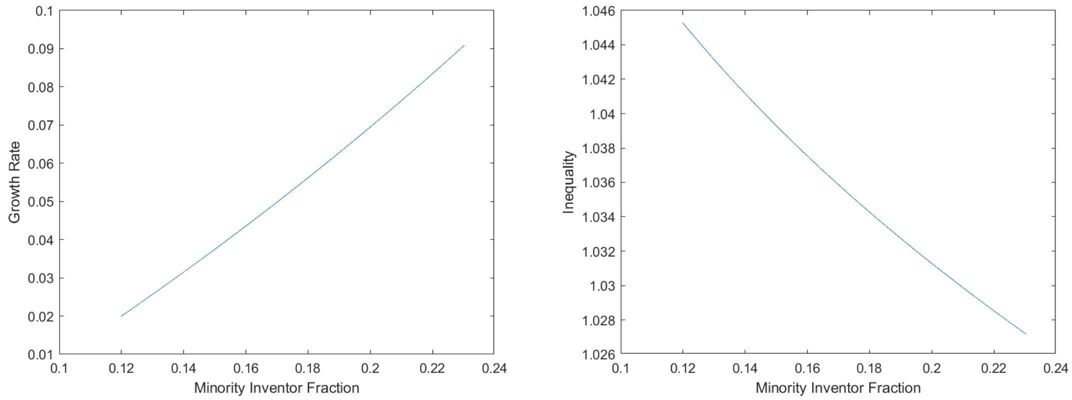
3. Simulate a reduction in τ, plotting how growth rates, average inequality in earnings for the two groups, and "minority" inventor fraction change

We plot the results of the calibration in Figure A1. The simulation shows that as we reduce τ towards zero, the growth rate increases to about 3.5%. Income inequality driven by inframarginal earnings for entrepreneurs also decreases, from 2% to 0%.

B.B *Basic Two-Sector Model*

Second, we calibrate the two-sector model with "real" barriers and higher own-sector research productivity. We take a similar approach to the one above:

Figure A2: Simulations of Multi-Sector Model



1. Fix the following parameters based on estimates in the literature: $\epsilon = 10, \rho = 0.95$. We also set preference parameters $\alpha = 0.8, \alpha' = 0.2, \delta = 0.5$. Finally, we set $\eta' = 0$ in order to focus on the case where inventors specialize.

2. We calibrate τ, η to again have $g^* = 0.02$ and the fraction of "minority" group in the pool of inventors is 0.12, recovering $\tau = 0.8636, \eta = 15.68$

3. We then reduce τ gradually and plot the response in growth rates, inequality through the expenditure channel, and "minority" inventor fraction.

Figure A2 plots the results of the simulations. Again, we find that reducing barriers increases growth rates and reduces inequality, this time across consumer groups through the expenditure channel. The quantitative results are more stark here, because we have assumed homogeneous research productivity.

B.C Two-Sector, Pareto Productivity

Finally, we calibrate a version of the two-sector model that has agents with Pareto research productivity. The setup is similar to that of the basic two-sector model, with the following changes:

1. There are again two groups of agents with different tastes: $\alpha = 0.6, \alpha' = 0.4, \delta = 0.5$

2. Agents exhibit a Pareto distribution in research productivity with $x_m = 1, \alpha = 1.5$

3. The minority group suffers a real productivity hit of $1 - \tau$.

4. Agents are randomly assigned a sector that they can innovate for. A parameter ϕ is used to capture the rate of specialization. Agents have a probability $\frac{1+\phi}{2}$ of being assigned to their

"own" sector, the one they relatively prefer as a consumer. This structure helps simplify the problem by eliminating sectoral choice.

Given this setup, we solve for real productivity cutoffs in each sector. Individuals with productivities above the cutoff in the sector they are assigned to become entrepreneurs. The cutoffs have to satisfy: i) allocation of research labor to ensure equal growth rates of varieties ii) the Euler equation condition iii) indifference of marginal agents in each sector between pursuing entrepreneurship and wage work. Given the cutoffs, we can compute the BGP growth rate and inequality.

We calibrate the parameters τ, η, ϕ by matching $g^* = 0.02$, the fraction of "minority" group in the pool of inventors is 0.12, and matching the regression coefficient β from Equation 1. For the results presented below, we match to a coefficient of 0.1, which is in line with the phone app estimates. We run 100 simulations and report average statistics. In addition, Figures A3 and A4 report detailed results from one such simulation, including the significant change in relative market size needed to achieve the same outcome as reducing barriers or reducing specialization.

In the calibration, we recover, on average, $\tau = 0.8603, \eta = 0.5173, \phi = 0.6850$. This suggests high barriers for minority entrepreneurs and high rates of specialization. Given the calibration, we compute the following statistics:

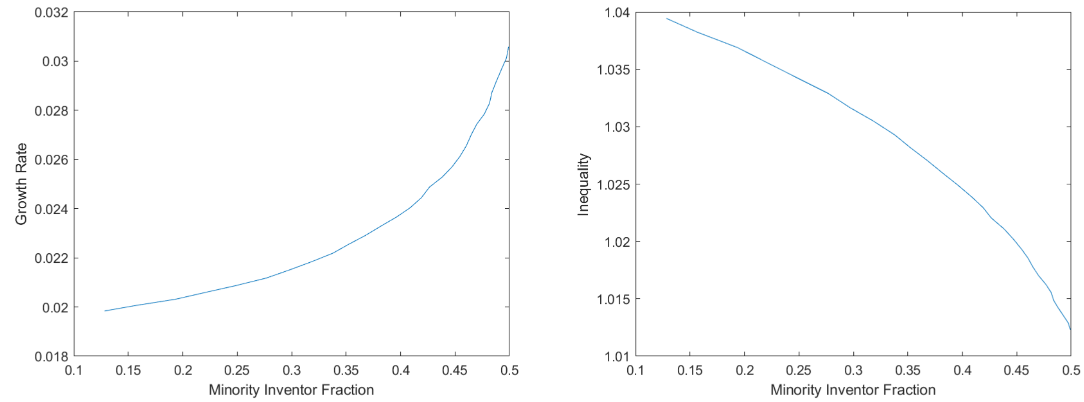
1. Reduce τ to zero: this is similar to the exercises above. We find that removing barriers leads to a BGP growth rate of 0.041, similar to the single-sector result above, and reduces inequality from 1.024 to 1.

2. Reduce ϕ to zero: this reduces the rates of specialization, so more productive majority group innovators are able to innovate for the minority group. Under this scenario, reducing specialization leads to a slight increase in growth rate to 0.0222, because the innovation productivity is more aligned with consumer tastes, and leads to a reduction in inequality from 1.024 to 1, because more productive individuals are able to innovate for the minority group.

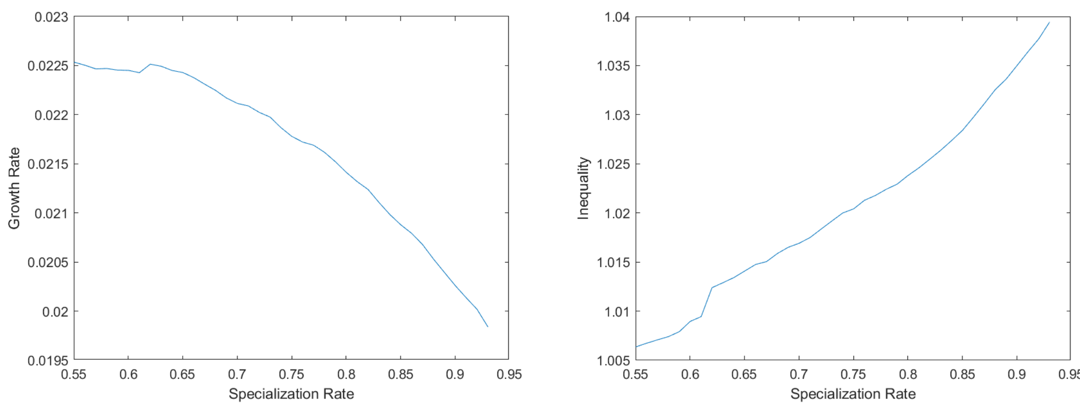
3. Increase δ to 0.6 while holding fixed the innovator pool: this is akin to expanding the relative market size of the minority group through redistribution. We find that, on average, the growth rate stays about the same 0.0203,while inequality falls from 1.024 to 1.006.

Figure A3: Simulations of Multi-Sector Model

Panel A: Varying Barriers



Panel B: Varying Specialization



Panel C: Varying Relative Market Size

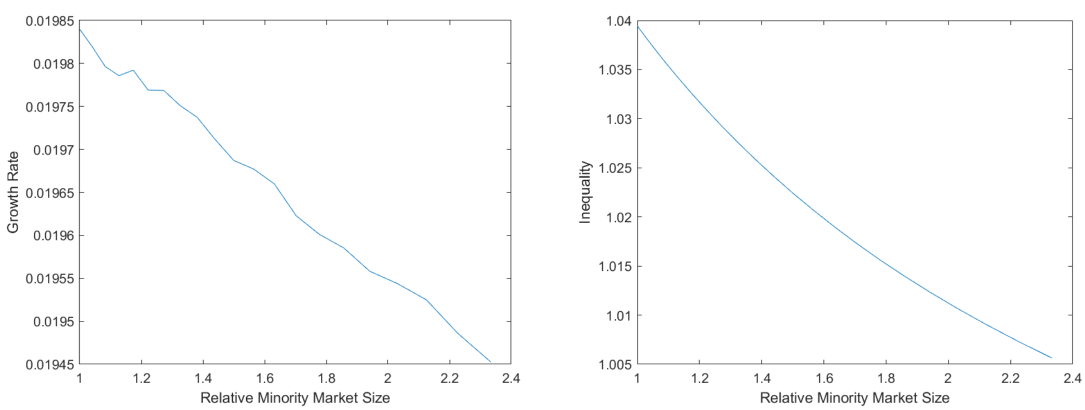
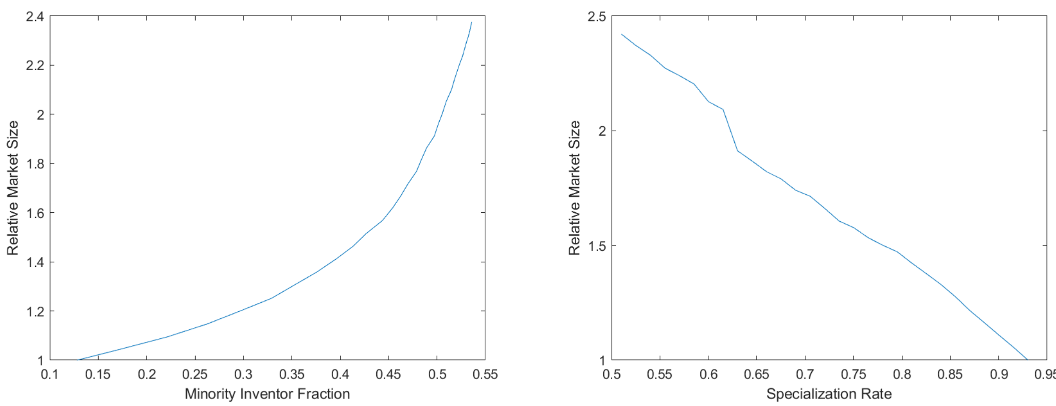


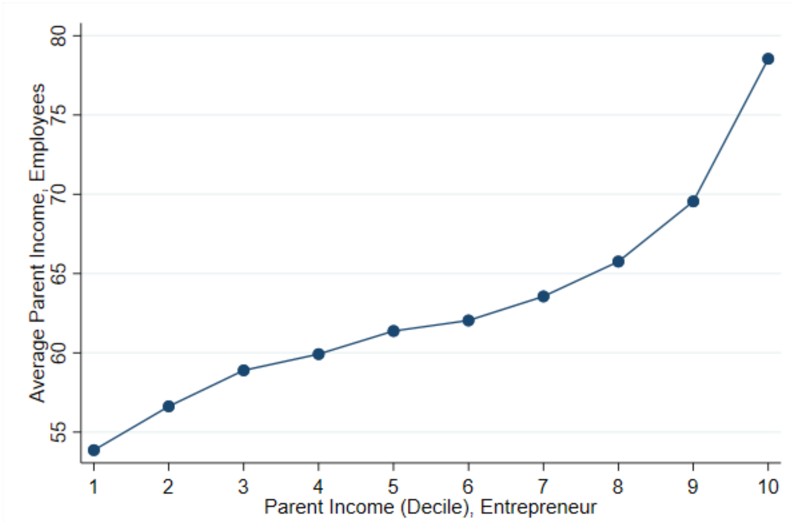
Figure A4: Equivalent Market Size Change to Reduce Inequality



Notes: graphs investigating the necessary change in relative market size in order to achieve the same reduction in inequality as: i) reducing τ to increase the minority inventor fraction ii) reducing ϕ (specialization).

C Additional Figures and Tables

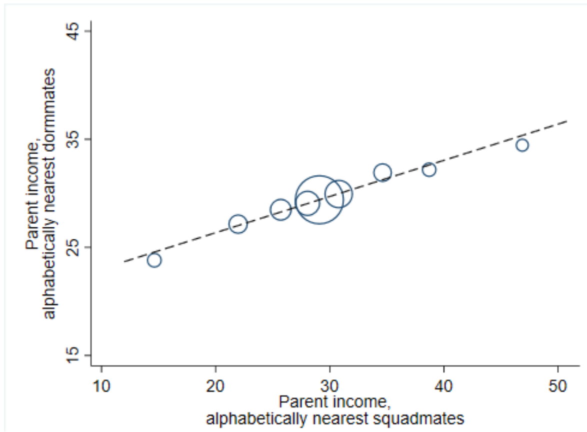
Figure A5: Parent income of employees by entrepreneur's parent income



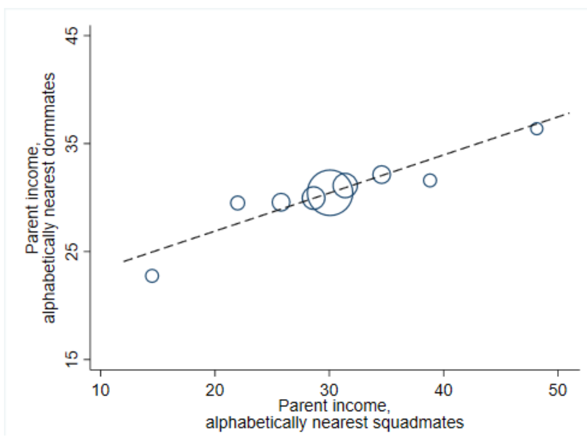
Notes: The coefficient (standard error) from a regression of log average employee parent income on log of entrepreneur's parent income is 0.1778 (0.0053). The number of observations is 127,627. Data sources: Finnish FLEED and FOLK datasets, years 2010-2016.

Figure A6: First stage regression of parent income of alphabetically nearest dormmates on parent income of alphabetically nearest squadmates

A. Entrepreneurs and managers



B. Entrepreneurs and managers with
own parent income in the highest tercile



Notes: The figure displays the first-stage relationship between parent income of two alphabetically nearest squadmates and dormmates in a sample including entrepreneurs and managers (panel A) and a sample restricted to entrepreneurs and managers in the highest own parent income tercile (panel B). The figure plots the residuals from separate regressions of dormmate and squadmate parent income on own parent income, squadron fixed effects, and year dummies. The size of the circle represents the number of conscripts within each bin. Data sources: Finnish FLEED and FOLK datasets and FDF conscript registry.

Table A1: Validation regressions

	Independent variable				
	Own parent income		Parent income, two alphabetically nearest squadmates		Dependent
Dependent variable	coeff.	s.e.	coeff.	s.e.	mean
Wage earnings	0.0041*	(0.0023)	-0.0020	(0.0025)	5.00
Employed	0.0419**	(0.0175)	-0.0180	(0.0220)	39.58
Years of schooling	0.0054***	(0.0005)	0.0009	(0.0006)	10.43
Married	-0.0044***	(0.0017)	0.0035	(0.0022)	0.27
Foreign	-0.0002	(0.0002)	0.0003	(0.0002)	0.00
Primary language Finnish	0.0256***	(0.0056)	-0.0045	(0.0073)	95.78
Unemployment benefits	-0.0045***	(0.0003)	0.0002	(0.0003)	0.28
General housing allowance	-0.0086***	(0.0003)	-0.0002	(0.0004)	0.26
Parent wage earnings	0.8079***	(0.0071)	-0.0011	(0.0050)	23.74
Parent employed	0.8180***	(0.0153)	-0.0132	(0.0132)	79.12
Parent years of schooling	0.0947***	(0.0011)	0.0007	(0.0011)	12.30
Parent pension income	-0.0210***	(0.0013)	-0.0001	(0.0015)	1.11
Parent unemployment benefits	-0.0299***	(0.0007)	0.0010	(0.0009)	0.94
Parent general housing allowance	-0.0080***	(0.0003)	-0.0002	(0.0003)	0.20

Notes: Table from Einiö (2019). Each row presents coefficients from a separate regression on the dependent variable indicated by the row label. Outcomes are measured in the year preceding the start of service. Income, earnings, benefits, allowances, and pensions are in thousand euro. All regressions include fully interacted service start year and squad dummies. N=50,578. Standard errors allowing for clustering at the level of management unit are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.